MORAL SAINTS RECONSIDERED

by

Vanessa Carbonell

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Philosophy)
in The University of Michigan
2009

Doctoral Committee:

Emeritus Professor Stephen L. Darwall, Co-Chair
Professor Elizabeth S. Anderson, Co-Chair
Professor Peter A. Railton
Associate Professor Mika T. Lavaque-Manty

ACKNOWLEDGEMENTS

First I would like to express gratitude for the guidance of my dissertation committee: Stephen Darwall, Elizabeth Anderson, Peter Railton, and Mika Lavaque-Manty. I am especially thankful to Steve Darwall for his wise and warm stewardship of my project over four years. This dissertation would not have existed at all if he had not shown enthusiasm for my budding interest in moral saints way back when I had absolutely no idea what I wanted to say about them. Nearly every page of this document was shaped by his thoughtful criticism. And he deserves special mention for what is surely the fastest feedback turnaround time of any advisor in the world! I am also especially thankful to Liz Anderson, whose guidance and feedback—not just on this dissertation but on every piece of philosophical writing I have sent her over the years—has been absolutely invaluable. I owe an immense debt to all her help, including her encyclopedic knowledge of philosophy and her keen insight into connections between philosophy and other fields of inquiry, especially the empirical and social sciences. Her positive feedback over the years has often kept me going when I was tempted to give up and start from scratch. I also want to thank Peter Railton for being a generous and charitable interlocutor, and for teaching me so much philosophy over six years. I remain in awe of Peter's ability to make my half-baked ideas sound smart and interesting, and to see the philosophical forest when I can only see the trees. I feel immensely lucky to have learned ethics from Steve, Liz, and Peter. And finally I want to thank Mika Lavaque-

Manty for generously agreeing to join my committee at a late stage and offering helpful feedback as I prepared for the job market.

I also want to thank Allan Gibbard and David Velleman, both of whom were wonderful teachers of ethics and gave careful and insightful feedback on work I did before my dissertation. Thanks also to Jamie Tappenden and Eric Lormand, for helping me work through ideas in Candidacy Seminar. I am also grateful to all the members of the Michigan faculty who gave feedback on my work at my Brown Bag talk, including Sarah Buss and Victor Caston. Last but not least, I owe a huge debt of gratitude to Louis Loeb, the hardest working Placement Coordinator in the world. His tireless, meticulous efforts were supererogatory to say the least, and I would not have survived the job market were it not for his kind and patient coaching.

Many current and past Michigan graduate students—too many to name—have offered me both astute philosophical criticism and much-needed moral support. In particular, my two brother-from-another-mother's, Ivan Mayerhofer and Dustin Locke, have been my most trusted partners in crime, and I am forever grateful to them for all the practice talks, endless emails and phone calls, burrito therapy sessions, travel companionship, jaded commentary and morbid humor. I also want to thank Erica Lucast Stonestreet for her moral support and friendship. I'd be nowhere without Howard Nye, John Ku, Aaron Bronfman, David Dick, and Tim Sundell. I thank them for helping me be better at philosophy, for great friendship, and for being their irreplaceable selves. And I must thank non-philosophers Cat Wu, Sarah Massey, Rebecca Cohen, Jo Russ, and Geoff Hill, for putting up with my gradual descent into professional nerd-dom over the past decade.

The members of Michigan philosophy department staff—without whom the place would simply fall apart—have helped me navigate the administrative minutiae of doctoral education, always with good cheer. In particular, I am indebted to Linda Shultes; over the last six years she made my life significantly easier in ways both seen and unseen. Thanks also to Molly Mahony for being the best philosophy librarian in the world. My work on this dissertation was supported by funding from the Michigan philosophy department, the Rackham School of Graduate Studies, and the Woodrow Wilson Foundation on behalf of the Charlotte Newcombe Foundation. I am extremely grateful for the support.

Finally, I would like to thank my parents, Jean and Armando, for their unconditional support. I also want to thank my amazing nephews Wyatt and Ethan, for being so much fun to hang out with during our all-too-brief visits, and for never holding my long absences against me.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Suppose there are people who deserve to be called "moral saints," and these people do not have any magical powers or disturbing pathologies. That is, suppose there are otherwise normal human beings whose lives are marked by extraordinary moral accomplishments: they do far more than we think morality requires of them, they exhibit resolve and tenacity when others would give up, and they bear heavy burdens of personal sacrifice. What would these people look like? How would they be motivated? And what does the existence of such people mean for the rest of us? Do moral saints serve a purpose other than to make us feel bad about ourselves?

The primary aim of this dissertation is to answer these questions. Along the way, I hope to begin to reestablish the importance of moral saints to ethical theory, and the importance of real-life moral saints as moral exemplars.

Preliminaries: Saints and Supererogation

Let me begin with a note about terminology. Philosophers refer to actions that are morally good, but not morally required, as "supererogatory." These actions are often described as "beyond the call of duty" or "good to do but not bad not to do." In what has become a standard definition, Gregory Mellema claims three conditions must be met in

order for an act to qualify as supererogatory: "First, it is an act whose performance fulfills no moral duty or obligation. Second, it is an act whose performance is morally praiseworthy or meritorious. Third, it is an act whose omission is not morally blameworthy" (1991, 3). But the first systematic account of supererogatory actions and their proper role in moral theory came much earlier, in J.O. Urmson's 1958 article "Saints and Heroes." Urmson, however, only uses the term "supererogatory" once (1958, 214). For the rest of the article he refers only to "saints" and "heroes" or "saintly" and "heroic" actions. For Urmson, "saintly" and "supererogatory" are apparently synonyms. A "saint" is simply someone who performs a supererogatory action.

Urmson distinguishes a certain technical sense of the term "saint" (or "hero") from various popular uses of the term with which it might be confused. He's not interested in the religious sort of saint, nor the nonmoral "hero of the soccer game." Nor is he interested in someone who overcomes fear, desire, self-interest, or a drive to self-preservation to do his duty. Such a person distinguishes himself by outperforming his peers, but does no more than what morality requires. Instead, Urmson thinks a stronger notion of sainthood is what ethicists ought to be interested in. A person is a hero or a saint in this strong sense when he (1) "does actions that are far beyond the limits of his duty;" (2) either against inclination (saint) or despite fear (hero); (3) either by self-control or without effort (201). The paradigm case of this sort of moral saint is the soldier who jumps on a stray grenade to protect his comrades, where we cannot say of the other soldiers who stayed put that they failed in some *duty*, and we cannot say that a superior officer could have "decently ordered [the soldier] to do it" (203). And although it might

be the case that "the deed presented itself [to the soldier] as a duty," it was not a duty (203).

It should be noted that Urmson writes very little about *the soldier himself*. That is, he does not give an account of the sort of agent who might be likely to smother a grenade, or discuss whether one grenade-smothering act is sufficient for sainthood (although he implies as much). Nor does he examine saintliness as any sort of disposition or pattern of behavior over time. Thus his use of the terms "saint" and "hero" might be considered misleading. His account is really an account of *supererogatory actions*, not the agents who perform them.

Urmson thinks supererogatory actions are important because they challenge the traditional conception of moral actions as being exhaustively categorized into the obligatory, the permissible, and the forbidden. Here Urmson is using "permissible" to refer to cases of what we might call *pure permissibility*, where either of two options is allowed by morality but nether is favored, morally, more than the other. Jumping on a grenade is surely neither required nor forbidden. However, jumping cannot be *permissible* in Urmson's sense because morality is not strictly indifferent between jumping and not jumping; jumping on the grenade must, in *some* sense, have a different moral status than merely standing idly by. Urmson thinks moral theories must make room for the class of actions that resemble grenade-smothering: actions that are not duties to "be exacted from persons as a debt," but rather exhibitions of the "higher flights of morality" (209, 211).

Urmson argues that utilitarian, Kantian, and intuitionist moral theories cannot explain supererogatory actions, insofar as these theories depend on the strict trichotomy

of obligatory, permissible, and forbidden. In making this claim, Urmson issued a challenge that has shaped the tone of the debate ever since. One way of responding to this challenge is to argue that one of these historical theories, properly interpreted, *does* allow for supererogatory actions. Thomas Hill takes this approach in his article "Kant on Imperfect Duty and Supererogation" (1971). Hill argues that proper attention to the different sorts of duty in Kant's theory reveals that "there is more room for choice in pursuing moral ideals, and not everything good is required" (1971, 55).

Another way of responding to Urmson's challenge, though, is to argue that a theory's inability to account for supererogatory actions is no weakness, either because supererogation is not a distinct deontic category or because accommodating supererogatory actions is not a desideratum of an ethical theory. Marcia Baron takes the latter approach in "Kantian Ethics and Supererogation" (1987). Baron thinks that to stretch Kant's theory so that it accommodates supererogation is to stretch it too far. Her approach can be called "anti-supererogationist," in that she downplays the value of supererogatory actions and is skeptical that they play any sort of interesting role in moral theory. Baron even suggests that to make parts of morality "optional"—as the supererogationist is apt to do—risks descending into an egocentric "yuppie ethics" (1987, 249).

The supererogation literature after Urmson seems to fall into two camps. In one camp are the philosophers who respond to his challenge by arguing either for or against expanding the trichotomy. This includes Hill (1971) and Baron (1987) as well Gregory Mellema (1991), David Heyd (1994), Roderick Chisholm (1986), and many others. Interestingly, the articles in this camp tend to have the word "supererogation" in the title

and tend not to make much reference to "saints." In the other camp, however, are philosophers who explore the *moral psychology* of supererogators, and the word "saint" can be found in nearly all of their titles. This includes Susan Wolf's famous paper "Moral Saints" (1982) as well as Robert Adams' "Saints" (1984), Daniel Haybron's "Moral Monsters and Saints" (2002), Edward Lawry's "In Praise of Moral Saints" (2002), Louis Pojman's "In Defense of Moral Saints" (2002), and many others. Of course, Wolf was following Urmson in using the term "saint" to refer to the agent who performs a supererogatory act, and the rest were simply responding to Wolf. What is problematic is that Urmson's scrupulous definition of the term "saint"—to refer to the agent who performs his precisely defined supererogatory action and not to refer to various other everyday language uses of the word—gets lost somewhere in the decades after his article "Saints and Heroes." The result is that philosophers studying the "moral saint" are studying various different conceptions of moral agents, including conceptions that are both explicitly and implicitly influenced by religious notions of sainthood, which Urmson specifically tried to avoid. Indeed, where Urmson intended "saint" simply to be the name for an agent who performs a one-time extraordinary deed (such as smothering a grenade), "saint" has since been appropriated for such diverse uses as an ideal moral agent or obsessive moral perfectionist (Wolf), "people in whom the divine can be seen" (Adams), and "good human beings" (Lawry). While each of these accounts might be interesting in its own right, it's not clear that these are accounts of the same type of agent, or even that any of these is an account of the most interesting type of agent.

By far, the most influential account of moral saints has been Wolf's (1982). Hers is a debunking account: it argues that moral saints, by their very nature, would be bland,

boring, obsessive and humorless, and that the rest of us should not aspire to be them, because they are not "unequivocally compelling personal ideals" (1982, 419). My account is meant to provide an alternative to Wolf's, and to be in direct dialogue with it. Of course, one might worry that, because I will define "moral saint" differently than Wolf, she and I are talking past each other—simply talking about different things. However, in Chapter 2 I aim to offer both an internal and an external critique of Wolf's account. By "internal critique" I mean I try to show that it fails on its own terms, taking for granted its definition and background assumptions about moral saints. And by "external critique" I mean I try to show that its definition and background assumptions are misguided, and that we have good reason to define moral saints differently. Where my critique is internal, it seems that the charge of talking-past cannot fairly be applied. And where the critique is external, we are talking past each other in a trivial sense—I mean something different by "moral saint" than she does, but I hope to give good reason to prefer my definition to hers.

A Non-Ideal Theory of Moral Saints

My aim is to give an account of moral saints that, like Wolf's, focuses on an agent's behavior over a lifetime rather than on a single action. Unlike Wolf's account, though, my account will not demand that a moral saint's every action be as good as possible. In short, my account of moral sainthood is meant to be (1) non-ideal; (2) psychologically realistic; (3) thoroughly secular; (4) informed by real-life case studies; and (5) neutral with respect to the major normative ethical theories. The account is non-ideal in the sense that it does not seek to describe an ideal—or even an *idealized*—moral

agent. That is, it is not an account of flawless, morally perfect agents. In setting aside discussion of the idea of moral perfection, I do not mean to suggest that it is an uninteresting or unimportant topic in moral theory. Rather, I have simply chosen to take a more pragmatic route and focus on a type extraordinary moral greatness that is answerable to facts about ordinary human psychology and manifested in real, living, non-pathological, non-robotic people.

With Urmson, I think it is important to distinguish moral sainthood from religious notions of sainthood. Some philosophical accounts of moral sainthood draw quite explicitly on religious concepts; Robert Adams, for example, conceives of moral saints as having a particular relation to the divine (1984). Even some philosophical accounts that are otherwise secular have been influenced by religious notions of sainthood, for example by taking virtues like purity and asceticism, which are not straightforwardly ethical notions, to be traits a moral saint must display. It's understandable that the moral and religious notions of sainthood have become entangled. After all, morality and religion have a long history of entanglement, and many religious saints achieved their sainthood through acts most of us would deem morally good. Indeed, many accounts of *moral* sainthood turn reflexively to examples—like Gandhi and Mother Teresa—who were both moral *and* religious figures, and for whom morality and religion were intimately connected. While it's possible someone like Gandhi or Mother Teresa would qualify for moral sainthood, it would be extremely difficult to evaluate their actions and motivations from a strictly moral standpoint given that religion loomed large in the background of so much of what they did. For the sake of clarity, then, I simply avoid examining cases where a person's moral motivations or actions are bound up with religious beliefs, or

where public admiration of a person (even purportedly *moral* admiration) is tied to a belief that the person holds some special religious status or significance.

Where can we find moral saints, then, if the usual suspects are ruled out, and if all ambiguously religious cases are set aside? Perhaps the next best place to look is in works of literature. Surely we could find a fictional character who qualifies for moral sainthood. However, I have chosen only to make use of real-life (in most cases, *living*) examples, for several reasons. First, just as anything that's *actual* is *possible*, we can also say that if a person is *real*, then she must by definition be at least minimally *psychologically realistic*. Of course, many fictional characters are highly realistic. And I don't mean to suggest we can't gain important insights about morality by examining works of fiction. Rather, I simply think that if one is particularly interested—as I am—in morality as a social enterprise that operates in communities of real people, and interested in whether moral exemplars and outliers of various sorts could potentially influence the moral demands faced by the rest of us, then one should focus on examining case studies of *real* people. After all, fictional characters are limited only by the imaginations of their creators; perhaps a highly convincing fictional character could perform a moral feat that would, in fact, be beyond the reaches of a real person. Insofar as we want our moral system to be answerable to empirical facts (physical, psychological, sociological, etc.), it seems unwise to make use of fictional characters as data points.

Using real-life, and in particular *living*, case studies also allows us to avoid mistaking lore or legend with reality. We don't have to worry that a person's moral accomplishments have been exaggerated by centuries of myth or obscured by unverifiable half-truths. Indeed, living cases are *especially* useful for my own argument,

because I am interested in the ways moral saints provide the rest of us with evidence about ways of living. Presumably, the fact that the moral saints live in the same historical era as we do will increase the quality of that evidence.

Finally, it should be noted that the account of moral sainthood I am proposing is at least in some weak sense neutral with respect to which normative ethical theory is correct. I intend for my account neither to presuppose nor to provide evidence for any particular ethical theory. What my account *does* presuppose is a relatively clear distinction between actions that are obligatory and those that are supererogatory. This distinction features prominently in ordinary morality, but is sometimes thought to cause problems for the historically significant ethical theories. Whether there is room for supererogation in a broadly Kantian deontological theory has been disputed.[1] Similarly, on a straightforward interpretation of utilitarianism, one is always *required* to perform the optimific action; this seems to leave little room for actions that are good but not required. It is even more difficult to understand supererogation from the standpoint of virtue ethics, since the central moral concept there is virtue and not obligation. Of course, there may be certain "large-scale" virtues (e.g., magnificence and magnanimity) the exercise of which gives an agent *extra* moral merit in some sense. And virtue ethics certainly could accommodate the equivalent of a moral saint—the most virtuous person, or the person who is more virtuous than necessary. But as I will discuss in Chapter 2, understanding moral saints in terms of particular virtues or character traits can be tricky.

Thus I approach the study of moral sainthood from the standpoint of ordinary or commonsense morality, which borrows principles from both deontological and

[1] See, e.g., Hill (1971) and Baron (1987).

consequentialist theories. Of course, it is impossible to remain *completely* neutral with respect to all questions of normative ethics. For example, I shall presuppose that helping others—contributing to their well-being—is morally good and in many cases morally required. This will be evident in my examples of moral saints—Holocaust rescuers, doctors, and other caregivers will feature prominently. However, one should avoid the temptation to read any deeper into the examples than this. For instance, one might think that because I have chosen examples that tend to involve helping *lots* of people—a doctor who helps thousands of patients, a mom who adopts a dozen children—that I am operating on some version of a consequentialist framework according to which a moral saint must help *as many people as possible*. However, in fact I think in principle a person could be a moral saint even if her deeds reach only one person.[2] It's simply harder to come by real-life examples of this sort. But we can easily imagine one: perhaps someone brings meals to a lonely disabled neighbor to whom she is not related, day in and day out for decades on end at great sacrifice.

Indeed, it might also be tempting to conclude from my examples that I think the *only* route to sainthood is through some sort of *helping* behavior. This, again, might seem to indicate an underlying consequentialist bias at the normative ethical level. However, I see no reason to think that helping others is a necessary condition for moral sainthood. Presumably one could be a moral saint by blowing the whistle on corporate corruption, at great risk to oneself, even if none of the victims of the corruption would directly benefit from your having blown the whistle, and even if it's not clear that you would be preventing anyone from becoming a victim in the future. Cases like this—where one's

[2] Indeed, one could presumably also achieve sainthood through deeds that affect only non-human animals.

moral accomplishment might be less tangible—are extremely interesting, but I have set them aside in favor of what I take to be less controversial cases.

Let me then give an overview of what I take a moral saint to be, with a brief preview of some of the case studies I shall use as examples in later chapters. Rather than giving strict necessary and sufficient conditions for sainthood, I aim to delineate features of a paradigm case. Roughly, on my view moral saints are people with more-or-less ordinary psychology who have devoted their lives to a moral project, and who consistently perform actions that are: (1) good but not required; (2) morally significant[3]; (3) undertaken at some personal cost; and (4) not outweighed by other morally bad or blameworthy actions. The two case studies to which I devote the most attention are Paul Farmer and Susan Tom. Farmer is a doctor who founded an organization that treats the poorest, sickest people across the world. I think his accomplishments—which I discuss in more detail in the chapters that follow—make him an uncontroversial example of a modern-day moral saint. Farmer is a good example not only because of the magnitude of his moral contribution, because also we have a richly detailed picture of his everyday life, his personality, and his motivations, in the form of the excellent book about him, *Mountains Beyond Mountains* by Tracy Kidder (2003).

Susan Tom is a single mother in California who has adopted over a dozen ill and disabled children. Her house is a vibrant, joy-filled place, an oasis for children that no one else wants. The daily moral contributions in Tom's life are in some ways more

[3] This constraint is necessary because some supererogatory (i.e., good but not required) actions are not morally significant. Michael Stocker (1968) offers a list of supererogatory actions that includes "buying an ice cream cone, on a hot day, for a child one does not know (56). With Stocker, I shall assume that not all supererogatory actions are saintly, and that there is more to being a moral saint than just many supererogatory actions strung together.

personal than Farmer's, because the people she is helping are her immediate family members. But her deeds are no less extraordinary. As with Farmer, Tom makes a particularly good example because an entire year of her family's life was chronicled in the excellent documentary film "My Flesh and Blood" (2003).

In the chapters that follow, I make use of examples like Farmer and Tom in order to make my account both more vivid and more psychologically realistic. In Chapter 2, "Sainthood and Character Traits," I appeal to Paul Farmer as a sort of counter-example to Susan Wolf's picture of moral saints as unattractive and annoying. In Chapter 3, "*De Dicto* Desires and Moral Fetishism," I contrast real moral saints like Paul Farmer and Susan Tom with "moral fetishists" or "moral imposters"—those who are motivated in the wrong way. In Chapter 4, "Sacrifice and Moral Obligation," I argue for an objective understanding of sacrifice, and along the way I look more closely at the nature of the sacrifices made by people like Farmer and Tom. And finally, in Chapter 5, "The Ratcheting-Up Effect," I argue that knowledge of people like Farmer and Tom can have the effect of increasing or *ratcheting-up* the level of moral obligation faced by the rest of us. This is perhaps the most significant and controversial claim argued for in the dissertation. Importantly, it is not an empirical claim; I am not arguing that learning about real moral saints will change what everyone else *does*. Nor I am arguing that teaching people about moral saints would be the most efficient way to change their behavior for the better. Rather, I argue that learning about moral saints can change what we can *legitimately demand* of people, thus changing their moral obligations. Of course, since reading this dissertation amounts to learning about the lives of moral saints, I should

warn readers: if my argument succeeds, then simply reading this document could potentially increase what morality demands of you!

CHAPTER 2

SAINTHOOD AND CHARACTER TRAITS

The main task of this chapter is to argue that moral sainthood need not constrain one's character traits or personality in the ways that Susan Wolf (1982) suggests. There may be some character traits—honesty, compassion, etc.—that are in some sense "essentially moral," that bear on or contribute to our moral assessment of an agent. But Wolf includes some less obviously moral traits in her account of moral sainthood. For instance, she thinks patience is mandatory in a moral saint, while a dark sense of humor is expressly forbidden. By building these and other requirements into moral sainthood, Wolf virtually guarantees that her conclusion—that moral saints are unattractive and unfit as personal ideals—will follow. Yet I shall argue that there is no good reason to build these requirements into sainthood in the first place.[4]

On Wolf's account, a saint is someone we want neither to *be* nor *be around*, because the saint cannot devote himself to hobbies, cannot tell certain jokes or laugh at them, cannot be pessimistic, and must generally be so engrossed in his moral mission as to be almost irritating. I respond to these claims in several steps. After giving a brief summary of Wolf's claims about character traits, I present the case of Paul Farmer as a potential counter-example to Wolf's argument. With the case of Paul Farmer in mind, I then challenge Wolf's claims about both the character traits and the activities of a moral

[4] Much of this Chapter—notably the first four sections—appear also in Carbonell (2009).

saint. Next I argue that if Wolf's account of these traits and activities were correct, then moral saints would be self-defeating. Finally, I examine whether the findings of the "situationism" research in social psychology bear on how we view the relationship between moral saints and character traits.

Wolf on Moral Saints

Wolf equates moral saintliness with "moral perfection" and thus defines the moral saint as "a person whose every action is as good as possible, a person, that is, who is as morally worthy as can be" (419). She further claims that our common sense, pretheoretical notion of moral sainthood necessarily includes that "one's life be dominated by a commitment to improving the welfare of others or of society as a whole" (420). This commitment can be discharged in two ways. Wolf's "Loving Saint" helps others because his happiness "would truly lie in the happiness of others, and so he would devote himself to others gladly, with a whole and open heart" (420). On the other hand, the "Rational Saint" helps others out of duty; he "sacrifices his own interests to the interests of others, and feels the sacrifice as such" (420).

These two conceptions of sainthood align quite nicely with our folk notions of saints as being either unusually compassionate or unusually dutiful. Indeed, Wolf's "Loving Saint" and "Rational Saint" mirror the two options first proposed in J.O. Urmson's seminal paper "Saints and Heroes" (1958). According to Urmson, there are two ways to commit a saintly or heroic action: "without effort" (like Wolf's "Loving Saint"), or through "self-control" in the face of countervailing self-interest (like Wolf's "Rational Saint") (Urmson, 1958, 201). Of course, a Loving Saint must also sacrifice personal

interests, but perhaps does not feel such sacrifices *as* sacrifices in quite the way that the Rational Saint does.

While Wolf acknowledges that the Rational Saint and the Loving Saint present two quite different pictures of motivation, she thinks their "public personalities" would be similar (421). Indeed, the bulk of Wolf's argument for the claim that moral saints are horribly unattractive proceeds without reference to the distinction between Rational Saints and Loving Saints. Moral saints may vary a great deal in certain cosmetic details, but their core character traits will be constrained by sainthood. It is only these essentially saintly traits that Wolf finds problematic. Joviality, garrulousness, and athleticism, for example, don't matter (421). What does matter, however, is that the saint "will have the standard moral virtues to a nonstandard degree" (421). As such,

> He will be patient, considerate, even-tempered, hospitable, charitable in thought as well as in deed. He will be very reluctant to make negative judgments of other people. He will be careful not to favor some people over others on the basis of properties they could not help but have (421).

These traits may seem uncontroversially saintly. As I shall argue in a later section, however, these traits are problematic if we interpret them in such a way as to make sense of Wolf's ultimate claim that the saint is unattractive.

Wolf makes a helpful distinction between *practical* obstacles to moral sainthood and *logical* obstacles. Having nonmoral interests or hobbies is merely a practical obstacle, because these hobbies would eat up time that would otherwise be spent benefiting others. So hobbies like "reading Victorian novels, playing the oboe, or improving [one's] backhand," which might seem (at least to Wolf) to play an essential role in a "life well lived," are in most cases prohibited for the moral saint, but only for

practical reasons, such as lack of time (421). If the moral saint could maximally benefit others *and* have hobbies, this would be unproblematic. In reality, though, Wolf thinks the moral saint will only have non-moral interests when doing so allows him to further his moral project, as when a "a good golf game is just what is needed to secure that big donation to Oxfam" (425). The moral saint cannot pursue the golf game for its own sake; the fact that getting to play golf sometimes goes along with saving the world is a mere "happy accident" (425).

There are other sorts of traits, though, that Wolf thinks present *logical* obstacles to sainthood. These traits are in "more substantial tension" with being a moral saint (421). For example, Wolf argues that certain sorts of humor would be off limits for the saint because they go "against the moral grain" (422).

> For example, a cynical or sarcastic wit, or a sense of humor that appreciates this kind of wit in others, requires that one take an attitude of resignation and pessimism toward the flaws and vices to be found in the world. A moral saint, on the other hand, has reason to take an attitude in opposition to this—he should try to look for the best in people, give them the benefit of the doubt as long as possible, try to improve regrettable situations as long as there is any hope of success. This suggests that, although a moral saint might well enjoy a good episode of *Father Knows Best*, he may not in good conscience be able to laugh at a Marx Brothers movie or enjoy a play by George Bernard Shaw (422).

These remarks about humor echo Wolf's earlier claims about other character traits: just as the moral saint must display positive traits like patience and charity, he must also have an overwhelmingly *positive* sense of humor. He must not only favor lighthearted humor but in fact *resist* and *oppose* dark humor. In the sections that follow I argue that this is simply not the case. Dark humor is not only permissible in a moral saint, but in some cases desirable. And even if certain kinds of humor were an obstacle to sainthood, it would be a practical obstacle, not the stronger "logical" obstacle Wolf describes.

After making these claims about the kinds of traits, activities, and humor that conflict with a saintly disposition, Wolf argues that the moral saint in no way resembles what we might call the "perfectly cool" person. "A moral saint," she writes, "will have to be very, very nice. It is important that he not be offensive. The worry is that, as a result, he will have to be dull-witted or humorless or bland" (422). She then argues that the moral saint *is* dull-witted, humorless and bland, because he does not embody the sort of nonmoral ideals we admire in "athletes, scholars, artists—more frivolously, […] cowboys, private eyes, and rock stars" (422). The moral saint cannot have "Katherine Hepburn's grace" or "Paul Newman's 'cool'" (422). These traits, Wolf argues, "cannot be superimposed upon the ideal of a moral saint" (422).

What Wolf is trying to show is that the "ideal moral agent" is not ideal insofar as he is not someone we would necessarily want to *be*. Wolf claims that "a person may be *perfectly wonderful* without being *perfectly moral*" (436). Here she is introducing what she calls the "point of view of individual perfection" (437). This point of view is not exactly moral, not exactly egoistic, but certainly contains elements of both of those perspectives, as well as perhaps an aesthetic component. She describes this point of view as follows.

> Like moral judgments, judgments about what it would be good for a person to be are made from a point of view outside the limits set by the values, interests, and desires that the person might actually have. And, like moral judgments, these judgments claim for themselves a kind of objectivity or a grounding in a perspective which any rational and perceptive being can take up. Unlike moral judgments, however, the good with which these judgments are concerned is not the good of anyone or any group other than the individual himself (436).

The problem, of course, is that we now have competing normative standards for evaluating lives. From the moral perspective, one ought to desire to be a moral saint, and

from the perspective of individual perfection, one ought to desire to be well-rounded in the ways Wolf describes.

It seems that Wolf's argument is made up of two claims: first, that moral saints are unattractive from the point of view of personal perfection, and second, that their unattractiveness from this point of view is *problematic*, because it means that the morally best life is not the best life all told. My main task is to challenge the first claim, by showing that moral saints can have all the traits that make for an attractive and well-rounded life. In challenging the first claim, though, I will implicitly challenge the second claim. After all, if saints are not unattractive, then we don't have to worry about their unattractiveness being *problematic*. Nonetheless, my argument leaves untouched much of what is philosophically interesting about the interplay between Wolf's two "points of view." While I challenge the claim that the personal point of view rules out certain morally extraordinary lives, I am *not* challenging the claim that certain personally interesting lives (e.g., that of the great violinist or single-minded athlete) might be ruled out from the moral point of view. I take it these are two different, though related, problems.

Paul Farmer: A Counter-Example to Wolf

Some commentators have argued that Wolf's conception of moral sainthood gets it wrong by favoring moral perfectionism over a relationship with the divine, or by setting the bar for sainthood far too high.[5] I shall claim, however, that we can challenge

[5] Robert Adams (1984) has argued that *real* saints (that is, Saint Francis, Mother Teresa, etc.) are not bland, and thus a conception of moral saints according to which they are bland cannot be correct. Moreover, he takes issue with what I shall call Wolf's "perfectionist rationality." He disagrees that perfection in moral value "depends on the maximization of that type of value in every single action of the person" (393). This

Wolf's thesis about the attractiveness of moral saints without departing from her underlying conception of a moral saint as a *truly extraordinary moral agent,* an agent whose uncommon qualities and achievements are not essentially religious.

Consider Dr. Paul Farmer, a man of extraordinary moral achievement who serves as a counter-example to some of Wolf's claims about moral saints. I take Farmer to be an uncontroversial example of a real-life moral saint, and yet he looks almost nothing like the person Wolf describes. Her saint is irritating, obsessive, and bland; Farmer is charismatic and funny. Her saint is holier-than-thou and no fun to be around; Farmer attracts friends and followers like a magnet. Wolf acknowledges that there are a "variety of types of person that might be thought to satisfy [the conditions for moral sainthood]," but claims that "none of these types serve as unequivocally compelling personal ideals"

---

maximization "lies behind much that is unattractive in Wolf's picture of moral sainthood; but I believe it is a fundamental error" (393).

Adams thinks the solution is to return sainthood to its religious roots. Real saints are not single-minded, he argues. Rather, "they commonly have time for things that do not *have* to be done, because their vision is not of needs that exceed any possible means of satisfying them, but of a divine goodness that is more than adequate to every need" (396). For Adams, saints are not moral perfectionists, but rather "people in whom the holy or divine can be seen" (398). But the sort of saint Adams describes is not necessarily a *moral* saint. While various historical examples of religious saints might turn out to be moral saints as well, moral sainthood itself is something we can describe without reference to the holy or the divine.

Whereas Adams responds to Wolf by appealing to a religious conception of sainthood, Edward Lawry goes in a different direction in his article "In Praise of Moral Saints" (2002). He argues for a much more lenient, inclusive conception of sainthood. "I have a sense," Lawry writes, " that [Wolf] is not talking about *real* moral saints" (1). In order to accommodate the several people he has known personally and believed to be moral saints, Lawry develops a theory according to which moral value is a "coming together of a life in integrity" (1). Displaying a high degree of integrity is what characterizes "good human beings, or, moral saints" (1). Notice that Lawry's notion of sainthood is so weak that "good human being" and "moral saint" seem to be taken as synonyms. This strikes me as mistaken. We all know some people who seem to display much more integrity than the rest of us. Nonetheless there seems to be a notion of truly *extraordinary* moral achievement that is worth exploring.

(419). Farmer, I shall claim, satisfies the conditions for moral sainthood as well as perhaps any living person can, and yet also serves as an "unequivocally compelling personal ideal." If there is any doubt that he is sufficiently compelling, whatever minor tweaks he would need are things that could be changed without sacrificing the quality of his moral achievements.

Paul Farmer is a doctor and medical anthropologist at Harvard Medical School. His non-profit organization, Partners in Health, runs clinics that treat the world's poorest, sickest patients. Farmer treats thousands of these patients himself, and he is world-renowned as an advocate for the poor and an expert on tuberculosis. In *Mountains Beyond Mountains* (2003), Tracy Kidder's celebrated book about Farmer, we are treated to a magnificently rich case study of a contemporary moral saint. We learn that Farmer grew up in an eccentric and unprivileged family. For much of his childhood the family of eight lived in an old bus parked in a trailer park, and later on a fishing boat moored in the shallows of Florida's Gulf Coast (Kidder, 2003, 47-54). Despite his odd upbringing, Farmer's stellar intellect drove him to Duke and ultimately Harvard, where he excelled and earned both an M.D. and PhD in anthropology, despite missing most of his classes to be in Haiti, working at the rural health clinic he built from the ground up (84). That clinic was the beginning of Partners in Health, which now oversees public health projects all over the world.

What is most interesting about Paul Farmer is not what he has accomplished but how he has accomplished it. Although he is almost maniacally driven by morally good pursuits, he does not think of these pursuits under the description "morally good." He simply wants to help the poor and the sick, and he does so not with the angelic purity

Wolf imagines, but rather with an acerbic wit and a willingness to do what is necessary to further his cause: curse the inaction of others, pay bribes to soldiers at checkpoints, and accommodate the dangerous mythologies of his patients. Farmer's life is, to be sure, marked by asceticism: he takes no salary (23); he sleeps no more than four hours a night (23); he lives alternately in a hut and in the basement of his office (151); he does not buy new clothes (255); he has little time to himself; he hikes for hours and hours to make housecalls (36-41). Nevertheless, we don't find his life unattractive on account of this asceticism; on the contrary, we admire him partly *because* of it.[6]

While Farmer may appear to have the drive of a perfectionist, he doesn't suffer from the sort of obsessive maximizing that makes Wolf's saint so unattractive. He certainly wants to help as many people as he can, and often lobbies for the sort of efficient measures that make this possible, like lowering the prices of drugs and following rational protocols for treating infectious disease. But he also works with people face-to-face and thus finds himself compelled to make gestures that are more heartfelt than efficient. He sends a Haitian boy to Boston for surgery at great expense (262-279). He spends hundreds of dollars to replenish a malnourished man with vitamin shakes when the money could have been spent otherwise (25). He signs an entire paycheck over to a patient who is facing eviction (95). He buys a six-pack of beer for a homeless, alcoholic patient, wraps it in wrapping paper and delivers it on Christmas Day (16).

[6] Farmer is not ascetic on *principle*. Though he does not seek them out, he seems to enjoy the finer things in life on occasion, as when Kidder takes him to a restaurant with fine wine (Kidder, 2003, 7). He also enjoys reading People Magazine in airports. He claims he reads it in order to stay in touch with his patients, but one gets the sense he finds some pleasure in getting lost in thoroughly unserious matters once in a while. We also learn that he likes action-adventure movies.

Farmer doesn't cultivate hobbies or personal interests to anywhere near the extent that Wolf seems to find necessary for a well-rounded life, but his life is far from "barren." Indeed, you might say that his work is so consuming that it creates its own hobbies: travel, foreign languages, and the study of religion and mythology in different cultures. While he denies himself most of the creature-comforts available to someone in a rich country, his self-denial is endearing; we can see that he gets satisfaction, and often great pleasure, from the kind of work his self-denial makes possible. He packs only three shirts for a two-week trip, but in so doing he frees up space in his luggage to act as a courier for all manner of objects that his patients ask him to deliver to family in the States, a task that surely brings him great joy (190-192). He's efficient without being robotic. "Traveler's tip number one thousand seventy-three," he tells Kidder, "If you don't have time to eat, and there's no other food on the plane, a package of peanuts and Bloody Mary mix are six hundred calories" (191). As unappetizing as this meal sounds, it doesn't make me "glad I'm not a moral saint," as Wolf might suggest. Rather, it makes me wish I were so motivated to help poor people that I were willing to subsist on airplane snacks, even for a day.

In fact, we ought to go further than simply to say that Farmer's life is *not barren*. On the contrary, he *flourishes*. What could be more interesting, more fulfilling, more deeply satisfying than a life devoted to using one's talent and intellect to improve the lives of thousands of people, indeed to prevent people from *dying*, and to do so in places where no one else is prepared to help them but you? Farmer's life is, of course, marked by great sacrifice, and we may wish, for his sake, that he had to make fewer sacrifices. But a self-sacrificial life can still be a good life on the whole. Indeed, we might even

think Farmer's lifestyle—in which self-sacrifice makes possible staggering moral accomplishments and countless meaningful human interactions—yields a *net benefit* of wellbeing. This doesn't mean that Farmer is blissful and content. Rather, he is chronically unsatisfied, and that is what keeps him going.

Thus far I have argued that the life of a moral saint does not have all the costs Wolf claims it does, and furthermore that the life of a moral saint is attractive *despite* its costs. In other words, I have tried to show that it would not be so bad to *be* Paul Farmer. But Wolf claims that we prefer not only not to *be* a moral saint, but not even to be *around* a moral saint. Farmer is a counter-example to this claim as well.[7] While Wolf's moral saint is annoyingly obsessed with "morality" so described, Farmer is obsessed with the content of his commitments. His obsession is more comical than unattractive. For example, he speaks in his own shorthand idiolect of acronyms and catchphrases. He wants to find an "O for the P" ("preferential option for the poor") and he often exercises an "H of G" ("hermeneutic of generosity") (174, 217). He gets annoyed when someone commits a "seven-three" ("to use seven words where three would do") or a "ninety-nine one hundred" ("quitting on a nearly completed job") (217). Yet his hyper-awareness of the endless task of healing the world's sick doesn't render him fanatical. Rather, it displays that he is vividly acquainted with a fact that escapes most people: that the desire for a clean shirt or an extra hour of sleep pales in comparison to the needs of the world's sickest and poorest people. When Kidder remarks on Farmer's insane schedule, Farmer

[7] There are several instances in Kidder's book where Kidder himself, and people he interviews, express frustration at the fact that being around Farmer often makes one feel guilty about one's own shortcomings as a moral agent. Nevertheless, it is clear that, on balance, Farmer is someone others want to be around. Indeed, he seems to fare better on this criterion than the vast majority of people; not only do others not mind being around him, but they seem to seek out time with him.

responds, “The problem is, if I don’t work this hard, someone will die who doesn’t have to. That sounds megalomaniacal. I wouldn’t have said that to you before I’d taken you to Haiti and you had seen that it was manifestly true” (191). Clearly, Farmer is obsessed with his work—and how could he not be, given that lives depend on it? Nevertheless, his obsesses about the *object* of his concern—the poor, the sick—and not the moral goodness of being so concerned. As I will argue in Chapter 3, this distinction is significant.

Farmer displays precisely the cynical and sarcastic sense of humor that Wolf thinks the moral saint is not entitled to. When a Haitian patient tries to pay him with “milk in a green bottle with a corncob stopper,” Farmer thanks her profusely in Creole and then turns to Kidder and says, “Unpasteurized cow’s milk in a dirty bottle. I can’t wait to drink it” (26). He regularly refers to his poor patients as “the shafted” and refuses to charitably accommodate the politically correct notion that all suffering is equal, believing instead that there are important differences in the “degree of hose-edness” of various groups (216). When Kidder asks him who is paying for his trip to Cuba, Farmer answers, “Capitalists, commies, and Jesus Christers” [the Soros Foundation, the Cuban government, and a church group] (184).

Farmer is also a counter-example to Wolf’s claim that moral saints are not pessimistic. Before giving two speeches in Cuba, he tells Kidder, “One speech is for clinicians, how to deal with HIV and TB coinfections. The other is why life sucks” (198). This pessimism doesn’t seem to make Farmer any less saintly. In fact, if anything it seems to make him *more* saintly, because it shows that he has a certain hardened realism in the face of grave challenges, rather than a cheery naiveté. Yet his pessimism doesn’t degenerate into *resignation*. Instead of giving up in the face of immensely difficult tasks,

he simply expects others to rise to the challenge. For instance, when other public health experts deem certain patients too difficult to treat, Farmer simply ignores the experts and finds a way. He takes day-long hikes to treat isolated patients in their huts, bringing small items they have asked him for, even when the link between these items and treating the disease is tenuous. "We can spend sixty-eight thousand dollars per TB patient in New York City," Farmer says, "but if you start giving watches or radios to patients here, suddenly the international health community jumps on you for creating *nonsustainable* projects. If a patient says, I really need a Bible or nail clippers, well, for God's sake!" (42).

As this brief glimpse has shown, if Paul Farmer is a moral saint, then he causes quite a few problems for Wolf's account. He is obsessed but not fanatical, ascetic but not self-righteous. He is sarcastic and cynically witty without being resigned. He is funny and fun, and no less morally admirable for it. He thinks unconventionally, in a way that seems only possible when a quirky, imperfect human mind is unleashed on a complex problem. What makes him so interesting is that he is a distinctly *human* moral saint, not a humorless robot. He proves that someone who exhibits all of the important features of a moral saint *can* be the sort of person we want to be.

The Traits of Sainthood

With the case of Paul Farmer in mind, we can now examine Wolf's account of the traits and activities of sainthood in more detail. Recall that she claims the saint must be "patient, considerate, even-tempered, hospitable, charitable in thought as well as in deed" and also "very reluctant to make negative judgments of other people" (421). Some of

these traits, like even-temperedness, strike me as uncontroversially conducive to benefiting others. Others may be problematic. Patience, for example, seems virtuous only when it is warranted. In my view, the moral saint should have no patience for the person preaching hatred on the street corner, unless patience could eventually conduce to changing his mind. Nor should the saint have patience for corruption or incompetence in government if outrage would better conduce to ending it. In fact, public displays of impatience might be *obligatory* in many cases. We would expect the moral saint to display impatience when fear or self-interest would cause most people not to. Paul Farmer, for example, seems to be an incredibly impatient person. (Of course, being impatient does not require being rude or belligerent; one could express one's impatience in a polite and patient manner, and perhaps we should expect a moral saint to do this.[8])

Similarly, it seems that the moral saint should be "charitable in thought" only when charity is warranted. To be sure, charity of thought is often helpful in guarding against premature dismissal of others' views or premature conclusions about their motives. But *automatically* or *universally* interpreting the words or deeds of others in the most favorable light seems downright naïve and certainly inimical to the project of benefiting others. The same goes for being "very reluctant to make negative judgments of other people." If Wolf means this as a *general* virtue, then perhaps she is failing to recognize that making negative judgments *when they are warranted* is an *essential* component in the project of benefiting others. Of course, there can be reasons not to display an attitude, even if it is warranted or fitting. For example, you might have a *moral*

[8] See Buss (1999) for more on the moral significance of manners. Buss argues that to be "impolite," "rude," "inconsiderate," etc., is to flout one's moral duty to treat others with respect.

reason not be angry at the person preaching hatred on the street corner, if your anger might provoke him to become violent against innocent bystanders. But this is merely a reason not to *display* your anger, not a reason to refrain from *feeling* it, as Wolf's view seems to demand.

We might charitably assume that Wolf means for all of the above exceptions to be built into her notions of patience, charity, and the like. Perhaps she means, not that a saint should *always* and *automatically* be patient and charitable, but rather that a saint would be particularly good at discerning when patience and charity are *called for*. Farmer, for example, surely exercises more patience when dealing one-on-one with a sick person than he does when navigating the bureaucracy that determines how quickly that person can get a needed drug or treatment. Perhaps this is all that Wolf means: the saint should be patient and considerate when it comes to legitimate needs (say, a sick person's need to understand important medical instructions and not be condescended to), but need not display these traits in response to illegitimate demands (say, the demands of the bureaucracy to receive multiply redundant paperwork). Yet this interpretation, according to which the moral saint displays these positive traits only when they are warranted, is simply inconsistent with the conclusions Wolf draws about her patient and charitable saint. For she considers these traits to be *so* saintly that they make the saint "too good" (421). Here we get the first taste of Wolf's ultimate thesis: that these saintly traits "are apt to crowd out the nonmoral virtues, as well as many of the interests and personal characteristics that we generally think contribute to a healthy, well-rounded, richly developed character" (421). But Wolf has not yet provided any good reasons to think that the moral saint *can't* be well-rounded. A well-honed ability to know when positive traits

are warranted and to display them accordingly doesn't seem to make a person *too good*, nor does it seem to be at odds with well-roundedness.

Perhaps Wolf is simply calling for a *reluctance* to be negative, as a way of compensating for the human tendency to be *too* negative. But this reading also fails to support Wolf's conclusion. For surely a reluctance to be negative—surely any trait that is fostered as a way of moderating tendencies toward extremism—is not the sort of thing that would prevent someone from being well-rounded. A healthy, compensating dose of reluctance would not make someone "too good," where "too good" refers to the notion that too much of a good trait can be a bad thing. So it seems that Wolf faces a dilemma. Either the saint is so positive and charitable that it interferes with his well-being and renders him irritating, or he is only moderately positive and charitable (that is, he displays these attitudes mainly when they are warranted). If the first is true, then Wolf's conception is too extreme, since knee-jerk positivity and unrestrained charity of thought do not seem to be requirements of sainthood, and indeed may be in tension with sainthood. Yet if the second horn of the dilemma is true—and I think it is—then the saint no longer comes across as unattractive, and Wolf's thesis suffers.

What is true of patience and charity is true also of most of the other character traits Wolf attributes to moral saints. Take, for example, the qualities of "looking for the best in people" and giving others "the benefit of the doubt." Either the moral saint *limits* how much he gives others the benefit of the doubt, in which case he does not appear unattractive, or he rampantly and indiscriminately gives others the benefit of the doubt, in which case his blind optimism is likely to undermine his ability to effectively pursue morally good projects.

It might be objected that I have interpreted Wolf's list of character traits too harshly. According to this objection, saints don't "look for the best in people" merely because doing so is conducive to performing good deeds. Rather, saints "look for the best in people" (and are patient, charitable, etc.) because to do so is the mark of a truly virtuous person. That is, there are certain traits that *essentially moral*, just as there are others that, as Wolf puts it, "go against the moral grain" (422). On this view, there are certain traits a moral saint must have even if it means, as I have argued above, that she is not accurately responding to the features of the world. In fact, *not* accurately responding to the features of the world might be an *essential* component of these virtues.

Julia Driver (1989) argues that some moral virtues *do* involve this sort of blindness to the facts of a situation. These virtues—modesty, blind charity, and the refusal to hold a grudge—"involve ignorance in an essential way" (Driver, 1989, 374). Modesty, for example, involves ignorance of one's worth. Driver claims that it would be "counterintuitive" to suggest that these traits are not virtues (384). "We value the virtues of ignorance" she claims, "because of the psychological states that underpin them, but we value these psychological states for instrumental reasons" (383). So, for example, we value modesty because it involves an underlying psychological state ("reluctance to take in one's own accomplishments fully" (383)) that tends to improve social interaction by reducing jealousy and envy. To be modest is to be "less troublesome" (384).

We ought to examine Driver's analysis of the virtue of "blind charity" in a bit more detail. If Driver's argument succeeds, it lends credence to Wolf's claim that a moral saint must "look for the best in people" and be "charitable in thought" and "very reluctant

to make negative judgments of other people" to such an extent that she becomes unattractive as a result (Wolf, 1982, 421-422). According to Driver,

> A person who is blind in charity with others is a person who sees the good in them, but does not see the bad. Blind charity differs from charity in that it is usually the case that, when one is merely charitable toward another, one favors that person in some respect, *in spite of* perceived defects or lack of desert. For example, in employing the principle of charity, one interprets a person's views in the best possible way, even though one perceives certain possible defects. Blind charity is a disposition not to see the defects, and to focus on the virtues of persons (381).

It is not immediately clear just why blind charity is meant to be a *good* thing, and even more difficult to understand how it could be a *morally* good thing. As Driver points out, blind charity requires a kind of ignorance: the inability to see the bad in people. This sort of charity "cannot be reflective," she says (382). On the face of it, unreflectiveness does not seem like the mark of a virtue. Of course, if we suppose that most people are mostly good, and that their defects are not usually important, then perhaps blind charity is useful in most cases. The blindly charitable person may have more pleasant, and more productive, interactions with most people; she won't perceive their flaws, and so she won't be hung up on them. Yet our appreciation or admiration for blind charity in such cases seems almost aesthetic rather than moral. We find the person who is blind in charity to be endearing, innocent, and pure; we admire such a person for her "sweetness," as Driver says of the character Jane Bennett in *Pride and Prejudice*. But what happens when the blindly charitable person interacts with people whose bad qualities are relevant, significant, and not to be ignored? It seems that blind charity could cause one to trust the untrustworthy, to rely on the unreliable, or worse, to ally oneself with evil. In such cases, blind charity seems indistinguishable from *naiveté*. Naiveté is only rarely an admirable

quality in an adult, and so it would be odd if it were supposed to count as a moral virtue. The naiveté of blind charity might enhance a large number of relatively insignificant social interactions, but this would be outweighed by the fact that it could spoil a small number of quite significant interactions, potentially causing a great deal of harm.

Wolf's version of charity might not be as "blind" as Driver's blind charity. Where Driver's charitable agent has a true perceptive *defect*—a *blind spot*, as it were—Wolf's moral saint may simply have a tendency to accentuate the positive and minimize the negative. But Wolf's moral saint faces the same problem whether she is completely blinded to the flaws of others or only partly blinded. As I argued earlier, Wolf's claim that positive character traits like charity make a moral saint unattractive leads us to a dilemma: either the saint is charitable in an *undiscriminating* way (roughly, blind charity), in which case she will be much less effective, and much less admirable, as a moral agent, or she is only charitable when charity is *called for*, in which case she is not unattractive on account of this trait. The first horn of the dilemma is only a problem if blind charity *does* cause a person to be less effective and less admirable as a moral agent. Driver has suggested the opposite: that virtues of ignorance, like blind charity, can ease social interaction, and that such traits are so admirable that it would be "counterintuitive" to suggest they are not virtues. But as I have argued, the considerations that count in favor of blind charity don't seem to be *moral* considerations, and in many contexts blind charity would in fact *undermine* moral goals. In any case, the second horn of the dilemma represents the more plausible account of moral saints. Moral saints should indeed be charitable, but not blindly charitable. Paul Farmer calls his own principle of charity his "hermeneutic of generosity" ("H of G"). He exercises it toward Kidder when he says, "I

know you're a good guy. Therefore I will interpret what you say and do in a favorable light" (214). Nevertheless, Farmer's "H of G" is not a type of *blind* charity. He does not lack the ability to see the bad in people. In fact, Farmer is a particularly astute moral critic. This allows him to shame others into action when leading by example does not work on its own.

Wolf's argument about character traits extends to humor. She claims that a cynical and sarcastic wit requires "resignation and pessimism toward the flaws and vices to be found in the world," and that a moral saint could not have these attitudes (422). It seems reasonable enough that cynical wit involves pessimism, but why can't the moral saint be pessimistic? Insofar as pessimism is simply a belief about how things tend to work out, it could actually be to the moral saint's *advantage* to be pessimistic.[9] Indeed, Paul Farmer is surely *more* saintly on account of his pessimism. He is motivated to help the poor precisely because, as he says, "life sucks," and his patients are "the shafted." He is pessimistic about how his patients will fare without his help, but this allows him to do *more* good.

We might also call into question Wolf's claim that cynical wit requires an attitude of "resignation." Is an attitude so extreme as resignation really necessary for cynical or sarcastic humor? Resignation seems to connote hopelessness or futility. But if cynical and sarcastic humor were built on hopelessness and futility, then cynical or sarcastic

[9] Now, perhaps we ought to clarify what sort of attitude pessimism is. If pessimism is just having the negative attitudes and beliefs that are warranted by the evidence, then this certainly poses no problem for the moral saint, unless it brings her down and thus makes her less productive. However, on this conception, pessimism is really just another world for "realism." Suppose pessimism is instead the tendency to have negative attitudes and beliefs *regardless* of the evidence. *This*, I concede, might be (practically, not logically) problematic for the moral saint. But only the first, weaker sort of pessimism seems necessary for a cynical wit.

commentary about bad situations would come out sounding sad or mean rather than funny. That is, if in referring to his patients as "the shafted" Farmer were *resigned* to the notion that their plight was permanent, or hopeless, we would consider his use of the term to be mean-spirited. But he's not resigned. Quite the opposite: he fights to treat his sickest patients when the world's public health authorities are resigned to letting them die.

Character Traits and Self-Defeat

By appealing to Paul Farmer as a case study, I have tried to suggest that Wolf's account of moral sainthood places unnecessary restrictions on the sorts of traits and attitudes a moral saint can have. Lurking beneath my analysis, however, is an argument for an even stronger claim: that if we *granted* all of these restrictions, Wolf's moral saint would be self-defeating. Here I want to make that stronger argument more explicit. Following Wolf's distinction between "logical" and "practical" obstacles to sainthood, I will argue that her moral saint is both logically and practically self-defeating. We'll call a saint "logically self-defeating" if two or more necessary components of sainthood cannot consistently be instantiated in the same person. A saint is "practically self-defeating" if, as a contingent matter, it often happens to be the case that the real-world exercise of one component of sainthood gets in the way of the exercise of another component thereof.

Wolf's moral saint is logically self-defeating insofar as this saint must display a vast repertoire of positive, optimistic traits, some of which conflict with each other. Wolf describes the moral saint in the following way: she is patient, considerate, even-tempered, hospitable, charitable (in deed and in thought), reluctant to make negative judgments, always looking for the best in people, very nice, not offensive, not pessimistic, not

resigned, not cynical, humorless, dull-witted, and bland. Wolf doesn't mention that moral saints must be *sincere*. But surely sincerity is a virtue that belongs on the above list, and perhaps on any plausible list of the virtues of a moral saint. After all, if a moral saint cannot be pessimistic or sarcastic, then she certainly cannot be insincere—surely insincerity goes "against the moral grain." But the virtue of sincerity is in tension with the other traits Wolf attributes to moral saints.

The problem, it seems, is that it would be quite difficult to display all of Wolf's positive traits without being insincere. Now, one might reply that I have not sufficiently idealized my moral saint; the *average* person, who is not perfectly positive and optimistic, could not display all of these positive traits without being insincere. But the moral saint really *would* be optimistic, patient, charitable, nice, and so on, and thus would not have to be insincere to display such traits. But this would dissolve the inconsistency *only* if the moral saint displayed the positive traits exclusively when they were *warranted*. As I suggested earlier, a sincere person can be considerate of the needs of others when those needs are deserving of consideration: when they are genuine, legitimate, consistent needs. So, for example, Paul Farmer ought to be considerate of his patients and their legitimate demands, but he ought not be considerate or patient when it comes to arbitrary and harmful bureaucratic practices. Thus, if Wolf demands that her moral saint *always* be considerate, regardless of whether this attitude is fitting, then she is also demanding insincerity. The moral saint cannot *sincerely* be considerate of illegitimate demands unless she is completely unaware of their illegitimacy. As I argued earlier about the "virtues of ignorance," it is difficult to see how being accommodating in such an unthinking, undiscriminating way can be a positive moral trait. It may be positive

in the sense that so accommodating a person would be very nice and agreeable, but it seems also negative insofar as such a person would be neither responding to the normatively relevant features of her environment nor helping to counteract the effects of harmful policies.[10]

My argument that Wolf's moral saint is logically self-defeating succeeds only if (1) sincerity is a necessary condition of sainthood, and (2) saints display positive traits even when they are unwarranted. The argument that Wolf's moral saint is *practically* self-defeating is more straightforward. Her moral saint is practically self-defeating because, in not allowing himself any of the pleasures of life that normally keep a person sane, he deprives himself of the very rejuvenation that would make extended good works possible. It is widely acknowledged that in order to maximize his pleasure, the hedonist ought not seek it, and that in order to maximize the good, the utilitarian ought not calculate the relative utility of every possible action.[11] Similarly, in order to maximize moral goodness, it is best that one not always aim at doing the most morally virtuous thing possible, if for no other reason than that the *deliberation* that must go into determining which action is as morally good as possible *itself* takes up time that could better be spent *doing* morally good actions. Wolf's moral saint, who exhibits a brute maximizing perfectionism, is thus self-undermining. After all, her saint is someone "whose every action is as morally good as possible" (419), and as I shall argue in Chapter 3, she expects the saint to be driven by this fact *de dicto*.

---

[10] I presume Edward Lawry (2002) is referring to this same phenomenon when he writes "But the inoffensive niceness of a person seems surely to be an objectionable moral trait when righteous indignation is called for. It seems that even trying to characterize the moral saint in this way is a self-defeating enterprise" (Lawry, 2002, 4).

[11] For a particularly lucid and concise explanation of the "Paradox of Hedonism" see Railton (1984), 140-141.

Thus there are perhaps two different ways in which Wolf's moral saint is practically self-defeating. The requirement of moral perfection is self-defeating *in the short run*, since it would result, at best, in a sort of deliberative inefficiency, and at worst in what Peter Railton (1984) calls a "paralyzing regress" of deliberating about whether to deliberate (154). In the short run, the saint would miss opportunities to do the morally right thing. The requirement that the saint "justify every activity against morally beneficial alternatives" is also self-defeating *in the long run*, because it would mean systematically deciding to forgo personal interests in favor of morally beneficial alternatives; over time, the saint would become burned-out, thus undermining her ability to do the morally right thing (Wolf, 1982, 422).

Now, it seems natural to argue that the moral saint must take time for personal hobbies and pursuits because she would need to account for the potential of becoming burned-out. Moreover, these personal hobbies and pursuits must be undertaken for their own sake—not instrumentally, as in the case of Wolf's donation-seeking golfer—otherwise they might not yield any genuine rejuvenation. Perhaps, however, it is better to argue that the saint can enjoy personal interests *because a more plausible picture of a moral saint will not require that she justify each action against morally beneficial alternatives.* For although it is true that a well-rounded life might better conduce to improving the welfare of others than an exhaustively single-minded life, it might be difficult for a moral perfectionist to alter her deliberative habits in light of this fact. For she would have little reason, in any *particular* instance, to prefer well-rounded pursuits to moral pursuits, as long as she were confident she could carry out the very next moral pursuit without dire personal consequences. In other words, preventing oneself from

becoming burned-out might require a sort of foresight and long-term thinking that is in tension with particular act-level decisions.

Peter Railton (1984) has offered a way out of these kinds of deliberative problems, which are often considered to be particularly problematic for consequentialists. He distinguishes between "subjective consequentialism" and "objective consequentialism". Subjective consequentialism demands that one aim to maximize the good in every action by using a distinctively consequentialist mode of decision-making—weighing the expected consequences of acts and carefully choosing the act that appears optimal (Railton, 1984, 152). Objective consequentialism, on the other hand, "is the view that the criterion of the rightness of an act or course of action is whether it *in fact* would most promote the good of those acts available to the agent" (152, emphasis added). It might turn out that the course of action that would *in fact* most promote the good is not the action that a subjectively consequentialist decision-making procedure recommends. So, for example, it might turn out that a moral saint ought in fact to allow himself the pleasure of personal hobbies, because it prevents him from becoming burned out and unable to do any good. This is true even if subjective consequentialism would not have recommended personal hobbies. Railton thus coins the term "sophisticated consequentialist." A sophisticated consequentialist is "someone who has a standing commitment to leading an objectively consequentialist life, but who need not set special stock in any particular form of decision making and therefore does not necessarily seek to lead a subjectively consequentialist life" (153). Sophisticated consequentialists might seem a little odd in any particular situation—they often choose actions that seem far from optimal—but they are more effective overall.

Railton's framework is useful because it allows us to think more carefully about whether Wolf's moral saint really is self-defeating. Perhaps Wolf's saint is actually a "*sophisticated* moral saint." A sophisticated moral saint can play golf even when a big donation to Oxfam does not depend on it, because playing golf keeps her sane, and staying sane means doing more good in the long-run. Recasting Wolf's moral saints as sophisticated in this way would certainly defeat my objection that these saints are self-defeating. But this would only push the problem back one step further. For sophisticated moral saints would not be *unattractive* in the ways Wolf claims, and so her argument would still be on shaky ground. After all, a large part of the saints' unattractiveness hinges on their decision-making procedure—their need to "justify every action against morally beneficial alternatives" (422). This causes them to forego personal interests, become bland, and turn into the sort of people we don't like to be around. If we replace this decision-procedure with a more sophisticated one, the unattractiveness disappears. Here we arrive at yet another dilemma: Wolf's saints are either "naïve" deliberators or "sophisticated" deliberators.[12] If naïve, they are self-defeating, and if sophisticated, they are not unattractive in the ways necessary for her argument ultimately to succeed.

In the end, it matters little whether Wolf's requirement that the saint justify all actions against morally beneficial alternatives can be shown to accommodate personal interests, or whether this requirement should be abandoned altogether, for Wolf includes the lack of personal interests as an *additional* feature of a moral saint. In other words, she seems to think that the moral saint is *essentially* dull-witted, humorless, and bland. He is

[12] As I've tried to indicate, it seems rather clear from the text that Wolf intends them to be "naïve" deliberators. So the dilemma here is really between the characterization given in the text and an alternative characterization we might offer as a charitable modification of her view. It's not a choice between two equally plausible interpretations of the text.

this way by his very constitution, not simply as a consequence of his perfectionist rationality. It is not merely that his project of benefiting others leaves him no *time* to be well-rounded, but rather that well-roundedness is itself antithetical to moral perfection. This is what Wolf means, presumably, when she says that certain pursuits are "against the moral grain" (422). Such pursuits don't run against the grain because they take up time that could be spent otherwise; they run against the grain by their nature. This is why her saint is self-defeating: he must be *both* ascetic and maximally beneficial to others, but being ascetic would likely undermine his ability to help others.

Character Traits and Morality, More Broadly

Thus far I have argued that moral sainthood does not constrain one's personality as severely as Susan Wolf claims. Indeed, *because* it does not, moral saints need not be as unattractive as she claims. And so we can be less worried about how the morally best life looks from the perspective of personal perfection. Of course, I concede that not all personality profiles or character traits are compatible with moral sainthood. There are, I think, some traits that are "essentially moral"—where by "moral" here I do not mean "morally *good*" but rather morally-laden or morally-loaded, that is, bound up with morality in some way, positively or negatively. For instance, our commonsense notion of "murder" is morally-loaded insofar as it is built in to the concept that a murder is not just any killing but an *unjustified* or *malicious* killing. In the vast majority of cases, it would be puzzling, perhaps even incoherent, to say, "She murdered him, but was it *wrong*?" while it is perfectly OK to say, "She killed him, but was it *wrong*?"

Few character traits have the morality quite so straightforwardly built-in to them as it is for "murder." It might seem that "integrity" is one. It would certainly be odd to say, "He's a man of great integrity, but how is he *morally*?" But integrity is in fact a rather vague case, and it's not clear that it really picks out an isolable character trait. On one extreme, we sometimes use "integrity" simply as shorthand for an overall moral assessment, in which case it is morally-loaded but in a trivial sense. On the other extreme, "integrity" may refer merely to a kind of consistency or transparency in one's character, in which case it could presumably be entirely non-moral—for instance, a career criminal could have integrity if he is disposed to commit his crimes in a consistent and transparent way, or if he is honest with himself about his criminality. Many senses of "integrity" may lie in the murky in-between territory. For example, we might think having integrity is a matter of consistently or self-consciously following *some* sort of moral code, even if it is the wrong moral code. In this sense, a mobster could have integrity if, say, he only kills people who have been disloyal to him.

I've discussed "integrity" here only to point out the difficulty of determining whether (and to what extent) moral considerations factor in a given character trait, indeed whether something ought even to *count* as a character trait. Even what is perhaps the most uncontroversially morally-loaded trait—honesty—is problematic when examined more closely. Surely, one would think, the disposition to be honest is a necessary (though certainly not sufficient) characteristic of the morally good person, and thus especially necessary in the morally best people. However, few would say that one ought *always and without exception* to be honest. Assuming there are circumstances in which one is morally required to lie, or at least morally permitted to do so, then what we really want in

the morally best person is not the disposition always to be honest, but rather the disposition to be honest *except when being dishonest is morally better*. Yet if our agent must know the difference between the cases where lying is and is not called for, then "honesty" begins to seem less like a trait that contributes to a person's moral character and more like a trait for which moral character is prerequisite.

Given how hard it is to pin down seemingly morally loaded traits like integrity and honesty, we should not be surprised that moral considerations have an even more tenuous connection with traits like patience, charity, optimism, and humor. I have tried to suggest that these character traits simply do not play as robust a role in determining a person's moral merit as Susan Wolf suggests. Indeed, I think our tendency to moralize these traits is a throwback to a more puritanical view of morality, one where notions like innocence and purity loom large.

Of course, the very idea that there *are* persistent and reliable character traits has been challenged by social psychologists. Philosophers have recently argued that this poses a problem for moral theory; presumably, the more our moral theory relies on notions of robust character traits, the more of a problem it is. Virtue ethics is particularly vulnerable to this critique. While Wolf's account of moral saints does not explicitly presuppose a virtue ethics framework—in fact, she focuses much of her argument on deontological and utilitarian theories—she does claim that the moral saint "will have the standard moral virtues to a nonstandard degree" (1982, 421). And as I have pointed out, she draws extensive conclusions about the types of character traits a moral saint would be required or forbidden from having. So it's worth taking a look at this vein in social psychology to see whether it bears on our understanding of moral sainthood.

Situationism and Character Traits

In recent decades, social psychologists have amassed a body of experimental data that purports to show that human behavior correlates much more strongly with features of situations than it does with features of an individual's personality or character. This school of psychological thought has come to be known as "situationism." In the most famous experiment, Stanley Milgram found that ordinary men who volunteered to participate in a study of learning and memory displayed a shocking willingness to inflict pain on other human beings. The subjects of Milgram's experiment were told by an authoritative looking figure to punish a "learner" for incorrect answers by flicking switches to administer increasingly painful electrical shocks that, the subjects were told, would not cause permanent damage. All the subjects complied initially, and two-thirds of them continued complying despite the cries and protests of the "learner" (actually a tape-recording) until they had reached the highest level, 450 volts, labeled on the device as "XXX" (Milgram, 1974).

The disturbing results of the Milgram experiment and countless others give us reason to call into question some of our oversimple folk psychological beliefs—for instance, that people's behavior is determined more by their "character" or personality than by the situation they find themselves in. Psychologists call this folk belief the "fundamental attribution error": we falsely attribute too much of human behavior to facts about the person.

Recently, Gilbert Harman (1999, 2000), John Doris (1998, 2002), and others have argued that the situationist literature (of which the Milgram example is just one small

part) has significant consequences for moral theory. Doris thinks the upshot of situationist findings is that "people typically lack character" (1998, 506). The account of "character" in play here is what Doris calls "globalism," according to which traits are *consistent* (across a variety of situations), *stable* (in multiple iterations of situations), and *evaluatively integrated* (particular traits of a given evaluative valence will be correlated with other traits of the same valence in the same person) (2002, 22). Globalism, Doris claims, factors in many character- or virtue-based moral arguments, but is in fact an empirically inadequate theory.

It would stray too far from our purposes in this chapter to attempt a full-scale analysis of whether Doris and others are correct about the relevance of the situationist findings to ethics. After all, this would require evaluating Doris's interpretation of reams of empirical data culled from experiments designed to test a wide variety of phenomena. Certainly, Doris's claims have generated much criticism. Recently, Rachana Kamtekar has argued convincingly that "the character traits conceived of and debunked by situationist social psychological studies have very little to do with character as it is conceived of in traditional virtue ethics" (2004, 460). And Gopal Sreenivasan has offerred several compelling criticisms of Doris's interpretation of the experimental data (2002).[13]

As it happens, I think both of these critiques present a serious challenge to the strong skepticism about moral character that Harman and Doris's work has spawned. Still, returning to the question of moral sainthood, I think we can draw the following

[13] For other interesting contributions to this debate, see also Besser-Jones, "Social Psychology, Moral Character, and Moral Fallibility" (2008) and Vranas, "The Indeterminacy Paradox: Character Evaluations and Human Psychology" (2008).

moderate conclusion: to the extent that the concept of "moral character" and the concept of a "character trait" are empirically problematic—and I agree there is reason to think they are at least *somewhat* problematic—then we should hope that our account of moral saints does not depend essentially on either of these concepts. We should favor an account based on axiological concepts (those pertaining to goodness) and/or deontological concepts (those pertaining to obligation or rightness) over an account based on virtue. This is one reason why my account of sainthood focuses on the features of actions that are not essentially related to the virtue or character of the person performing them—features like the action's being good but not required, its involving self-sacrifice, its being morally significant, etc.

Still, I take it to be an unresolved empirical matter whether there are such things as stable character traits. Correspondingly, I think it's an empirical matter whether moral saints would turn out to have distinctive profiles of character traits. Certainly we would expect that if there is such a thing as, for instance, honesty, moral saints would be more likely than average to display it. And even Doris's account is compatible with there being character traits; he just thinks they will be more finely-grained than folk psychology normally allows. Perhaps moral saints will be more likely to have certain of these fine-grained traits.

Indeed, there could be an interesting compromise between the situationists and those who think character traits factor prominently in morality. Even if it is true that the *situation* is the most powerful determinant of a person's behavior, it may also be true that people are *differently disposed* to winding up in certain situations. Obviously, social factors affect what sorts of situations people end up in, but the idea here is that

personality or character-based factors must play a role as well. Perhaps one distinctive thing about moral saints is that they are disposed to put themselves in the types of situation in which they will be compelled to act to the benefit of others. For instance, just as two-thirds of Milgram's subjects were willing to flip the switch, perhaps two thirds of doctors who see first-hand the level of poverty and disease in rural Haiti would choose to spend their lives tirelessly working to help people in need. But surely it is no *accident* that Paul Farmer found himself in the situation of seeing first-hand the suffering in Haiti. Indeed, he chose to go there while still a college student, before he had even begun medical school. Even if robust character traits don't factor causally in how Farmer reacts to a given situation on a given day, it seems plausible that character traits or some facsimile of them could factor causally in the types of situations in which he finds himself. Whereas the experimental literature on situationism focuses largely on *artificial* situations encountered by people who know they are participating in an experiment, perhaps an equally interesting subject of study is the decidedly *non-artificial* situations that real people find themselves in over the course of a lifetime.

CHAPTER 3

DE DICTO DESIRES AND MORAL FETISHISM

Consider two kinds of moral agent. First, consider the morally best agents, or the "moral saints": those who sacrifice their own interests or risk their safety in order to help others, who perform actions that are good but not morally required, who remain committed to a moral mission even in the face of opportunities to give up, and who do all of this for the right reasons. Now consider on the other hand, not the morally worst or most evil agents, but rather the agents whose moral motivation seems to be phony, superficial, pathological, or misguided. Broadly speaking, we can call these agents "moral imposters." One species of moral imposter is the moral fetishist: the person who is unhealthily obsessed with morality.

Now, a very well-disguised moral imposter might sometimes be mistaken for a moral saint. But we would not expect a true moral saint to share the negative qualities of a moral imposter, and certainly we would not expect a moral saint to treat morality as a fetish. Indeed, the moral saint would be the *last* type of person we would expect to make a fetish of morality. It is puzzling, then, that according to the most prominent account of moral saints—Susan Wolf's account—moral saints treat morality in exactly the way that Michael Smith has criticized as being a "moral fetish or vice." On Wolf's account, moral saints are motivated not by the right-making features of acts, but rather by the rightness

of those acts itself, as an abstract concept. In other words, these agents want to do what is right, but where this is read *de dicto* and not *de re*. It is the *description* of an act as right, and not the act itself, that seems to be driving these agents. I will call this sort of desire or motivation "*de dicto* moral motivation." If Susan Wolf is right that moral saints are motivated in this way, and if Michael Smith is right that this type of motivation is a fetish or moral vice, then it seems that the morally best people are not so good after all, since surely moral fetishism is incompatible with moral sainthood.[14]

Fortunately, I think there is a way around this problem. In this chapter I argue that we should understand *de dicto* and *de re* moral motivation as complementary rather than competing. Moral saints need not favor *de dicto* moral motivation *at the expense of* a corresponding *de re* motivation, as Wolf's account seems to require. Moreover, once we no longer view the two types of motivation as mutually exclusive, we see that *de dicto* motivation need not be a fetish or moral vice. In fact, I argue that *de dicto* moral motivation can play an important role in regulating our moral behavior.

This chapter is divided into several sections. First, I address the question of whether moral saints are really as dominated by *de dicto* moral motivation as Susan

[14] The distinction between *de re* and *de dicto* moral desires has garnered attention since Michael Smith mentioned it, briefly but controversially, in an argument about the debate between metaethical internalism and externalism. Smith made two central claims: that externalists are committed to positing a *de dicto* desire to be moral, and that such a desire is fetishistic. Smith's argument spawned a vigorous philosophical discussion, but that discussion has focused mainly on the first claim. While several philosophers have disagreed with *both* of Smith's claims, the amount of attention devoted to refuting or even understanding the second claim has been comparatively tiny. Here, however, I hope to engage with Smith's "fetish argument" *independently* of the debate over internalism. That is, I want to examine whether a *de dicto* desire to be moral *is* fetishistic, and what this means for moral saints, without taking sides in the debate between internalism and externalism. In some cases this will involve ignoring interesting arguments in the literature on *de dicto* desires—arguments about whether externalism entails these desires.

Wolf's account seems to imply. Next I address the question of whether *de dicto* moral motivation is really a "fetish or moral vice." I then suggest a few different ways that *de re* and *de dicto* moral motivation might coexist and productively interact in moral agents. Finally, I argue that a "non-buck-passing" account of rightness would support the division of motivational labor I've proposed.

Moral Saints and *De Dicto* Desires

Let us first get clear on the difference between *de re* and *de dicto* desires. The *de re/de dicto* distinction helps to resolve an ambiguity in statements about mental states like belief or desire. Consider, for example, a sentence like this one given by Jamie Dreier:

(K) Kalista desires to do what is right.

As Dreier explains, "(K) is ambiguous. It could mean that for each thing that is in fact right, Kalista desires to do that thing. Or it could mean that Kalista has a desire whose content is: to do whatever is right" (Drier 2000, 621-622). With Dreier and Michael Smith, I will call the first reading the *de re* reading, and the second one the *de dicto* reading. Notice that the *de re* reading attributes to Kalista a large number of *specific* desires directed at particular actions. We can think of these desires as unmediated or "original" as Dreier puts it (622). On this reading, Kalista is moved by the right-making features of the actions, not the rightness itself. On the other hand, the *de dicto* reading attributes to Kalista an abstract *standing* desire—a desire to do *whatever* happens to be right. "This desire," Dreier claims, "is one she could have even if she has no idea of what

the right thing to do is, or if she is uncertain" (622). And of course, it's a desire she can have even if she *does* have an idea of what the right thing is, but it's the *wrong* idea. If she thinks beating up homeless people is the right thing, then she'll desire to beat up homeless people.

With this distinction in place, we can now ask whether the morally best people—the moral saints—are more likely to be motivated to do what is right in the *de re* sense or the *de dicto* sense. Susan Wolf's (1982) account of moral saints seems to entail the latter. As we found in Chapter 2, Wolf argues that the people who lead the morally best lives are unattractive.[15] Most of us would want neither to be them nor be around them, she argues, because their commitment to morality would dominate their lives, leaving no space for humor, hobbies, or general well-roundedness. Part of what might be causing Wolf's saints to be so unattractive is the fact that they are driven by *de dicto* moral motivation seemingly at the expense of *de re* motivation. Of course, Wolf does not draw the distinction in these terms. But if we look at the way she describes moral saints, we find that the distinction plays an important role.

Wolf's moral saint is a perfectionist, "a person whose every action is as morally good as possible" (419). The saint's life, Wolf claims, is "dominated by a commitment to improving the welfare of others or of society as a whole" (420). This commitment to welfare might more naturally be read *de re* rather than *de dicto*, since caring about a person's welfare seems necessarily to involve caring *for that person*—caring about her welfare for its own sake, that is, for her sake.[16] But most of Wolf's other descriptions of

[15] Portions of the rest of this section also appear in Carbonell (2009).
[16] Of course, it is still *possible* to find a *de dicto* reading of, say, "Kalista wants to improve the welfare of others." Suppose, for example, that things like health, happiness,

the moral saint seem to indicate a *de dicto* moral motivation. She says, for instance, that the moral saint's life will be "dominated by explicitly moral commitments" (423) or, as she later puts it, "dominated by the motivation to be moral" (431). To have one's life dominated in this way is a bad thing, Wolf argues, and we can see this by looking at the heroes of history and literature. Since we prefer to read about mixed characters rather than people who are uniformly saintly, "there seems to be a limit to how much morality we can stand" (423). Of course, the fact that morally conflicted characters make for better stories does not necessarily mean that part of our conception of what makes for a good human life is that one be morally conflicted. Nevertheless, we can see Wolf's point: a life that is *dominated* by moral commitments begins to look tiresome, unhealthy, barren, perhaps even "pathological" (424).

Wolf returns to this point several times, and we can begin to see that what is unattractive about her moral saint is that he is committed to morality *de dicto*. It's not just that his moral commitments *themselves*—say, helping the poor and healing the sick—leave him too *busy* for a well-rounded life (though this is also one of Wolf's concerns).

---

and freedom are among the things that constitute a person's welfare. It could be the case the Kalista isn't committed at all to health, happiness, or freedom. Instead, perhaps Kalista follows the school of thought according to which one ought to deprive people of things like health, happiness, and freedom because such deprivation "builds character." If *that* is what she is committed to, then she is only committed to welfare *de dicto*. Still, it seems that with a concept like welfare, the *de re* and *de dicto* readings diverge less than they do with more abstract concepts like moral rightness. For even if Kalista is so deluded as to think the kicking a person when he's down is a way of fostering his welfare, she still seems to care *for him*. She's just wrong about what caring for him entails. Insofar as *de dicto* motivations are ever worrisome, they seem less worrisome when the object of the motivation is something like "welfare" than when it is "moral rightness."

Rather, there is something troublesome about the way in which he is committed to *morality*. Wolf writes that

> [T]here is something odd about the idea of morality itself, or moral goodness, serving as the object of a dominant passion in the way that a more concrete and specific vision of a goal (even a concrete *moral* goal) might be imagined to serve. Morality itself does not seem to be a suitable object of passion (424).

The distinction Wolf makes here between "morality itself" and a "concrete moral goal" is the same distinction I made earlier between the *de dicto* and *de re* readings of Kalista's desire to do "what is right." On the *de dicto* reading, Kalista is committed to morality itself, and on the *de re* reading she is committed to one or several concrete moral goals. Thus Wolf's discomfort with a commitment to "morality itself" is probably the same discomfort we feel, at least initially, when pondering the idea of a *de dicto* desire to do what is right.

Consider, for example, Wolf's claim that "there is something odd about the idea of morality itself, or moral goodness, serving as the object of a dominant passion" (424). We can see just how odd this would be by imagining a person who, in response to the question "What's your life's passion?" answers simply, "Morality" or "Moral goodness," and in response to the question "What are you going to do today?" answers, "Morally good things." There is something about these answers that makes us uncomfortable, even though presumably we are not uncomfortable with a person whose life is dominated by pursuits that are, in fact, morally good. That is, the way in which the respondent is *describing* his life's passion, which is surely a reflection of the way in which he is *committed to* his life's passion, strikes us as not only strange but suspect.

While I agree with Wolf that there is something fishy about a person so committed to an abstract concept, I'm not sure she's right that a moral saint would look like this. Instead, I imagine a moral saint who answers the above questions by giving the *content* of his moral commitments rather than the fact that he is so committed. So, to "What's your life's passion?" he would answer "Healing the sick" or "Eradicating tuberculosis," and to "What are you going to do today?" he would say, "See patients" or "Raise money."

Indeed, while I can imagine a person who answers instead "do morally good things," I can only imagine him giving this answer facetiously or ironically, or as some sort of pep talk to himself, to stay motivated in the face of obstacles. To give the answer "do morally good things" *sincerely*, one would have to be, oddly, a little cold. After all, describing one's commitments to oneself as moral is perhaps a consequence of being committed to them *as moral*. Yet it seems that what makes a person saintly is a commitment to various projects *for their own sake*. We'll return to this question later, though, since it may be the case that being committed to something *as moral* is in fact benign or even beneficial, assuming one is at the same time committed to the project for its own sake.

One problem with people who are committed to or obsessed with "morality" so described is that whatever it is they are describing as "morally right" may not actually *be* morally right. Many of us associate an obsession with morality *de dicto* with the tendency to have false beliefs about what morality demands. Think, for example, of Jerry Falwell's "Moral Majority" or any organization that describes itself as being on a "moral crusade." At the very least, there is a correlation between *de dicto* moral obsession and the

tendency to have *fanatical* moral beliefs, if not false ones. Indeed, Wolf acknowledges that the saint as she depicts him somewhat resembles a "moral fanatic" (425). So even if we stipulate—as it seems reasonable to do—that a moral saint will not have *false* beliefs about what morality demands, there is still something problematic about a moral saint whose desire to do the right thing is read purely *de dicto*.

In fact, Wolf concedes that her view seems to paint the moral saint as a "disgusting goody-goody or an obsessive ascetic" (425). But what's disgusting about a goody-goody is not that she *does* morally good things, but rather that she describes them as morally good, takes their moral goodness as a mark of her own superiority, basks in the glory that accompanies doing good things, and is concerned more about the appearance of moral goodness than the actuality. We don't find a goody-goody by looking for someone who makes a significant moral impact on the world; we find her by looking for someone whose self-image is unhealthily tied to her reputation for making such an impact. Thus in addition to confusing *de re* moral motivation with *de dicto* moral motivation, Wolf's saint has the further problem of substituting concern for her moral self-image with concern for the moral projects themselves. Perhaps her superficial *de dicto* moral motivation is even *derived* from her desire to improve her moral self-image.

The same goes for the obsessive ascetic. What is unattractive about such a character is not that he sacrifices his own desires to help others—helping others is rarely unattractive in itself—but that he elevates the self-sacrifice to the level of an obsession, and that he thinks the sacrifice *itself* intrinsically good. Thus one worry about a commitment to morality *de dicto* is that, even if the person with this commitment has obsessions that actually are morally good, he seems to have them for the wrong reason.

He seems to be benefiting others because it's the "moral thing to do" and not for the beneficiaries' own sakes. Paul Farmer, a doctor who I have argued is a real-life moral saint, calls these misguided people "do-gooders." The do-gooders tend to behave in such a way as to provoke others to refer to them as "saints," where "saint" is an insult. But these aren't the moral "saints" we're interested in. Looking for real moral saints among the goody-goodies and do-gooders would be like looking for excellent fathers by rounding up men who wear "World's Greatest Dad" t-shirts.

What I've tried to show so far is that Wolf's account clearly ascribes to moral saints a *de dicto* moral motivation, and it may be in part due to this type of motivation that these moral saints come out looking so unattractive. Indeed, the moral saints in Wolf's account seem to make a fetish out of morality, or at least to be in danger of doing so. After all, they are obsessed with morality; their lives are dominated by this abstract concept, making them irritating and rigid. So the question naturally arises, is it the *de dicto* moral motivation *itself* that is causing the moral fetishism? Are the motivation and the fetishism in fact one and the same thing? Certainly, we should seek an account of moral saints according to which they are not moral fetishists. Does this mean we need an account that strips them of all *de dicto* moral motivation? In what follows I argue that the answer is no. There are many ways to have a *de dicto* motivation to be moral, and Wolf has offered us a picture of the most objectionable way. Indeed, Wolf's account seems to presuppose the idea, pervasive in the literature on *de dicto* desires, that *de re* and *de dicto* motivations to be moral are to some extent mutually exclusive. We can see this in her discussion of the difference between having as one's dominant passion "morality itself" or "moral goodness," on the one hand, or a "concrete moral goal," on the other hand

(424). It seems that she disapproves of a person whose dominant passion is for *morality itself* at least partly because she is assuming that this person cannot *also* have a passion for concrete moral goals. But why must this be the case?

Indeed, there is another reason to be skeptical of the apparent connection between *de dicto* moral motivation and fetishism that arises when we look at Wolf's saints and what makes them unattractive. When we read about someone obsessed with morality itself in the way Wolf describes, and we see that this person comes dangerously close to being a "disgusting goody-goody or obsessive ascetic" or even a "moral fanatic" (425), we begin to conjure up an image of a certain sort of moral imposter. This imposter comes in several varieties, including: (1) a person who is obsessed with "morality" but is wrong about which actions are in fact morally good; (2) a person whose concern for morality is insincere or phony; (3) a person who is really only interested in the *appearance* of being morally good; (4) a person who is really only interested in the praise or admiration that comes with being thought by others to be morally good; (5) a person who is so fanatical about enforcing moral standards in one arena that she stomps on moral standards in another arena; and (6) a person whose interest in morality waxes and wanes, for example when moral causes go in and out of fashion.

Surely I've left out some ways of being a moral imposter, and many people may embody more than one of the above characteristics. For now I simply want to make note of the fact that *none of the above necessarily follows from having a standing,* de dicto *desire to do what is right*. Indeed, there is no reason to think any of the above follows even from having morality itself as the *object of a dominant passion*.

Instead, perhaps what is going on is that, in our collective experience as observers of various moral personalities, we have found that an obvious, identifiable *de dicto* moral motivation is quite often *correlated* with one or many of the above varieties of moral fakery. Indeed, just *imagining* a person for whom "moral goodness" itself is the object of a dominant passion may automatically cause us to see this person as an imposter, for we have found that people who are up to no good often claim to be on a moral mission, while people who are doing more than their share may hesitate to identify "morality" as their primary motivational force. But these relationships may be only contingent.

In order to better understand the relationship between motivation and fetishism, we need first to look more closely at Michael Smith's argument.

Smith's "Fetish" Argument

As part of his famous argument against externalism, Michael Smith argues that a *de dicto* motivation to do the right thing is fetishistic. The only way an externalist can explain the "reliable connection" between a change in our moral judgments and a change in our motivation, Smith claims, is to posit "a motivation to do the right thing, where this is now read *de dicto* and not *de re*. At bottom, the strong externalist will have to say, having this self-consciously moral motive is what makes me a good person" (1994, 74). Suppose, for example, that I used to judge littering to be permissible and I now judge it to be wrong. Suppose further that I used to be motivated to litter and now I am motivated not to. If my change in motivation is explained externally, then it must be the case that my concerns about littering are indirect: they are derived from a general concern to do

"the right thing," whatever it may be. I'm not concerned about the littering itself; I'm merely concerned about littering insofar as it falls under the description "morally wrong."

This sort of self-conscious, *de dicto* moral motivation is "quite implausible," Smith argues. According to Smith, our moral concerns should be direct and non-derivative.

> For commonsense tells us that if good people judge it right to be honest, or right to care for their children and friends and fellows, or right for people to get what they deserve, then they care non-derivatively about these things. Good people care non-derivatively about honesty, the weal and woe of their children and friends, the well-being of their fellows, people getting what they deserve, justice, equality, and the like, not just one thing: doing what they believe to be right, where this is read *de dicto* and not *de re*. Indeed, commonsense tells us that being so motivated is a fetish or moral vice, not the one and only moral virtue (75).

Why would it be a fetish or moral vice to care only about doing "what one believes to be right," *de dicto*? As we discussed earlier, one reason is that what one believes to be right might diverge from what is *in fact* right. Another reason is that a person who is more concerned with the fact that her commitment is described as "morally right" than with the content of the commitment itself might be simply keeping up appearances. Of course, as I have already noted, keeping up appearances might simply be something we *associate* with *de dicto* moral motivation rather than something that is necessarily connected to that type of motivation.

Smith's own explanation for what's vicious about *de dicto* motivation is found in the following passage.

> For the objection in this case is simply that, in taking it that a good person is motivated to do what she believes right, where this is read *de dicto* and not *de re*, externalists provide the morally good person with 'one thought too many'. They alienate her from the ends at which morality properly aims. Just as it is constitutive of being a good lover that you have direct concern for the person you

> love, so it is constitutive of being a morally good person that you have direct concern for what you think is right, where this is read *de re* and not *de dicto* (76).

So Smith's objection to *de dicto* motivation is that to be motivated in this way is to have "one thought too many," and as such to be alienated from the proper ends of morality. Whereas our earlier concerns about *de dicto* motivation—that it could be misguided or insincere—seemed only contingent, Smith's accusation is stronger. His claim that this sort of motivation is alienating is meant, apparently, to hold for *all* cases of *de dicto* motivation, even those that are neither misguided nor insincere. The worry is that *de dicto* moral motivation is alienating to agents insofar as it acts as a filter through which their otherwise direct concern for moral projects must pass, rendering that concern merely derivative and hence illegitimate.

But what's wrong with having "one thought too many"? This phrase comes from Bernard Williams' famous argument about impartiality (1981, 18). Williams imagines a scenario in which a man could save only one of two people, and one of them was his wife. Those who think the man's motive ought to be morally impartial would say that, if he saves his wife, his reasoning must include *two* thoughts, that *it's his wife* and that *even an impartial morality permits him to favor her in this case*. Williams thinks the second thought is unnecessary and, worse, that if you need to consider the second thought then you haven't fully appreciated the force of the first thought. The second thought is one too many.

Now just as Williams thinks the extra thought alienates the man from his wife, Smith thinks that needing to consult a general standing motivation to do "the right thing" *de dicto* manifests an alienated concern for the object of right action. So, when I change

from judging littering to be morally permissible to judging it morally wrong, the externalist, according to Smith, would have to say that my newfound motivation not to litter is merely derived from my motivation to do whatever happens to be "the right thing." While we may not yet agree with Smith that this is a *fetish* or *vice*, we can see what might be troubling about it by considering the difference between morality and mere social convention. If tomorrow I wake up and learn that the American convention of driving on the right side of the road has been reversed, I will be moved to drive instead on the left. But my motivation will not be direct or "original"—it will not respond to any intrinsic features of driving on the left. In the case of a mere convention, this is perfectly fine. But when we're dealing with morally significant actions like whether to help the injured animal I've just hit with my car (because I was so distracted by having to drive on the left), it's expected that I will have *direct* concern for the animal.

One question that immediately springs to mind is why having an indirect or derived moral motivation rules out *also* having a direct, non-derivative motivation. The easy answer to this question is that Michael Smith has just *posited* that externalism entails motivation *de dicto* but *not de re*. And indeed he says exactly this when he claims that the externalist thinks the "good person is motivated to do what she believes right, where this is read *de dicto* and not *de re*" (76). But this doesn't seem to be exactly what is going on in the Williams' drowning example. Williams says we would hope that the man's "motivating thought, fully spelled out, would be the thought that it was his wife, not that it was his wife and that in situations of this kind it is permissible to save one's wife" (18). But if Smith wants us to believe that the thought "*it was his wife*" represents a *de re* concern for doing the right thing (that is, a concern directly for the wife), and the thought

"*in situations of this kind it is permissible to save one's wife*" represents a *de dicto* concern for doing the right thing, then what we have is a person motivated to do what he believes to be right *de dicto* and *de re*, not *de dicto* but not *de re*.

Perhaps what is going on is that both Williams and Smith see the "it was his wife" thought as being insufficient, on its own, to motivate the man to jump in the water. That is, a man whose concern for his wife only motivates him to save her life *once he has consulted a more general principle* must be a man who is not *sufficiently* concerned for his wife, or perhaps not even *genuinely* concerned for his wife.

But surely there is another way of understanding this situation. Perhaps the man's concern for his wife *was*, on its own, sufficient to motivate him to rescue her, but he consulted the general principle as a way of being extra scrupulous. In other words, because he had two thoughts rather than just one, the action was motivationally overdetermined.[17] This approach turns Smith's thesis on its head: instead of seeing the man as fetishizing morality, we can see him as being particularly morally diligent.

Williams asked us to consider what things would look like from the wife's perspective. The wife, he claims, would want the man to stop thinking after he got to "it's my wife." But I wonder whether we only ascribe this impatience to the wife because she is drowning and doesn't want the man to waste too much time deliberating. In other circumstances she might in fact *prefer* that he husband consult a general moral principle, because doing so does not necessarily demean the concern he has for her.

For example, what if the wife were being attacked by a herd of endangered elephants, and in order to save her the man would have to kill several of the elephants?

[17] In "Are Desires *De Dicto* Fetishistic" (2002), Jonas Olson suggests that cases of "motivational overdetermination" are "quite common" (91).

That the man has the thought "killing these elephants, while terribly unfortunate, is morally permissible" doesn't seem to weaken or cheapen the thought "oh dear, my wife!" The thought "oh dear, my wife!" is no less genuine here than it would have been had it been the man's *only* thought. Indeed, we can see that the wife might even *appreciate* this sort of extra deliberation if we alter the case even more. Suppose that in order to save his wife the man would have to drive his boat through ten swimming children, drowning them. Surely in this case it is better that, instead of stopping at "oh dear, my wife!" he proceeds to consider whether it is permissible to save his wife in these circumstances. While we all want our loved ones to care about us non-derivatively, this doesn't mean we want them to act on this care *unreflectively*. Having a standing, general desire to do the right thing, *de dicto*, is often a way of mediating our unreflective motivations.

Surely it is compatible with being morally extraordinary that a person be motivated to do the right thing both *de re* and *de dicto*. Of course, it may be the case that one way of being a moral saint is to be unusually accurate in directing one's non-derivative concern toward beings who need that concern. For example, we can imagine a person who, in the original drowning case and the elephant case, thinks simply "my wife!" and does the right thing, and in the ten-swimming-children case thinks simply "ten children!" and does the right thing, never needing to subject his concern to further scrutiny by consulting an abstract principle or desire. In some respects, this sort of saint resembles what Susan Wolf called the "Loving Saint"—a person who is extraordinarily compassionate, a person for whom good deeds come easily.

But there are many ways of being morally extraordinary; some moral saints are probably quite cerebral: they are constantly in the business of weighing reasons and

consulting general principles, never trusting their gut reactions. This is the sort of saint who employs both *de re* and *de dicto* motivation. In these cases, the desire to do the right thing *de dicto* is not a fetish but instead a sophisticated regulatory mechanism. While this is not exactly what Susan Wolf meant when she coined the term "Rational Saint," we can see how this sort of saint is particularly rational. The question that remains is whether this sort of saint could have *just* the *de dicto* motivation, or if *both* types of motivation are necessary. Intuitively, there would seem to be something rather empty or fake about a "saint" whose *only* motive was to do "the right thing" *de dicto*. Interestingly, in the next section we'll look at an argument that *de dicto* motivation alone is sufficient, if not for the moral saint then at least for "morally good person." But for now it should suffice to say that, at a minimum, we should allow the moral saint a little leeway in getting her *de re* concern for her moral projects to align with her *de dicto* commitment to moral rightness.

I've tried to offer an alternative to Smith's analysis, an alternative in which *de re* and *de dicto* moral motivation complement each other rather than compete. Of course, this account depends on abandoning the notion, implicit in Smith's argument, that *de re* and *de dicto* motivation are mutually exclusive. We'll return to this question later when we examine the complex interaction between *de re* and *de dicto* desires. In the next section I'll look briefly at how philosophers have responded to Smith's fetish argument.

Responses to Smith

Several philosophers have recently challenged Smith's fetish argument, and their criticisms can be useful in developing a view about how *de re* and *de dicto* motivation

might interact. Hallvard Lillehammer (1997) challenges all three parts of Smith's argument: that internalism entails a desire to do the right thing *de re*; that externalism entails a desire to do the right thing *de dicto*; and that *de dicto* desires are a moral fetish. His criticism of the fetish claim is found in the following passage:

> The claim that it is a fetish to care about what is right, where this is read *de dicto*, is false. It is false even for Smith's basic case, where an agent changes his most fundamental values. Consider the case of someone who has always believed that morality is not very demanding in terms of individual sacrifice. Suppose he comes to believe that he is morally required to sacrifice everything he has, perhaps even his life. Suppose further that he does not directly acquire a *de re* desire to do what he now thinks is right, but that a standing desire to do what is right *de dicto* provides the causal link which motivates him to sacrifice everything he has. It is not a platitude that this person is a moral fetishist. Maybe it would be admirable if he eventually came to care about what is right in an underived way. But given what he now considers morality to demand, he might be forgiven if his immediate concern for what is right is not direct (191-192).

Here Lillehammer seems to be making a point similar to the one I made earlier in the ten-swimming-children case. Just as the man might be forgiven for not having a direct concern for the children he would have to drown in order to save his wife (not because he is callous but because he is overwhelmed by his concern for his wife), Lillehammer's agent can be forgiven for giving his direct concern a little time to catch up with his standing *de dicto* commitment to doing the right thing, especially given how much of a sacrifice that commitment is asking him to make.

Lillehammer offers another example, one meant to capture that *de dicto* motivation is not fetishistic even in cases where a person changes her moral judgment. Suppose a woman who is "tired of her husband" and "temporarily indifferent" to his feelings contemplates having an affair with a stranger, but judges it to be wrong.

> However, she has a standing *de dicto* desire to do what is right which, together with her moral judgment, causes her to do the right thing, in spite of the absence

> of a *de re* desire to do the right thing and the presence of a *de re* desire to do the wrong thing. If there is anything in this case which prevents this person from being good it is not her standing desire to do what is right, where this is read *de dicto*. For this desire is playing the role of an internalized norm that prevents her from being tempted to do wrong. Such norms are not in contradiction with the platitudes that are definitional of moral discourse. Their benefits are all too obvious (192).

This is a good example of why a *de dicto* desire to do what is right does not have to be fetishistic. Indeed, in this example *de dicto* desire plays an indispensable role in leading the person to do the right thing. But I'm not sure the example gets us all of the conclusions Lillehammer wants. For one thing, his claim that the *de dicto* desire "prevents her from being tempted to do wrong" seems to contradict his earlier claim, in the same paragraph, that she "is tempted to have an affair." While it is plausible that the *de dicto* desire here prevents her from *doing* wrong, it seems a stretch to say that it prevents her from being *tempted* to do wrong.

Moreover, we might wonder whether Lillehammer's premise that the woman is "temporarily indifferent to her husband's feelings" is too convenient. It's because of this stipulation that he can claim that she both has a *de re* desire to do the wrong thing (cheat) and lacks a *de re* desire to do the right thing (not cheat; be considerate of her husband's feelings). But perhaps we should have a more nuanced view of *de re* desires. Surely she still *has* a *de re* concern for her husband's feelings; one explanation is that it is consciously inaccessible for the moment. Another possibility is that she has both a *prima facie* desire to cheat and a *prima facie* desire to protect her husband's feelings, but because she is in the throws of temptation she is unable to competently assess what she desires *all things considered*. Instead, she uses her general commitment to morality *de*

dicto as a shortcut, but if she were able to weigh her *de re* commitments she might find that, in fact, her husband's feelings outweigh the temptation to cheat.

In any case, we can see that the sense in which Lillehammer's tempted spouse lacks the relevant *de re* desire is rather trivial; she lacks a concern for her husband's feelings locally—right now, tonight—but not globally. Keep in mind that Smith and Lillehammer are discussing "good and strong-willed" people, not necessarily moral saints. But surely even a moral saint will have moments when *de re* concern for the right thing will be masked, suppressed, overwhelmed, or even temporarily extinguished. In these moments, a general commitment to doing the right thing *de dicto* can act as a patch, filling in the gaps that interrupt an otherwise strong direct concern. Notice, though, that this is compatible with the view that in order to be a moral saint one must generally have a *de re* desire to do the right thing. When the *de re* desire is momentarily absent, the *de dicto* desire plays a vital role, but never to have the *de re* desire in the first place would be problematic.

Lillehammer is not the only philosopher to take on Smith's fetish argument. In "Belief, Reason, and Motivation: Michael Smith's 'The Moral Problem'" (1997), David Copp offers an extensive critique of Smith's case against externalism. Along the way he addresses the fetish claim. "I do not think it is fetishistic to have the *de dicto* desire," he writes. "A good person could have this desire along with a variety of direct desires, such as the desire for the good of her loved ones" (49-50). Unfortunately, Copp does not offer any sort of argument for this claim, but we can imagine that his thoughts are similar to those outlined above.

On the other hand, Sigrun Svarvarsdottir takes up the fetish claim in some detail in her article "Moral Cognitivism and Motivation" (1999). With Lillehammer and Copp, Svarvarsdottir doesn't think externalism needs to posit a *de dicto* desire to do the right thing.[18] This desire need only be "*a part of* the motivational structure of the good person" (199). Svarvarsdottir claims further that "insofar as the force of Smith's objection trades on ascribing to externalists a commitment to such a monolithic conception of the good person, it is not to be taken seriously" (199).

Recall that earlier I listed a number of ways in which a person might be a "moral imposter"—that is, a number of types of "moral fetishism" *broadly construed*. The list included traits like insincerity, fanaticism, and concern for appearances. But since none of these traits seemed to follow—at least not straightforwardly or intuitively—from having a *de dicto* desire to do what is right, we decided that it was not these flaws in particular that Smith must have been referring to when leveling the charge of fetishism. The only clue we have as to what Smith may have actually meant by "moral fetish" is his discussion of Williams' "one thought too many" and what it means to be alienated from the proper ends of morality. But to call a position a "moral fetish" strikes me as much stronger than to call it alienating. Svarvarsdottir seems also to think that moral fetishism would be a rather severe affliction. She defines it as follows:

> It would be characteristic of holding oneself and others to very rigorous moral standards, while being completely unwilling to entertain any reflective question about their nature or grounds. It would be accompanied by a fear of any skeptical

[18] One might think that all these arguments against the claim that externalism entails a *de dicto* desire to do the right thing would make it unnecessary for us to even examine the further claim that the *de dicto* desire is a fetish. But, setting aside the debate over externalism, we have independent reason to think that some people, including moral saints, might have this *de dicto* desire, so we need to know whether it is fetishistic even if Smith's argument against externalism fails.

> questions about morality, and a refusal to take them seriously enough to even attempt a thoughtful answer. The question 'Why be moral?' would be branded as irreverent and illegitimate (200).

This is perhaps an extreme version of moral fetishism, and thus it may be uncharitable to think it is what Smith meant. However, since Smith meant for the "fetish" charge to be a decisive blow to externalism, it's plausible he meant something as bad as this. The problem, of course, is that it is a great leap indeed from a person who "wants to do the right thing" *de dicto* to a person who is as militantly obtuse about morality as the person described above. Indeed, when we think of Svarvarsdottir's fetishist, we think of the paradigm moral imposter: the goody-goody who is obsessed with "morality" but not very good at detecting which actions are in fact morally right, and who is overly concerned with appearances. But as we determined earlier, while we might be tempted to associate *de dicto* motivation with this constellation of moral failings, we would be wrong in thinking this association is anything more than a correlation.

Svarvarsdottir agrees, writing that "a concern for being moral should not be confused with a rigorous obsession with morality or a resistance to examine hard reflective questions about morality" (200). To think that a concern for morality would entail anything so unsavory as "obsession" or "resistance" strikes Svarvarsdottir as profoundly wrong; indeed, she thinks it would be wrong to think there could be *anything* bad about having a genuine concern for morality:

> Now, I cannot see how it would make one a worse person if the disposition to care about what one deems morally valuable were due to a desire to be moral. I would think that the crucial thing is that one has such a motivational disposition (as well as certain others) and that it is not due to some desire such as to impress other people or to stay out of trouble or to obtain some other personal gain from being perceived as morally conscientious (200-201).

Ultimately, Svarvarsdottir settles on the same defense of *de dicto* moral motivation that we arrived at earlier in this section: that this sort of motivation does not rule out also having a direct, non-derivative concern for the object for its own sake, and that this motivation conveniently fills in the gaps when such direct concern is missing, especially after we've changed our moral judgments.

> Admittedly, we expect a good person to develop a deep commitment to an end she has come to see as morally valuable and to pursue it for its own sake. … The presence in the good person of the desire to be moral certainly does not prevent her from forming such a commitment. *Although her desire to ϕ may initially be derived from her desire to be moral, it may subsequently come to operate psychologically independently of the latter* (205-206, emphasis added).

In the morally *best* people, we will want to say not just that the desire to ϕ *may* develop into an independent, underived desire, but that it *must*. That is to say, a *de dicto* motivation to do the right thing *by itself* is not sufficient in the morally best people. As Svarvarsdottir says, if someone never converted the *de dicto* motivation into *de re*, "I'd hesitate to hold his personality up as a moral ideal" (205).

There are two more responses to Smith's argument that we ought to at least briefly address. James Dreier, in "Dispositions and Fetishes: Externalist Models of Moral Motivation" (2000), is one of the only responders to at least putatively *agree* that *de dicto* motivation is fetishistic. "An agent is a moral fetishist, in our sense," he writes, "just in case what appeals to the agent about the moral actions that she wants to perform, is that they are moral actions" (629). That is, to desire the right thing *de dicto* just *is* to be a fetishist, because you are desiring it for the wrong reason. (This, obviously, is a weaker notion of "fetishism" than Svarvarsdottir's, but nonetheless it is strong enough that we would be troubled if it allowed moral saints to come out as fetishists.) "The good moral

agent, by contrast," Dreier writes, "finds attractive the properties that *ground* the rightness, the right-making properties, or what she takes to be the right making properties" (629). This rightness-responsiveness is, essentially, the same idea as a *de re* motivation to do the right thing.

I agree with Dreier that the *de dicto* motivation as he describes it would be fetishistic if it were the *only* motivation the agent had, and if it were the dominant motivation driving most or all of her actions throughout her life. But I think Dreier would find this sort of motivation unproblematic in the context of the picture we've been developing thus far, according to which this *de dicto* motivation is only present *in addition to* the *de re* motivation, and serves the important role of filling in the gaps in *de re* motivation that will occur in any psychologically normal human being. In any case, Dreier doesn't think the *de dicto* motivation is what the externalist needs; instead, he proposes that externalism entails a "second order desire" model, a model where good moral agents have a standing desire to desire to perform actions with right-making features. On this model, the second-order desire is unproblematic, because it interacts with *de re* desires in the way we discussed earlier: "While the second order desire does play an original causal role in generating the admirable first order motivations, it needn't play any maintenance role once the first order motivation is formed" (636).

Finally, Jonas Olson, in "Are Desires *De Dicto* Fetishistic" (2002), is perhaps the only philosopher explicitly to question why an agent cannot have *both de re* and *de dicto* desires to do the right thing.[19] Olson mainly summarizes the ground covered by earlier

[19] Indeed, Olson offers up some cases to show why he thinks *de dicto* motivation is sometimes *preferable* to *de re*, but I don't find his cases terribly compelling, certainly no more compelling than the similar cases put forward by Lillehammer.

responses to Smith, but he considers not only the "morally good person" but also the "paragon of moral goodness"—presumably, a moral saint. With Sobel (2001)[20], Olson argues that the *de dicto* desire to do the right thing—the gap-filler, as we've characterized it—would not be necessary in the paragon of moral goodness, because "her *de re* desires to perform acts with right-making characteristics would in each and every case provide her with sufficient motivation to act on those desires (that's what would make her a paragon)" (92). As I said earlier, I think this is *one way* of characterizing the moral saint: as a hyper-accurate rightness-responder, a "Loving Saint." But I think it is unrealistic to expect a person—even a moral saint—to have *no gaps* in her *de re* moral motivations. As such, I think a moral saint can fall back on a standing *de dicto* desire to do the right thing at no threat to her sainthood. Indeed, it could be the case that one way of being a moral saint is to have a *particularly active* standing *de dicto* motivation: that is, to be conscientious about aligning one's actions with the abstract property of rightness.

How *De Re* and *De Dicto* Desires Interact

Thus far we've encountered the beginnings of an account of exactly how *de re* and *de dicto* motivations interact. The first and most obvious feature of the relation between the two is that *de re* desires are primary in certain senses: all else held equal, we find it morally preferable for a person to be motivated to act rightly *de re* rather than *de dicto*; *de re* desires are those that respond to an object for its own sake, and this is the most direct and therefore most legitimate way of responding to an object; when someone

[20] Olson references a 2001 version of Sobel's unpublished manuscript "Good and Gold: Metaethics from Moore through to Mackie" but the version available at the time of this writing is newer: 2006.

is doing the right thing *de re* she is necessarily succeeding in performing actions which are in fact right, whereas a desire to do "the right thing" *de dicto* could potentially lead her astray; and finally, desires *de re* are deliberatively prior, meaning they require no higher-order thought processes and as such they prevail as the primary motivating force in the vast majority of low-deliberation moral actions.

However, a desire to do what is right *de dicto* plays an important role in regulating and, in some cases, replacing *de re* desires. There are at least four roles *de dicto* motivation might play in complementing *de re* motivation.

The first is what I'll call "*Higher-order moral reasoning*." When a man is faced with saving either his wife or a stranger and, in deciding to save his wife, he has not just one thought—"it's my wife!"—but *two* thoughts—"it's my wife and in this situation I am permitted to save her" he is exhibiting higher-order moral reasoning. Setting aside questions about whether his exercise of higher-order moral reasoning demeans his relationship with his wife, one thing is certain: a person who subjects his first-order moral motivations to a test against general moral principles is, at least in many cases, exhibiting a more sophisticated sort of deliberation. Just as those who think impartiality is central to ethics would laud the man for subjecting his initial thought to systematic scrutiny, those of us who think that sometimes it's better to be moved reflectively rather than unreflectively would laud a person who thinks, not just, "that's littering!" but "that's littering, and littering is morally wrong." In such a case, we think that the commitment to morality *de dicto* (as exhibited in the thought "littering is morally wrong") does not detract from the original direct motivation but in fact enhances it. After all, surely it is an asset to a moral deliberator to have the ability to step back from a situation and not only

perceive his *de re* desires but be able to examine them in light of more general principles. Whereas Susan Wolf painted a picture according to which *de dicto* motivation came out looking pathological, in a more charitable light this sort of motivation seems more like strength of moral conviction.

The second role is what I'll call "*Normative governance*." At its best, *de dicto* moral motivation acts as a check on *de re* desires, weighing them when they conflict and giving us a reason to choose amongst them. So, for example, if I find myself motivated directly to save my daughter from the cheetah attacking her and also motivated directly not to harm the cheetah, I may need to resort to my standing *de dicto* desire to do the right thing, which may, for example, tell me that in general I am permitted to show partial concern for my daughter even when doing so violates my obligation not to harm cheetahs. Presumably it is this phenomenon that Lillehammer was referring to when he said that *de dicto* desire sometimes serves as an "internalized norm" (192).

The third role is as a "*Motivational impetus*." As discussed earlier, sometimes when we change our moral judgment we do not immediately acquire a *de re* desire to do whatever our revised judgment tells us to do. Or perhaps we acquire a motivation but it is initially quite weak, whether because we are not completely certain about our new judgment, or because of weakness of will. In such cases, a standing *de dicto* desire to do the right thing can be a motivational substitute, or as Olson calls it, a "safety device" (92). For example, suppose that for a long time I judged it morally permissible to drive a car even when I hadn't slept in 36 hours, and I drove this way regularly. But suppose after some thought I changed my mind and decided that this practice was dangerous and morally forbidden. After a change in judgment like this, there might naturally be a grace

period in which I have a hard time being motivated directly not to drive my car. However, if I have a standing (*de dicto*) desire to do what is right, and I judge refraining from driving to be obligatory, then I can at least be motivated *derivatively* in accordance with my judgment, if not yet directly. In fact, it is in cases like these, where our desire to do what is right *conflicts* with our baseline desires, that we can be sure we are acting out of moral reasons and not just convenience.

The final role is what I call "*Epistemological stopgap*." *De dicto* moral motivation can act as a catch-all for the many contexts in which we want to do what is right but do not (yet) know exactly what is right, and therefore have no direct *de re* motivations toward any particular course of action. For example, suppose that I want to do what is right with respect to the question of whether to raise the minimum wage. Suppose also that I have learned that prominent economists disagree about whether raising the minimum wage is morally good, indeed this disagreement persists even among just those economists with a demonstrable interest in helping the poor. Since I don't know what course of action is right with respect to this question, I don't have any direct motivation one way or the other, but nonetheless I certainly have a standing motivation to do whatever happens to be right in the end. This is different than simply being *indifferent* with respect to what to do. Whereas the indifferent person might not care which way she votes on a ballot measure to raise the minimum wage, the person with the standing *de dicto* motivation to do what is right cares a great deal. Indeed, it is *because* she cares about doing the right thing that she will seek out information and deliberate intensely about how to vote. Once her epistemic gap has been closed and she has arrived at a

judgment, the *de dicto* motivation will move her to act even if she has not yet acquired an original desire to vote one way or the other.

When we consider the various ways in which *de re* and *de dicto* moral desires can interact, we discover that any position according to which one type is considered genuine and the other illegitimate is oversimplified. Indeed, that the distinction is oversimplified is made all the more evident when we consider cases in which it appears that *de re* and *de dicto* desires amount to one and the same thing. Earlier I suggested that perhaps something like this is the case with respect to a value like welfare. Now I'll consider justice.

When we respond to justice or injustice, it may be impossible to respond directly (*de re*) to the features that make an act just or unjust without at the same time responding to the act under the description "just" or "unjust." If, for instance, I am working to counteract voter disenfranchisement, I may be motivated directly by a sense of justice, but at the same time I may be responding to what Wolf would call the "abstract and impersonal consideration" *because it is just*. In other words, the content of my motivation might be *that people in one neighborhood had ample opportunity to vote, but people in another neighborhood were intimidated from voting*. But how is this motivation any different than if the content were simply "*an injustice*"? The disenfranchisement would obviously be unjust whether I described it that way or not, and my so describing it does not *make* it unjust, but in responding to it, I cannot help but respond to it under the description "injustice." That's because it's not obvious just from the descriptive fact that "these people could vote and those people could not vote" that I'm responding to right-making features. We need the added fact that the difference in treatment was *unfair*—that

it was an *injustice*—to make my direct motivation intelligible. To respond, *de re*, to "unfair differences in treatment" and to respond, *de dicto*, to "an injustice," seem to be one and the same thing. Indeed, if I fought injustice *de re*, without being able to describe it or conceptualize it as "injustice," I would seem to be lacking a crucial deliberative mechanism—the mechanism that picks out, in a non-arbitrary way, the special force of this wrong.

Given this and surely other cases in which the line between *de re* and *de dicto* motivation is blurry, the only reasonable account of the morally best people would have to be an account that accommodated both.

Moral Saints and Desires *De Dicto*: Two Models

I hope I've succeeded in demonstrating that, although it originally appeared that what was repulsive about Susan Wolf's moral saints was that they were concerned about morality *de dicto*, there is in fact no conflict between moral sainthood and a desire to do what is right *de dicto*. Moreover, I have even tried to show that in certain respects a *de dicto* desire to do what is right can be an asset. This leaves us with the following nagging question: are moral saints more or less likely than the general population to have a standing *de dicto* desire to do what is right? Or, assuming that all good people have this standing desire to some extent, we can ask: do moral saints tend to exercise their standing *de dicto* desire to do what is right more or less than the general population? If I were to meet a moral saint, what might the role of the *de dicto* desire be in his life?

The answer, I propose, is "it depends." Since there are many ways of being a moral saint, we would expect there to be a wide spectrum spanning the different relative

influences of *de re* and *de dicto* motivation a person might exhibit. Let me offer two paradigm cases of the sort of saints who might fall at either extreme end of the spectrum. While we can stipulate that both of these saints are very smart and both of them are highly active in performing good actions, I will call one the "Cerebral Saint" and the other the "Action Saint."

> **The Cerebral Saint**. This is the sort of saint whose behavior is most influenced by a standing *de dicto* desire to do what is right, because it is right. The cerebral saint still *has* desires *de re* to perform morally right actions, but she has perfected the art of using *de dicto* motivation to act as a regulator, in the ways I described in the previous section above. She does not use the standing *de dicto* desire as a way of being stubborn or obsessive, and its use does not cause her to be alienated from her ends. Rather, it's simply the case that the Cerebral Saint has always had an unusually strong concern for doing the right thing, but she has never completely trusted her direct, organic responses to the objects of moral concern, perhaps because her childhood training in philosophy taught her to be skeptical of these first-order motivations and always to subject them to rigorous scrutiny. While her first-order motivations tend anyway to match with what it is in fact right to do, the abstract notion of rightness is always in the back of her mind. As such, the Cerebral Saint is the consummate researcher. She always consults general principles and empirical data before undertaking an action. If she were working in the field of public health, she would have a reputation like the one Bill Gates has developed: reading voraciously to learn about all the latest research in epidemiology. The Cerebral Saint would have an unusual degree of comfort with the abstract notion of "moral rightness"—she would think about it and talk about it frequently, but without exhibiting any of the signs of moral imposterism I discussed earlier. She might think, for example, that it's very important that people have access to decent healthcare. Indeed, she will have developed this and many other first-order moral motivations, but her attachment to them will be highly flexible and unsentimental, open to change if it turns out they are not morally good ends after all.

The notion of the Cerebral Saint might seem a little incoherent. That's because it's a necessary part of sainthood, we seem to think, that one's desire to do what is right be read mainly *de re*. That is, we want it to be true that, for most or nearly all of the actions that are in fact right, our moral saint desires *directly* to do them. Notice, though,

that this constraint can be fulfilled by the Cerebral Saint while still giving her leeway to be moved by the abstract notion of rightness. On the other hand, the Action Saint finds the abstract notion of rightness virtually unnecessary.

> **Action Saint**. We can take as a model of the Action Saint someone like Dr. Paul Farmer. While he does not completely lack a standing desire *de dicto* to do what is right, his abundant desires *de re* are almost always motivationally efficacious on their own. While Farmer is certainly capable of understanding and manipulating abstract concepts (he is, in his own way, incredibly cerebral), he finds that generally they get in the way of action. It's as if Farmer sees everything in the world through the lenses of *de re* moral concern, and so he is constantly overwhelmed by the people he ought to help, the hospitals he ought to build, the diseases he ought to treat. There is hardly any room left over for their to be "gaps" that require him to fall back on a standing *de dicto* desire to do what is right. Contemplation of whether his actions can be described as "morally right" does not keep him up at night; but worrying about the dying people whom only he can help *does* keep him up at night. Indeed, Farmer seems almost distrustful of anyone too caught up in "morality" *de dicto*, for he knows that such people tend to be concerned mainly about appearances. That's why he often uses the term "do-gooder" as a sarcastic insult: as he sees it, to wear one's moral goodness on one's sleeve is to betray insincerity.

Surely this is something of an oversimplification, and surely the majority of moral saints will fall somewhere between the Cerebral Saint and the Action Saint. Nonetheless, we can see that there is a place for *de dicto* moral motivation in a theory of moral sainthood. This sort of motivation may not strictly speaking be *necessary* in order to live a morally extraordinary life, and certainly this kind of motivation *alone* is both insufficient and dubious. But once we have carefully separated *de dicto* moral motivation from the many varieties of moral imposterism with which it is often associated, we find that it can play a subtle but vital role in the moral psychology of the morally best people.

Rightness, Reasons, and *De Dicto* Motivation

Here are some of things I hope to have established thus far about being motivated to do the right thing *de dicto*: (1) that this sort of motivation is not at odds with moral motivation *de re*; (2) that it can, and in many cases ought to, coexist and interact with moral motivation *de re*; (3) that it is not fetishistic; and (4) that all moral saints need to have this sort of motivation, but they will rely on it to varying degrees, depending on individual differences in the interaction between *de re* and *de dicto* motivations.

In this section I briefly suggest that the importance of a standing *de dicto* moral motivation is made more evident when we look closely at the nature of rightness and wrongness. Of course, the nature of rightness and wrongness is too big of a question to address with any satisfaction here. But I want to at least introduce the idea that our account of rightness will bear on our account of *de dicto* moral motivation, and vice versa.

Consider the following questions: Is the fact that an action is right distinct from the fact that it has certain right-making features? Is the fact that it's right an *additional* reason to do it, or is its rightness redundant with the existing reasons? It should be quite clear why the answers to these questions bear on my argument about moral saints and moral fetishism. If rightness is an *additional* reason-giving property of actions, then it might in many cases be not only permissible but obligatory that a morally good person be motivated by the rightness of an action *de dicto*. Correspondingly, if rightness is *redundant* as a property of actions—that is, if there is nothing more to an action's being right than the totality of the particular reason-giving features the act already has—then that might partly explain the charge of fetishism. Perhaps a moral fetishist is just

someone who gives rightness *de dicto* more attention (and more influence in her deliberation) than any redundant property could possibly deserve.

When we ask whether rightness is a reason-giving property, we are asking whether the correct account of rightness is a "buck-passing" or a "non-buck-passing" account. As R.J. Wallace has put it, buck-passing accounts are "summary accounts" and the concepts these accounts analyze are "summary concepts" (2006, 332, 335). A buck-passing account of *value*, for instance, would show that the property of being valuable does not itself provide reasons, but merely *summarizes* the other reason-giving properties an object has, such as pleasantness. So, for example, according to T.M. Scanlon's buck-passing account of value, "being valuable is not a property that provides us with reasons. Rather, to call something valuable is to say that it has other properties that provide reasons for behaving in certain ways with regard to it" (1998, 96). Suppose, for example, that a discovery "is valuable because it provides a new understanding of how cancer cells develop" (96). The *value* of this discovery does not provide us with a reason, say, to spread the word about the research paper in which the discovery is revealed. Instead, our reasons to spread the word arise from *other* facts about the discovery, like the fact that it could save lives. Scanlon's account of value is a "summary account" insofar as it claims that the property of being valuable does not itself provide reasons ("is not a normatively significant kind or property in its own right"), but rather summarizes the *other* reason-giving properties that an object has, like pleasantness or the potential to cure cancer.

Jonathan Dancy has explained Scanlon's buck-passing account of value this way: there are some natural properties (e.g., pleasantness) that ground reasons, and value lies in that grounding relation. In other words, "being of value is having features that ground

reasons" (165). The value does not add any reasons to the mix; so "the badness of the pain cannot add to the reasons generated by the features that make the pain bad (i.e., effectively, by the way it feels)" (165).

According to buck-passing accounts of *rightness* (or wrongness), the fact that something is right (or wrong) does not provide any additional reason (not) to do it. Consider, for instance, Jonathan Dancy's view of rightness:

> In deciding whether an action is right, we are trying to determine how the balance of reasons lies. Our conclusion may be that there is more reason (or more reason of a certain sort, perhaps) to do it than not to do it, and we express this by saying that it is therefore the right thing to do. The rightness-judgment is verdictive; it expresses our verdict on the question of how the reasons lie. It is incoherent, in this light, to suppose that rightness can add to the reasons on which the judgment is passed, thus, as one might say, increasing the sense in which, or the degree to which, it is true. And the same is true of wrongness (2000, 166).

So long as our concept of rightness is "verdictive" in the way Dancy describes, the buck-passing view about rightness is appealing. After all, it makes sense that in deciding whether an action is right or wrong, we are in a way "adding up" the various reason-giving factors (harm, well-being, rights, promises, etc.). Our final judgment is like a verdict or summary; it presents the correct weighing of the reasons without influencing or contributing to their weights at all. In other words, the buck-passing account draws on a notion of rightness that means something like "on balance, morally choiceworthy," where what we are responding to is the considerations that make it choiceworthy, not its choiceworthiness itself. These specific considerations would be, as Scanlon says, "sufficient in themselves" (2007, 6). The fact that "it would be right" is not only unnecessary but redundant.

But things are surely more complicated. Even if rightness and wrongness judgments just are verdicts, it's not obvious that this precludes them from giving us *additional* reasons. Furthermore, whether our concepts of rightness and wrongness are verdictive in this way (or *merely* verdictive, we might say) is open to debate.

There are more and less plausible versions of the thesis that rightness and wrongness are verdicts. Consider the following two possible versions:

(1) *Verdict as placeholder/signal*. An action's rightness or wrongness is just a placeholder or signal for the fact that the reason-giving considerations add up in a certain way. We use rightness and wrongness as short-cuts or heuristics, as products of a straightforward decision procedure. Consider the following analogy: you take a blood test to check for the presence of certain antibodies that indicate the possibility of disease. If the pathologist sees those antibodies under the microscope, he sends back the test results marked "positive." But this verdict is merely his way of signaling to your doctor the presence of the antibodies. The "positive" result does not provide additional reason to begin treating the condition over and above whatever reason is provided by the existence of the antibodies themselves. The verdict is just a signal.

(2) *Verdict as normative weighting scheme*. The fact that an action is right or wrong involves a weighting of the various first-order reason-giving considerations, and the nature of the weighting is normatively relevant. Rightness is like the spreadsheet formula we use to calculate a student's final grade. The grades on all the various tests and papers, taken individually, give us reasons to rate the student's performance as excellent, mediocre, or poor. But the verdict we get after applying the spreadsheet formula gives us an *additional* reason to approve or disapprove of the student's performance as a whole, because it represents the results of a normatively relevant weighting scheme. The "verdict" in this case is not simply a transparent window through which to view the first-order reason-giving considerations. Rather, the verdict represents a way of interpreting those reason-giving considerations as a whole so as to create a new reason.

The distinction I've just drawn is rough but at the very least I hope it lends some plausibility to the idea that we can have a "verdictive" conception of rightness without being forced to accept that *the fact that an action is right* provides us with *no additional reason* (for an action or attitude) than was present before the judgment was made. If we

think of rightness as the second sort of verdict rather than the first, we see that the way in which the first-order reason-giving properties are combined is itself a reason-generating process. Perhaps this more nuanced verdictive conception no longer qualifies as a "buck-passing" account, since it is a built-in feature of buck-passing as normally defined that there is "no additional reason." But this strikes me as a problem for buck-passers who try to motivate their view by claiming it trades on an intuitively appealing notion of rightness-as-verdict, rather than a problem for nuanced interpretation of what counts as a verdict. In any case, we need not accept a "verdictive" conception of rightness or wrongness.

Consider, for example, Scanlon's view of wrongness, according to which an act is wrong "just in case any principle that permitted it would be one that someone could reasonably reject" (2003, 160). On this view, the "normative basis of right and wrong" lies fundamentally in the idea of "justifiability to others" (160). Clearly, there is more packed into this notion than the simple idea of a verdict. A wrong action is one that we cannot justify to others, one that goes against "what we owe to others," and this fact provides us with an additional reason not to do it, over and above the mere fact that the principle permitting it is a rejectable principle.

Another non-buck-passing account of rightness is Stephen Darwall's (2006). From his "second-personal standpoint," wrongness is fundamentally a matter of what we're responsible or accountable to others for. Again, this is a much richer notion than simply a verdict. On Darwall's account, the fact that an action is wrong provides an additional reason against doing it. We are, and ought to be, motivated by this notion of

accountability when we would not be motivated merely by the presence of certain reason-giving features like harm or pain.

On non-buck-passing accounts like these, rightness is fundamentally a matter of what we're responsible for or what is demanded of us. Construed this way, we can say not only that the rightness of an action provides an additional reason to do it, but that *recognizing* this additional reason and responding to it *de dicto* might be an essential skill of the morally good person and thus the moral saint.

In his more recent work, Scanlon has offered some examples that help to show just what it might mean for rightness or wrongness *itself* to influence our action. There is a gray area between rightness or wrongness constituting an extra reason and constituting no reason at all; in the middle lie cases where the rightness or wrongness "shapes" deliberation. Consider, for example, the following case:

> I have been hired as a guard, by someone who has good reason to believe he is likely to be attacked. While standing guard, I see someone else about to be injured by a thug. I could run from my post and prevent this, but I would be leaving my client exposed to attack. So it might not be wrong for me to refuse to go this person's aid, despite the fact that he will be injured if I do not. (2007, 7)

Scanlon thinks the idea of wrongness influences the guard in this case, but not primarily by way of "providing a new direct reason for a certain course of action" (7). Instead, wrongness "shap[es] the way I should think about the decision I face, and [determines] which other considerations I should take to be reasons" (7). Complementing this "shaping" role of wrongness is what Scanlon had earlier called a "backstop" role: in cases where we are tempted to do something we judge to be wrong, we attend to its wrongness and ask ourselves "How much weight should I give to the fact that doing this would be wrong?" (1998, 157).

In both the "backstop" role and the "shaping" role, wrongness needs to "provide reasons (or invoke them)" in response to the same question: "Why take the results of thinking about what to do in the way morality prescribes as authoritative and conclusive?" (2007, 10). It is because we can ask this important question that buck-passing accounts of rightness are implausible. On a buck-passing account, the fact that an action constitutes breaking a promise might seem to be the only morally relevant reason-giving feature. But Scanlon points out that there are higher-order questions we often need to ask about our reasons, like "Why should these reasons include the fact that one made a promise but exclude the fun of breaking it?" (2007, 9). In order to know whether to avoid a certain action, we need to know whether it is wrong, and in order to know whether it is wrong, we need to attend to subtle weightings of the lower-order reasons against and in favor of it.

But it's not that the wrongness of an action *just is* its lower-order properties. Rather, the wrongness arises out of the way those reason-giving properties add up. Scanlon seems to think the role wrongness is playing here falls somewhere in between buck-passing and reason-providing. He calls it a "reason-referring" property (2007, 10 note 7). He later explains that while, according to his account, the property of moral wrongness is not *itself* reason-providing, he *is* positing a higher-order reason-providing property, and it's a property of "one way of having [the property of moral wrongness]"—namely, via justifiability to others on grounds they could not reasonably reject (2007, 17).

We can now begin to see parallels between the "shaping" role of rightness (or wrongness) and the various ways in which I earlier claimed *de dicto* moral motivation can interact with *de re* moral motivation. Recall that I considered four ways *de dicto*

motivation might influence our deliberation: higher-order moral reasoning, normative governance, motivational impetus, and epistemological stopgap. The first three, in particular, draw on notions of rightness itself being, if not "reason-providing," at least "reason-referring." The "higher-order moral reasoning" role was meant to explain situations in which a *de dicto* motivation enhances a moral judgment, by adding a layer of reflectiveness to what would otherwise be a crude (even if correct) *de re* judgment. If Scanlon's argument about rightness is plausible, then these higher-order reasoning cases are cases where the agent responds to the rightness (or wrongness) of an act over-and-above its right-making features.

The "normative governance" role of *de dicto* motivation was meant to describe cases in which the agent needs to consult a judgment of rightness because the *de re* considerations conflict. This seems to be what Scanlon referred to as the "backstop" role of rightness, though the backstop cases have the added feature that the agent is otherwise motivated to do the action that is in fact wrong. Finally, the "motivational impetus" role of *de dicto* motivation was meant to explain cases in which a standing general desire to do the right thing acts as a "patch" to fill the gaps in *de re* motivation that occur when we have a change in moral motivation. This, again, seems to be analogous to Scanlon's "backstop" role for wrongness.

If Scanlon is correct and rightness and wrongness are not mere "verdicts" but rather reason-shaping or reason-referring properties that represent an important form of higher-order moral reasoning, then we should expect moral saints to be interested not just in the right-making features of acts, but in the rightness itself. That is, we should expect moral saints to have *de dicto* moral motivation. However, that doesn't mean moral saints

need to be motivated in the way Susan Wolf describes. Her picture of *de dicto* moral motivation is flawed in two ways. First, it posits *de dicto* moral motivation *at the expense* of *de re* motivation, when in fact the two are complementary. And second, it conflates a *concern* for the rightness of one's action with an *obsession* with the rightness of one's action. Both of these features cause Wolf's moral saints to appear irritating, fanatical, or even pathological. Indeed, her saints seem to make a fetish of morality. Fortunately, we need not worry that real moral saints would look like this, since there is no reason to build a monolithic obsession with rightness *de dicto* into our account of moral sainthood, at the expense of the corresponding *de re* desires. There are moral imposters and moral fetishists out there, and sometimes the best way to find them is to look for the person obsessed with "rightness." But it's not the concern for rightness itself that's to blame for their behavior.

CHAPTER 4

SACRIFICE AND MORAL OBLIGATION

The principal aim of this chapter is to isolate and illuminate a notion of *sacrifice* that is relevant to a discussion of moral obligation. The aim in the next chapter is to argue that the behavior of moral saints plays a unique and important role in actually increasing our moral obligations—that is, in transforming actions that would otherwise be supererogatory into obligations. The two aims are related, of course. We commonly think that we cannot be obligated to perform a given action if it involves an unreasonable degree of sacrifice. In this chapter, I give an account of sacrifice that proves useful in determining when a sacrifice is unreasonable. In Chapter 5, I argue that by taking on lives of great sacrifice, moral saints provide the rest of us with evidence about what is and is not possible, and that this evidence has the effect of actually *altering* our moral obligations. What we can reasonably expect others to do depends partly on their justified beliefs—beliefs about what is physically possible, what is emotionally possible, and what degree of sacrifice is entailed by certain actions and lifestyles. If an agent justifiably believes that a given action is too burdensome, then it would be unreasonable if morality required her to perform that action. But exposure to moral saints may cause us to see certain lifestyles as less burdensome than we thought, thus updating our stock of justified

beliefs, and ultimately changing what we are obligated to do. I call this the "Ratcheting-Up Effect."

As discussed in earlier chapters, by "moral saint" I don't mean a moral perfectionist or an idealization of a flawless moral agent. Nor am I referring to a person who commits a one-time act of heroism, though some philosophers have called such persons moral saints. Instead, as before, I'll use the term "moral saint" to refer to an extraordinary and rare person who, over the course of a lifetime or a significant portion thereof, repeatedly behaves (and indeed, cultivates habits of behavior and perhaps certain deep commitments) in such a way that is, intuitively, beyond the limits of what morality requires, often at great personal sacrifice. I will offer two case studies of people I take to be real-life moral saints.

A Puzzle about Agents and Observers

The importance of sacrifice can be illustrated by appealing to a certain puzzle. It is often the case that someone who roughly meets the definition of moral saint I gave above, someone whose actions are intuitively good but not required, will protest, "I was just doing my duty," or "I would have felt guilty if I hadn't done it," or "I'm not special. Anyone else in my position would have done the same thing."[21] One real-life example of

[21] A related, but distinct, phenomenon is described by Gregory Trianosky (1986): "Sometimes we are challenged to perform acts that are good to do but not required, by individuals who plainly are already committed to performing them" (27). Trianosky is interested in the fact that we tend to provide *excuses* when we decline requests to do supererogatory things like join the Peace Corps, yet excuses are commonly thought only to be appropriate when we have omitted an *obligatory* act. See also Susan Hale, "Against Supererogation" (1991). Hale claims that not only do we give excuses for failing to perform supererogatory actions, but we also distinguish between appropriate and inappropriate excuses for the omissions.

this phenomenon is John Weidner, an agent for the Netherlands Security Service who worked for the Resistance, helping Jews escape during World War II. He put himself in grave danger, endured severe beatings, and narrowly escaped execution after being captured by the Gestapo. When asked whether what he did was "an extraordinarily good deed," John replied "No. Absolutely not. I did my duty. That is all."[22] Similar statements by other Holocaust rescuers have been extensively documented.[23] These statements have an odd ring to them. We *expect* a firefighter who runs into a burning building to say "I was only doing my job"—indeed, such statements are now a cliché of media coverage of fires and other disasters, and they are unremarkable because they are true. After all, it *is* his job, even if he only says so out of false modesty in an attempt to elicit praise. But when an *ordinary* person runs into a burning building, the statement "I was only doing my duty" takes on a different meaning. In some cases, of course, the ordinary person is being insincere. Perhaps she does not really believe that rushing into the burning building was a duty, but says so out of modesty—she does not want to be, or to appear to be, self-congratulatory. Or perhaps she is not genuinely modest but feigns modesty—she seeks the praise and attention that tends to follow displays of modesty after extraordinary accomplishments. So, interestingly, both modesty and *false modesty* could cause a person

---

[22] Monroe (2003), p. 117.

[23] The phenomenon of moral saints refusing to acknowledge that their actions are supererogatory is also discussed by Urmson (1958) and Hale (1991). For more examples, see Fogelman (1994), Gilbert (2003), and Gies (1987). Miep Gies, who helped hide Anne Frank's family, said in her memoir, "I am not a hero…I was only willing to do what was asked of me and what seemed necessary at the time" (11). Another example is Paul Farmer, whose case is discussed below. In one scene in Kidder's book about Farmer, a patient calls Farmer a "saint," something that evidently happens to him quite often. "I don't care how often people say, 'You're a saint.'" Farmer says, "It's not that I mind it. It's that it's inaccurate. People call me a saint and I think, I have to work harder. Because a saint would be a great thing to be" (16).

to utter, "I was only doing my duty," when she does not in fact believe it. But these are not the cases I'm interested in. Rather, I'm interested in the cases where the moral saint genuinely believes it to be the case that her actions are required, but outside observers come to the opposite conclusion.

In his important article "Saints and Heroes" (1958), J.O. Urmson acknowledges these cases of peculiarly modest agents. "I have no desire" he writes, "to present the act of heroism as one that is naturally regarded as optional by the hero, as something he might or might not do; I concede that he might regard himself as being obliged to act as he does" (203). Nevertheless, even though this hero[24] *regards himself* as obliged, Urmson claims that on some level he knows he is not:

> But if he were to survive [smothering a grenade] only a modesty so excessive as to appear false could make him say, "I only did my duty," for we know, *and he knows*, that he has done more than duty requires. Further, though he might say to himself that so to act was a duty, he could not say so even beforehand to anyone else, and no one else could ever say it. Subjectively, we may say, at the time of action, the deed presented itself as a duty, but it was not a duty (203, emphasis added).

[24] Urmson distinguishes saintly actions from heroic actions (and saints from heroes) in the following way: saintly actions involve "resistance to desire and self-interest" and heroic actions involve "resistance to fear and self-preservation" (200). But this distinction doesn't play much of a role in his philosophical arguments, and he seems to think that a single action is sufficient to make someone a saint just as it would be to make them a hero. In this paper I use a conception of moral sainthood that sets a higher bar: moral saints don't just commit a single extraordinary deed, but display a pattern of behavior over a significant period of time. However, though Urmson and I are discussing slightly different sets of agents—he mixes heroes and saints, and he counts one-off actions—I think his discussion about whether these agents believe their actions to be obligatory is still relevant. Someone who habitually performs supererogatory actions could have the same sort of attitudes about her behavior that Urmson describes in the agent who performs just one supererogatory action, and I have cited evidence that this is in fact the case with some real-life moral saints.

Urmson draws our attention to a useful distinction in perspective. On the one hand, we have the agent's subjective attitude toward the action—how he *regards* it, how it *presents itself* to him, what he *says to himself* about it. On the other hand, we have the fact of the matter about whether the action was obligatory, a fact about which the agent can be mistaken. Surely observers can also be mistaken about whether the action was obligatory, but Urmson seems to think that they at least have a more impartial perspective, and that in the case of grenade-smothering, they correctly mark the action as supererogatory.

It might initially be puzzling how Urmson can claim that the soldier both "regards" the action as obligatory and "knows" that it is not. The simplest and least satisfying interpretation is simply that the soldier believes *both* that it is obligatory and that it is not—in other words, he has two utterly inconsistent beliefs, as people sometimes do. Another possibility is that the soldier "regards" the action as obligatory in some more emotionally-laden sense (he *feels* obligated), as opposed to the more rational sense in which he believes it not to be. Along these lines, we might say that in the "heat of the moment" the soldier regards the action as obligatory, but in a cooler, more reflective state of mind he would admit that it is not. This aligns well with Urmson's claim that the soldier could not tell *others* that the action was obligatory (even beforehand), because presumably these bystanders, not themselves forced to make the decision of whether to risk their lives, are in a relatively "cooler" state of mind, and thus would not accept him making such a claim.

Now, Urmson's soldier case is unique in that the soldier cannot survive the heroic action, and thus cannot evaluate it after the fact. Urmson seems to think that if the soldier could survive, he would admit he was a hero. The cases I'm interested in are those where

the agent maintains that she is required to act—either upon reflection afterward, or beforehand if she has time to deliberate—and yet observers disagree. Below, I'll suggest that the problem lies in the relationship between obligation and sacrifice. It's not that there is a double-standard of obligation—one that the agent applies to herself, and one for everyone else—but rather that we have asymmetrical access to facts about the *sacrifices* involved in discharging various obligations.

First, though, it's important to point out that not *all* cases of peculiarly humble saints and heroes are ground for puzzlement. Some of the actions of a moral saint may well be obligatory as she insists they are. Indeed, many actions commonly regarded as supererogatory may in fact be obligatory. Moreover, just because an agent says she feels guilty, or anticipates feeling guilty, when she considers *not* performing some extraordinary action—say, a dangerous rescue—this does not mean that she is morally required to perform the action. Indeed, it does not even mean that she *believes* she is morally required to do it.[25] After all, guilt can be recalcitrant. Even in a case where an agent judges an action to be neither obligatory nor supererogatory, but merely morally indifferent, she may still feel guilty if she fails to act. Moral saints in particular might have an overly sensitive guilt-response, something which serves them well most of the

[25] As Elizabeth Anderson has pointed out to me, we might interpret Darwall's "Second-Person Standpoint" as generating a kind of indirect "warrant" for guilt when one omits a supererogatory action. If an action is obligatory and there is someone who can hold me second-personally accountable for it, then it seems I ought to feel guilty if I omit that action. With a supererogatory action, it seems that no one has standing to hold me accountable for an omission, since the action is by definition not obligatory. But what if I can hold *myself* accountable—what if I can occupy the second-person standpoint with respect to myself? Is this what is going on when I "hold myself to a higher standard" or when I am obligated *in light of my ideals* (see Little 2007) even when others would not be? The problem, it seems, is that the mere fact that I could or do hold myself accountable does not seem to entail that I am obligated.

time but occasionally misfires. Indeed, we ought to be less interested in reports of guilt or anticipated guilt than we are in *judgments* about whether acts are required or not, since clearly the two can come apart. The cases of interest to us here are those in which the agent sincerely judges her action to be obligatory, yet observers do not. An agent's feelings of guilt or anticipated guilt may sometimes gives us a clue as to what she judges, but in general these feelings cannot be expected to reliably track her judgments, or to track the actual rightness or wrongness of the action, and so are of only secondary interest.

Thus there are many ways to interpret reports like "I'd feel guilty if I didn't do it," and "I was only doing my duty," and there is no reason to think one interpretation applies in all or even most such cases. Yet it still seems that the disparity between how agents and observers classify an extraordinary action is puzzling and perhaps symptomatic of some deeper tension. Let us call this disparity the *puzzling data*:

Puzzling Data: Moral saints habitually perform actions that are, intuitively, beyond moral obligation. Yet (some of) the moral saints consider (some of) these actions to be obligatory, not supererogatory. There is a *persistent agent-observer disparity*.

How might we resolve this puzzle? One the one hand, these moral saints (and other occasional heroes and rescuers) might simply be mistaken: their actions are in fact not required, and morality demands less of them than they thought. On the other hand, the observers might simply be mistaken, and morality requires more of us that we thought, including sometimes putting our lives in grave danger. As I will argue in what follows, perhaps things are more complicated than these two possible explanations would allow. While it may be true that the saints are mistaken about what morality demands of them, they are not *simply* mistaken. They have a perspective on the action that is not readily

available to observers, and we should take their perspective seriously. One of the reasons for the agent-observer disparity is that agents and observers have asymmetric access to facts about *sacrifice,* and facts about sacrifice influence the moral status of actions.

The boundary between the obligatory and the supererogatory is sensitive to facts about sacrifice because, just as it is implausible that something one *could not do* could be morally required ("obligated implies can"[26]), it is also implausible that something one *could not do without certain sacrifices* could be morally required if those sacrifices are significant relative to what is at stake (obligated implies no unreasonable sacrifice). In what follows I aim to shed light on this sacrifice constraint by offering an account of what sorts of sacrifice rise to the level of generating a legitimate claim—that is, a claim to be absolved of a given moral obligation. For now, it seems sufficient to say that the

[26] I'm avoiding the phrase "ought implies can" here because "ought" is frequently used with a broad meaning that is ambiguous between an "ought" of obligation and an "ought" of supererogation. For instance, some might think it makes sense to say that you ought, morally, to perform an act even if it is not obligatory or required. Since I'm interested in precisely the distinction between obligation and supererogation, I avoid the term "ought" for the sake of clarity. It's also worth noting that the "can" in "ought implies can" might be weaker than the "can" in "obligated implies can". As Stephen Darwall (2006) argues, there is a weak sense of "ought implies can" found in Kant, where to say that an agent "can" do something is just to say that it is "an open deliberative alternative, that is, something such that one's abilities and opportunities with respect to it do not preclude intelligible consideration of whether to do it" (240). This weak sense of "can" is consistent with the agent not even knowing or being able to know that she ought to perform the action. However, if we are interested in *obligation* as something that members of a moral community can authoritatively demand of one another, and hold each other responsible for, then we may need a stronger sense of "can." According to Darwall, the very idea of blaming someone or holding her responsible requires that "we presuppose that she must have been in a position to know [what she should have done]" (241). In other words, for us to hold someone accountable for an action, not only must the action be an "open deliberative alternative," but also there must be a "process of reasoning" by which the agent could come to see herself as obligated. In a later section, I argue for what is perhaps an even *stronger* sense of "can," in the form of a "knowledge constraint" on obligation: if an action is obligatory, then it must be the case that it would be reasonable to believe (or a reasonable person would believe) that it is obligatory.

following rough principle holds: if a given action is a candidate for moral obligation, but the action *cannot be performed without giving up X*, then, as the significance of X increases relative to what is at stake, the seriousness of the obligation decreases. There is some level of X at which the obligation is completely defeated, and the action is supererogatory.

Agents and observers sometimes assess actions differently because different elements of sacrifice are salient to different judges. For example, an aid worker might view lost time with her family as insignificant in light of the lifesaving work that she can do with that extra time, and so from her perspective a certain busy lifestyle might seem obligatory. To the aid worker, the *felt* aspect of the sacrifice might be most salient, and on a day-to-day basis she might not *feel* all that bothered by it, if only because she is too busy to notice. But an observer might conclude that the aid worker's loss of family time is more significant than she herself perceived, and that in fact her lifestyle cannot possibly be obligatory.[27] The outside observer might take less notice of the aid worker's *felt* loss and more notice of the importance of family connections in one's life. What underlies this agent-observer disparity, it seems, is the notion that there is more to a loss than how it *seems* from the point of view of the person incurring that loss.

If agents and observers make different judgments about sacrifice, and if we cannot be obligated to make unreasonable sacrifices, then it is no surprise that agents and observers sometimes make different judgments about moral obligation. Suppose, for example, that we have an underlying moral obligation to save other human beings from

[27] No doubt the opposite asymmetry is just as common—an agent thinks it is too much of a sacrifice to give up X, and concludes the action cannot possibly be required, but observers see things differently. But it's the first asymmetry which seems to arise more frequently with moral saints.

an unjust death, unless doing so requires an unreasonable sacrifice. Agents and observers may disagree on whether a sacrifice is *unreasonable* because they occupy fundamentally different vantage points on a situation that is complicated and dramatic, especially if it involves the prospect of unjust death. Miep Gies, who denies that she is a hero for helping to hide Anne Frank's family, may have considered her own sacrifices to be entirely reasonable in light of the grave danger faced by the Frank family. It's true that she risked imprisonment and even death by helping them. But on a daily basis, the mundane sacrifices—a bit less to eat, a bit less sleep, a bit more secrecy—no doubt seemed entirely bearable compared to what the Frank family was going through. One reason Miep Gies denies having done a remarkable deed is perhaps that her sacrifices did not seem all that *bothersome* to her. Outside observers, however, can examine not just the *bothersomeness* of her sacrifices but the real value of the safety and security Gies was giving up. In other words, the outside perspective can account for losses that were bad for Gies even if she did not notice them or value them highly. As I will argue, there is more to sacrifice than just *felt sacrifice*. On this view, we can make sense of cases like that of Miep Gies, where the agent and observers disagree about whether a sacrifice was unreasonable and whether an action was obligatory.

In order to understand how this agent-observer asymmetry could arise, we need a better understanding of just what sacrifice *is*. Is *feeling* something *as* a sacrifice enough to make it so? Surely not, for then agents who give up outlandish luxuries could claim these losses as *sacrifices* so long as giving up those luxuries caused them distress or seemed like a big deal. Do certain losses count as sacrifices regardless of whether the agent feels them as such? If so, then we must be willing to accept that an agent need not

be *aware* of a given cost for that cost to count as a sacrifice. Agents, it appears, have access to the first phenomenon; for now, I will call this *felt sacrifice,* or the *subjective* component of sacrifice. Observers, on the other hand, have only limited and indirect access to the subjective component of sacrifice, but quite a bit of access to certain objective facts, such as whether giving something up will affect the agent's wellbeing in ways that even she cannot predict or may not care about. In what follows I argue that the account of sacrifice relevant to arguments about moral obligation is an objective account, one that defines sacrifice in terms of losses of wellbeing, where wellbeing is understood in agent-neutral terms. The notion of "felt sacrifice" plays only an indirect role in this account; that someone *feels* a loss *as* a sacrifice is a good indicator that something important has been given up, but it itself is neither necessary nor sufficient for a loss to be a genuine sacrifice.

Sacrifice and Moral Obligation

In everyday thought and language, we use the term "sacrifice" rather flexibly, perhaps even sloppily. We tend to conflate its subjective and objective meanings. Our task here is not to give an analysis of a generic broad concept that underlies each and every use of this tricky term, but rather to isolate a particular, coherent concept of sacrifice that is properly suited to thinking about moral obligation. We begin with the basic thought that moral obligation seems to be a matter of what we can legitimately demand of one another. Just as there is a limit to what we can demand of others morally, there is a limit to which of our own needs, wants, and preferences we can claim to count as defeaters of our own moral obligations. For instance, just as you cannot demand that I risk my life in order to save you from losing a tooth, *I* cannot demand that you take my

love of milkshakes so seriously as to release me from the obligation to save your life when doing so would require giving up my milkshake. Thus cooperation, reciprocity, and fairness seem to point us toward a view of sacrifice that is closer to the objective pole than the subjective. In order to be able to make moral demands that others will take seriously—demands that will sometimes involve asking others to give up things that matter to them—I need to be willing to limit the things that *I* put forth as defeaters of my *own* obligations to just those things whose importance I can justify to others. As mutually accountable agents in a moral community, we make demands on one another, and these demands must be both public and publicly justified.

In "Preference and Urgency" (1975), T.M. Scanlon argues for a way of understanding benefits and sacrifices that incorporates justifiability to others. He argues that the "urgency" of moral claims is best understood against the background of an objective criterion of wellbeing. Whereas a subjective criterion entails evaluating well-being from the "point of view of the person's tastes and interests" (656), an objective criterion requires evaluation that is independent of those tastes and interests, so that "an appraisal could be correct even though it conflicted with the preferences of the individual in question, not only as he believes they are but even as they would be if rendered consistent, corrected for factual errors, etc." (658). The objective view can make room for much of what is intuitively appealing about the subjective view. It can, for example, account for the importance of individual tastes and interests by declaring it objectively important that social institutions be set up in such a way as to allow people to develop such tastes and interests (658). Unlike the subjective view, though, the objective view does not take the strength of those tastes and interests as the *primary* locus of well-being;

after all, it seems that in evaluating what is good for a person there needs to be room not just for what she *does* take an interest in, but what she *should* take an interest in.

The objective view also does a better job than the subjective view at capturing the fact that a criterion of well-being which is going to serve as the basis for claims people make against each other must represent a kind of "consensus" about what matters. A subjective view can represent a consensus only in a weak sense: the consensus that everyone's individual tastes are valued and should count equally, whatever they happen to be (657). The objective view, however, seems to capture a more robust type of consensus: it requires that individual tastes and preferences be at least somewhat *intelligible* to others, and it accords them value *in virtue of* that intelligibility. Thus, when comparing the "urgency" of two person's interests,

> [W]hat we do is not compare how strongly the people in question feel about these interests (as determined, perhaps by what they would be willing to sacrifice to get their way) but rather inquire into the reasons for which those benefits are considered desirable. Even if the goods in question are quite foreign to us and of no value in our society *we can understand why they are of value to someone else if we can bring the reasons for their desirability under familiar general categories* (660, emphasis added).

The sorts of categories that Scanlon has in mind are things like "material comfort, status, or security" and "health or protection against injury" (661). (This is by no means an exhaustive list, of course.) Insofar as we can interpret a given preference or interest as falling under one of the general categories, we can understand it, and this seems to make it more viable as a sort of *currency* in interpersonal exchanges. For instance, you might think of a rescue situation as a sort of moral exchange in which we try to determine whether what's at stake for you (say, your life) is worth more, or is more compelling as a

source of a claim, than whatever I would have to give up in order to save you. I don't mean to imply that these considerations would appear in conscious deliberation at the time, only that these considerations seem to bear on our evaluation of whether the rescue was obligatory or supererogatory.

Following Scanlon and similar veins found in Rawls's notion of "public reason," I think it makes sense to impose what we can call an "intelligibility constraint" on the sorts of losses that could count as a sacrifice. The harder it is for the rest of us to find your attachment to miniature toy soldiers *intelligible* as a member of one of the "familiar general categories," the less likely we are to consider it to be urgent enough to rise to the level of generating claims. By imposing such a constraint, we don't automatically discount certain attachments just because they fail to fit neatly into one of the general categories that emerge in consensus. Rather, in such cases we seek more information—we seek to make the attachment intelligible in terms we can understand. Thus *any* attachment is *potentially* the source of a claim. This view is objective in the sense that all potential sacrifices must meet a criterion of adequacy (namely, intelligibility as a member of a consensus category) that is independent of the person's interests and preferences. Nothing counts as a sacrifice *simply in virtue* of its being valued; but the fact that it *is* valued gives us reason to step back and examine *why* it is valued.

The intelligibility constraint operates on two levels: it first requires that we be able to understand a loss as falling under a general category, and it then requires that we be able to understand the significance of the loss as being *proportional* to the real value of what is given up. So, even if we can understand your attachment to toy soldiers as falling under a consensus category—say, "hobbies" or "leisure" or more generally

"mental health"—we need also to understand the *strength* of your attachment. If you care more about your toy soldiers than you do about feeding and clothing yourself, we would likely find that unintelligible. This is because even amongst Scanlon's urgency categories, some categories are more urgent than others, and some items are more urgent than others even within the same category. Meeting a minimal standard of urgency is necessary, but not sufficient, for intelligibility. If an agent is treating something as more urgent than it is, we will have trouble understanding the magnitude of the loss.

Sacrifice as Normative Rather Than Descriptive

Notice that this conception of sacrifice, in addition to being objective in the sense I have outlined so far, is inherently *normative*. We cannot determine whether something is a sacrifice by simply reading off certain descriptive facts about the situation, like the fact that the person *takes* it to be a sacrifice. In everyday talk, we sometimes use the term "sacrifice" loosely, in a purely descriptive sense, to refer to the giving up of something that one in fact values, whether or not we think it ought to be valued. Suppose Mary likes having at least one dozen pencils on her desk at all times, even though she only ever uses one. We might say "Mary sacrificed her pencils in order to make room for a desk lamp." We are acknowledging that Mary gave up something she valued without necessarily endorsing valuing it. Notice, however, that the less we can make sense of someone's valuing a given thing, the more likely it is that we're being sarcastic when we say that losing it was a "*sacrifice*". This is often the case when someone appears to *overvalue* something. If Mary has six televisions and is upset about having to move to a house with room for only five, we would say "Ooh, a mere *five* televisions, what a *sacrifice*."

Sarcasm is appropriate in such cases because, when we are not speaking loosely, we find that the concept of sacrifice is not descriptive but normative. We don't think downgrading to five televisions is a *real* sacrifice because we don't think anything important or urgent has been given up. It is not that we find having a sixth television to be *intrinsically* trivial; it's just that we have a very hard time understanding how Mary could be so attached to it as be distressed by its loss. The mere fact that she *is* distressed is not sufficient to convince us that a legitimate claim exists. We would need a better story—such as that Mary is a political consultant who must watch 6 channels at once—before her loss would be intelligible to us. Of course, if her distress reaches a certain level, it might generate a claim indirectly. For example, if the stress of losing her sixth TV makes her physically ill, then her loss seems to enter into a general category like "health" that we can make sense of with little effort. Nevertheless, this sort of claim seems only to have force when she is *in the grip* of the distress. Insofar as we ought to make accommodations for her, it is on account of the illness and not the lack of televisions. Once the illness subsides, it seems we would have trouble making sense of the strength of her attachment as generating a general claim unless she had a *very* compelling story about the significance to her life of having six televisions. Without such a story, the appropriate response would be to wean Mary from her attachment to six televisions, rather than indulge it.

Sacrifice and Odd Cases

Notice that this view seems to yield a particularly reasonable and humane way of dealing with certain fringe cases. Some people are simply attached to very odd things, or

attached to things seemingly far out of proportion to their worth. On a purely subjective view of sacrifice, their mere attachment to those things would be sufficient to generate claims, and the rest of us would feel, in a sense, like suckers. Why does Mary's severe aversion to the taste of adhesive glue relieve her of her obligation to put stamps on her letters? Why should *she* get free mail? Why does John's unusual attachment to his Italian shoes relieve him of his obligation to rescue the drowning child in the muddy puddle? If sacrifice were purely a subjective notion, and if John valued his shoes more than the life of a small child, then it would seem that giving up his shoes could be an *unreasonable* sacrifice for John. It would follow that he was not obligated to muddy his shoes and save the drowning child. Clearly, this is unacceptable. On the other hand, an objective view that simply decreed certain things not worthy of value would have trouble with cases at the margins as well. For even though we initially have a hard time understanding how someone could have as strong an attachment to, say, a collection of toy soldiers, as other people have to their health, we still would have some sympathy for the toy collector if he were asked to give it up and this caused him great distress. We would want to at least *try* to make sense of his loss. Given the rich variety of human interests, he would seem to deserve at least the benefit of the doubt. An objective view with an intelligibility constraint allows us to do just that.

Consider, for example, a vivid case in which someone values something far out of proportion to its apparent worth. In the movie *Rain Man*, Dustin Hoffman's character Raymond likes to see certain television shows at certain times. Raymond is autistic and entrenched in his routines; if he cannot see "The People's Court" or "Wheel of Fortune", he gets very, very upset, so much so that he might even be in danger of hurting himself.

Surely if one were to forcibly take Raymond's television away it would be considered inhumane. And if Raymond were to give up his television—even for something else worthwhile—it would clearly be a sacrifice. While we may not understand how someone could be so attached to a TV under ordinary circumstances, a closer examination of Raymond's life would reveal that his attachment could be interpreted as falling under certain general categories that the rest of us can make sense of. He likes order, regularity, ritual, and predictable forms of stimulation. Without them, he feels insecure, perhaps even *unsafe* in a sense. We all understand the need to feel safe, even independently of the need to *be* safe. With a little effort we could all come to understand the legitimate connection between Raymond's TV and his sense of safety. Thus even an objective view could make room for a case in which something that could not be said to be terribly important in general, to *most* people, might nevertheless play an important role in a particular person's life. The same principle applies in cases where something that *would* play an important role in *most* people's lives does *not* play such a role in a particular person's life on account of his unusual personality characteristics. For example, while Raymond has a special attachment to his television shows[28], he also appears to have a striking *lack* of interest in romantic relationships, at least as they are commonly understood. While we might think it would be good for him if he could form such relationships, the fact that he cannot or will not seems to indicate that the lack of such

[28] I don't mean to claim that Raymond is in any way *essentially* attached to his television shows. He seems more attached to the predictability and regularity that the shows offer than the shows themselves, and I imagine that over time, this attachment could be transferred to a different object or pastime. In the short-term, however, it seems clear that to give up his shows would be a real sacrifice, one that could not be compensated for by simply offering him a new hobby.

relationships is not a *sacrifice* for him, or at least certainly not as much as it would be for some other people.

Thus far, we have examined how a thing's being *worth valuing*, in the sense of being intelligibly a member of certain general categories, is a necessary condition for the loss of that thing to count as a sacrifice in the sense relevant to moral obligation. In fact, I will argue that a thing's being *worth valuing* is not just necessary but *sufficient* for sacrifice. On this view, there is no further necessary condition; in particular, it is not strictly necessary that the loss be *felt as a sacrifice*, in the subjective sense. The subjective feeling of sacrifice cannot be *necessary* because, as I will argue later, the most plausible account of sacrifice must allow for cases in which a person is unaware or unconcerned about the costs she is incurring, despite the fact that these costs can legitimately be said to be sacrifices. Nevertheless, felt sacrifice generally tracks objective sacrifice rather well, so when we encounter a case of severe felt sacrifice we can take it as a reason to suspect that something urgent has been lost. Still, much more needs to be said about sacrifice and how it relates to moral obligation. But before starting that discussion it's worthwhile to take a look at a few case studies in order to steer our thinking back toward moral saints in particular. It helps to illustrate these cases in a bit of detail so that we can see exactly what it looks like when a person makes significant sacrifices to engage in a morally good project.

Case Study: Paul Farmer

Recall that Paul Farmer is doctor who founded Partners in Health, an organization that runs medical clinics serving the world's poorest and sickest people. Farmer's clinics

treat tuberculosis and HIV regardless of the patients' ability to pay, and they do it for a fraction of what it costs in developed countries. What began as a single clinic in Haiti is now a worldwide operation with an annual budget in the tens of millions of dollars. Farmer himself treats countless patients directly, sometimes hiking for hours to make house calls. In the early years of the organization, Farmer alternated between living in a hut next to his clinic in Haiti and sleeping in the basement of his office in Boston. He had his entire salary deposited directly into the organization's budget, leaving the office staff to make sure his modest bills were paid. He didn't buy new clothes or take vacations, and he went for long periods without seeing his wife and daughter. Despite all this, Paul Farmer maintains an upbeat disposition marked by a quirky, dark sense of humor, and he constantly feels that he should be doing more.[29]

Yet, while he bears what appears to be a grave and exhausting moral burden, Paul Farmer is an upbeat, engaging person with an infectious sense of humor. He speaks in a bizarre idiolect of catchphrases and acronyms; he makes sarcastic remarks about his work and his patients that can be simultaneously "politically incorrect" and loving. People seem to gravitate toward him, taking up his cause as their own. To be sure, there is an adventurousness, even a sort of backward glamour, in his lifestyle—the jet-setting international voyages, the allure of seemingly insoluble problems, the acclaim of your peers, the attention of world leaders and dignitaries, the superhero status your patients bestow on you, the joy of success.[30] This is not a self-sacrificial life, you might think, but a rather attractive alternative to the monotony of a typical day job. I think, however, that

[29] See Kidder (2003) for More on Paul Farmer and Partners in Health.

[30] I thank Jamie Tappenden for pointing out this way of looking at Farmer's life.

this interpretation is mistaken. It is one thing to see beauty amid extreme poverty and to get joy from helping the sick, but it is quite another to see *glamour* in a job that requires expending vast amounts of physical and psychological energy to make gradual improvements in intrinsically depressing situations—a job that would not exist if not for a horrifying backdrop of suffering and injustice. I doubt that Paul Farmer would agree that his work is glamorous in any sense, even a backward sense. What is perhaps more important, though, is that even if Farmer's lifestyle produces a net positive benefit for him (in wellbeing, satisfaction, glamour, or whatever), *the sacrifices he makes are not thereby diminished*. This is a point I will return to later: the existence of counterbalancing benefits does not diminish the force of sacrifices.

Case Study: Susan Tom

Susan Tom is a California woman featured in the 2003 HBO documentary "My Flesh and Blood." She is the single mother of fourteen children, two naturally born with her ex-husband and twelve she adopted. All of her adopted children suffer from some degree of illness or disability: one is mentally retarded from fetal alcohol syndrome; one is wheelchair-bound from spina bifida; two were born without legs; one is horribly scarred from a crib fire in her infancy that resulted in her birth parents being sent to prison; another has never been able to bend her knee or elbow joints.

Susan never planned to have such a big family. She adopted one child at a time, saying no to several children recommended to her by social workers, and picking only those she thought would fit well in the household. Eventually, she realized that since she was already caring for five or six children, it would not be too difficult to accommodate

five or six more. (When you already require more grocery carts than a single person can push, the addition of a third or fourth cart no longer seems like a big deal.) The film reveals Susan's unrelenting down-to-earth compassion, especially as she cares for the two of her children who are terminally ill: Joe, an emotionally disturbed fifteen year-old with cystic fibrosis; and Anthony, a twenty year-old with Epidermolysis Bullosa (EB), a rare and fatal disease that causes one's skin to fall off with the slightest trauma, creating horrible sores and infections.[31]

It is clear from the film that Susan Tom gets immense satisfaction from the companionship of her children, and that they serve to relieve a deep loneliness that has plagued her since the breakup of her marriage and possibly since she was a child. Nonetheless, her devotion to them is otherwise remarkably selfless. She has no life savings, no retirement account, no time to work outside the home, and no time for a personal life. (The income for household expenses comes from California's "Adoption Assistance Program.") More significantly, the day-to-day care of the children requires stunning feats of physical and emotional strength, especially the care of Anthony. People who suffer from EB need to undergo a cumbersome ritual that involves changing the sterile dressings that cover much of their bodies and bathing them in a solution of water and bleach several times a week, for several hours, to stave off infection. Susan performs Anthony's bathing ritual herself. It is not only time consuming and physically taxing, but it requires her to see her child in immense pain, indeed to cause the pain herself. Although the ritual is meant to prolong Anthony's life, it also serves to make vivid on an

[31] Joe dies during the film. Anthony died in 2004, after the film was made. Anthony was the second of Susan Tom's children to die from EB, and she has since adopted another baby with the disease.

almost daily basis the fact that he is slowly dying. On top of this, Susan Tom provides nursing care for the other children who require it, accompanies them to hospital stays and surgeries, and does all of the cooking, cleaning, shopping, and laundry that one can imagine is generated by a household with a dozen children. (She gets some help from her eighteen-year-old daughter Margaret and a babysitter, but is otherwise on her own.)

Despite all of this, one of the most remarkable things about Susan Tom's household is that it is on the whole a joyful and vibrant place, filled with laughter and play.[32] For Halloween, Susan organizes an impressive backyard spectacle; all of the children wear costumes, and Joe performs a magic trick in which he saws his sister in half, a trick that works especially well since she has no legs to begin with. One gets the impression that these kids are especially fortunate, not simply for having been given a permanent and safe home, but for ending up in this *particular* family with this particular mother.

Sacrifice and Well-Being

Paul Farmer and Susan Tom lead lives marked by great sacrifice. Yet their lives also seem unusually meaningful and fulfilling. They seem to *flourish*. Just what does it mean, then, to say that a life is marked by great sacrifice?

Thus far, I have argued that we should look at sacrifice in the following ways. First, the notion of sacrifice we are interested in is tied to moral obligation. So we are not interested in *all* the ways in which someone could be said to give something up, but in just those cases where what they are giving up is *urgent* enough to generate a legitimate

[32] The only exception to this is the disruption sometimes caused by Joe's behavioral problems, which cause conflict between him and his sisters. Susan Tom seeks professional help for him when Katie, who is mentally retarded, reports that he attempted to engage her in inappropriate sexual behavior.

claim. In order to satisfy this constraint, following Scanlon I have argued that it makes sense to restrict "sacrifices" to only those losses that fall under certain general categories about which there is some consensus. We determine whether something falls under such a category by looking at the role it plays in a person's life and trying to *understand* it as a member of one of those categories—this I called the *intelligibility constraint*. Next, I claimed that this account of sacrifice is *objective* in the sense that potential sacrifices are evaluated independently of whether the agent takes them to be sacrifices. Moreover, on this account sacrifice is a fundamentally *normative* notion. For something to be a sacrifice, it is not sufficient that someone has given something up; rather, what is given up has to *matter* in the right way.

One rather useful way to accommodate all of these constraints is to think of sacrifice as *losses of well-being*. A similar proposal is made by Liam Murphy in *Moral Demands in Nonideal Theory* (2000). Murphy is interested in the "overdemandingness" objection to the principle of optimizing beneficence—that is, the claim that a moral principle requiring us to help others until it is counter-productive is too *demanding*. So Murphy needs an account of just what it is for a moral theory to make a *demand* in the relevant sense. He defines "demands" as the losses of well-being (relative to the "factual status quo") that are sustained when an agent complies with a moral theory (17). So, for example, if a moral theory requires that I donate a certain amount of money to famine relief, and in complying I sustain a loss in well-being, then that theory has imposed a *demand* on me—not just in the sense of "demand" in which every moral requirement is a demand, but in a thicker sense that includes the effect *on me* of complying with the requirement. Murphy argues that demands should be measured against the "optimally

prudent life" (51). So he claims that "prudence, which requires a person to optimize his own self-interest over time, makes no demands" (17, note).

Notice, however, that while Murphy's "demands" and the sacrifices I am discussing are both losses of well-being, not all sacrifices are demands. For "demands" in Murphy's sense are just those losses of well-being that are generated by the requirements of a moral theory. *Sacrifices*, on the other hand, can accompany actions that are not *required* by a moral theory. After all, someone can certainly make a sacrifice in the course of performing a supererogatory action—part of what *makes* it supererogatory is the fact that it involves sacrifice. Indeed, one purpose of this chapter is to figure out just what role sacrifice plays in the lives of moral saints, who are in a sense chronic supererogators. Sacrifice also differs from Murphy's notion of "demands" in that, whereas demands are a feature of moral *theories*, sacrifice could presumably play a role in our moral lives even if there were no single correct systematic moral theory but just a collection of obligations, permissions, recommendations, etc. (Perhaps Murphy would be willing to think of a "demand" as the cost in well-being of complying with a moral *obligation* rather than with a moral *theory*. If so, then this point would be moot, but the above point about supererogation would still hold.)

Finally, there is one more difference between "demands" and sacrifice, and if I am interpreting Murphy correctly, it is a very important difference. In referring to a loss of well-being, Murphy seems to be thinking of a *net* loss. So, if a moral theory requires that I volunteer my time at a homeless shelter, and I lose some amount of well-being because I have to wake up extra early to get there, but I gain an equal amount of well-being because the work is very enjoyable, then it seems I have no *net* loss of well-being

and so this theory imposes no "demand" in Murphy's sense. However, it still imposes a *sacrifice* in requiring me to wake up extra early, for as I shall argue later, the relevant notion of sacrifice is one that is not diminished by the fact that certain losses are compensated by other gains. To be sure, someone with a demanding lifestyle—say, an astronaut—would no doubt rather give up certain goods (comfort, time at home, safety) and also *gain* certain goods (adventure, a place in history) than to give up certain goods and gain *nothing*. But his job still involves *sacrifices* even if the goods that are gained are as valuable as the goods that are lost.[33]

An Objective Theory of Well-Being

I hope that this contrast with "demands" has brought out some of the important features of the notion of sacrifice we have been examining. Let me return to the idea that sacrifice can be thought of as a loss of well-being. While Murphy does not endorse any particular theory of well-being, I would like to suggest that an account of sacrifice ought to be based on an agent-neutral theory of well-being like the one offered by Stephen Darwall in *Welfare and Rational Care* (2002). Darwall's main task is to give an account of well-being at the *conceptual* level; his account can be paired with various theories at

[33] Michael Stocker discusses a related phenomenon in his book *Plural and Conflicting Values* (1990). In the chapter "Plurality and Choice," he gives an account of what it is to lack a good. His account allows for cases in which a life can coherently be said to *lack* a particular kind of value even if the absence of that value makes possible the presence of other important values. Stocker argues that value is essentially plural, and therefore we cannot say that one value could ever really *compensate* for the loss of another value. Thus he claims that "a life can be lacking in pleasure even if the only way to have additional pleasure would involve losing so much wisdom and honour that the life with greater pleasure would be overall worse" (171). I think this is correct, and I think my notion of sacrifice is similar to Stocker's notion of *lack*. Paul Farmer is *sacrificing* time with his family even his life is overall better on account of this sacrifice, because the goods generated by his lonesome lifestyle do not *replace*, as it were, the time spent with family.

the normative level, and I will argue that for the purposes of thinking about sacrifice, it makes sense to consider an "objective list" theory at the normative level.

Darwall argues that a person's well-being (or her welfare, her "good") is "constituted, not by what that person values, prefers, or wants (or should value), but by what one (perhaps she) should want *insofar as one cares about her*" (4). In other words, a person's welfare is whatever it is rational for *us* to desire *for her* insofar as we care about *her* (9). On this view, "the normativity of welfare is not agent-relative but agent-neutral" (20). That is, Mary's preference for a sixth television does not essentially factor in her well-being in virtue of its being a *preference* or even in virtue of its being *her* preference. Rather, if it does factor in her well-being at all, it is only contingently, if it happens to be the sort of thing that we should want *for her sake* insofar as we care for her.

What makes this view so compelling is that it can accommodate the astonishing variety of idiosyncratic tastes and interests that can be seen to play a meaningful role in a person's life and yet it does not make well-being depend essentially on the agent's point-of-view. In other words, the fact that the agent values something does not *make* that thing part of her good; what is good for her can come apart from what she cares about. This is important because it leaves open "the possibility of pursuing values one cares deeply about at some cost to oneself" (3). It should be immediately clear why this possibility is going to be crucial to an account of sacrifice that bears on the behavior of moral saints. As Darwall puts it, "if there were no difference between what a person valued and what benefited him, self-sacrifice would be impossible, except through weakness of will. […] [I]t would be impossible for pursuing one's values ever to cost one *on balance*, since realizing a value would be the same thing as benefiting from it" (3). While I am not

committed to the claim that self-sacrifice is a *necessary* characteristic of *every* moral saint, it is difficult to see how we could talk about moral saints at all if our theory of sacrifice left no room for actions that run counter to the agent's welfare.

We can now see what it would look like for an account of sacrifice to be indexed to an agent-neutral theory of well-being: *A sacrifice is a loss of something that contributes to the agent's good, where something contributes to her good just in case it is the sort of thing it would be rational for us to want for her, for her sake, insofar as we care for her*. For example, it would be rational for those who care about Susan Tom to want her to have a retirement account, adequate health insurance, meaningful social interaction with other adults, and a life free from watching her children endure painful medical procedures. All of these things are sacrifices she incurs in order to give her children a decent life. And because well-being is defined independently of what the agent actually desires or enjoys, these losses count as sacrifices even if Susan Tom doesn't *miss* having a retirement account or health insurance. Nevertheless, this notion of well-being does leave space for the agent's desires and interests in the following way: when such interests play an important role in a person's life, they are very likely to be among the things it is rational for us to want the agent to have for her own sake.

For Paul Farmer's sake, we want him to have a decent night of sleep, plenty of time with his family, and a life free from exposure to dangerous diseases. To be clear: these are not things we want for Paul simply because we want him to be happy, or because his happiness brings us happiness. Rather, these are things we should want for him for his *own sake, insofar as we care about him*. As Darwall points out, care is rooted in sympathetic concern; it is a "natural psychological kind" (12). Unlike the mere *desire*

that Susan and Paul be benefited or be happy—which desire can be only satisfied or thwarted—*care* is accompanied by an emotional response on the part of the carer. If I found out that Susan Tom was perfectly happy despite having to watch her children suffer, and if well-being were simply happiness or benefit, then it would seem that I would have to be satisfied with her situation. If, on the other hand, Darwall's theory is correct, then we can make sense of the sadness I might feel for Susan insofar as I care about her, a sadness that persists even if it turns out that she is happy by her own lights.

It is worth noting once again that Darwall's account of well-being as presented so far operates at the *conceptual* level. It is not meant to provide a test, as it were, for us to apply in each and every case. Rather, it is meant to be paired with another theory at the normative level: a theory that tells us, as a substantive matter, what sorts of things actually promote well-being. We cannot tell just by looking at the conceptual level whether a life filled with health, or wealth, or safety, or marvelous aesthetic experiences, or ignorant bliss, or in fact none of the above, would be the sort of life that it would be rational for us to want for someone for her own sake out of care for her. A theory of well-being at the normative level will spell that out for us. In keeping with the constraint we borrowed from Scanlon—that benefits and burdens need to be intelligible as members of certain general consensus categories in to generate legitimate claims—it makes sense to pair Darwall's theory with some variant of an "objective list" theory at the normative level.

What is an objective list theory? Derek Parfit (1984) offers a pithy definition: "On *Objective List Theories*, certain things are good or bad for us, whether or not we want to have the good things, or to avoid the bad things" (493). Objective list theories of

wellbeing meet the constraints on sacrifice I developed earlier: they make sense of the idea of *urgency*; they meet the intelligibility constraint; they are normative rather than descriptive; and they are objective insofar as they do not essentially depend on the agent's point-of-view. Other theories of well-being (e.g., hedonism) tend to fail on one or more of these criteria.

Different objective list theories give different lists, of course. Though Parfit does not offer his own list, he does suggest some examples of the things that might appear on such a list:

> The good things might include moral goodness, rational activity, the development of one's abilities, having children and being a good parent, knowledge, and the awareness of true beauty. The bad things might include being betrayed, manipulated, slandered, deceived, being deprived of liberty or dignity, and enjoying either sadistic pleasure, or aesthetic pleasure in what is in fact ugly (499).

Something can contribute to a person's good either by providing one of the good things on the list or by thwarting one of the bad things on the list. So, for example, we might ask how a situation like the one from *Rain Man* discussed earlier would fare when analyzed in terms of an objective list theory. It might be that watching his television shows provides Raymond with opportunities for *rational activity* and the *development of his abilities*. Or perhaps taking his television away would constitute a *betrayal* of one's relationship with him, or constitute *depriving him of his liberty*. If none of these seems

exactly right, it may be that we need a more fine-grained list, or we simply need a longer list or a different list altogether.[34]

In his book *Well-Being* (1986), James Griffin offers a list of just five items: accomplishment; the components of human existence; understanding; enjoyment; and deep personal relations (67). The list is deceptively short. Quite a bit is packed into some of the items on the list. "Understanding" refers both to knowledge in general and to the capacities of "being in touch with reality, being free from muddle, ignorance, and mistake" (67). The second item, "the components of human existence," requires a bit more explication:

> What makes life 'human,' in the distinctly normative sense the word has here, is not a simple thing. The systematic way to understand its complexities is to understand the complexities of 'agency'. One component of agency is deciding for oneself. Even if I constantly made a mess of my life, even if you could do better if you took charge, I would not let you do it. Autonomy has a value of its own. Another component is having the basic capabilities that enable one to act: limbs and senses that work, the minimum material goods to keep body and soul together, freedom from great pain and anxiety. Another component is liberty: the freedom to read, to listen to others; the absence of obstacles to action in those areas of our life that are the essential manifestations of our humanity—our speech, worship, and associations (67).

Ultimately, it is difficult to think of anything that matters that cannot be found in Griffin's list either explicitly or implicitly. Some areas, of course, require greater interpretation. For instance, Griffin elaborates on "deep personal relations" by saying that we need "deep, authentic, reciprocal relations of friendship and love" and that such

[34] Another objective list theory is Martha Nussbaum's 'capabilities' as described in *Sex and Social Justice* (1999, 41-42). In brief, her list is: 1) Life. 2) Bodily health and integrity. 3) Bodily integrity. 4) Sense, imagination, thought. 5) Emotions. 6) Practical reason. 7) Affiliation. 8) Other species. 9) Play. 10) Control over one's environment.

relationships "have a value apart from the pleasure and benefit they give" (67-68). One wonders whether the value of such relationships is coarse grained enough that, say, parent-child relationships and romantic relationships have essentially the same kind of value, or if they actually represent two different goods. If there is something valuable in romantic relationships that is fundamentally distinct from what is valuable about parent-child relationships, then someone like Susan Tom, who has one but not the other, is missing out.

Lurking behind any objective list theory is the idea that some goods are more vital or more basic than others. Whatever Griffin means by "accomplishment," it is unlikely that it turns out to be as important as the things he lists under "the components of human existence." The beauty of the list, however, is that anything that is on it satisfies the first part of the "intelligibility constraint"—it's a member of a "familiar general category." These are not luxury items; every human being can lay a claim to these items and hold others responsible for not interfering with the exercise or aquisition of them. Our claims to these items are not unlimited, though. Not every good that can be traced back to the list is necessarily *urgent* enough to act as a defeater of moral obligations. Recall that the intelligibility constraint requires not only that items be understandable as falling under the general categories, but also that we can understand the significance of the cost of giving up those items as being *proportional*—proportional to their role in our lives, proportional to other important goods, and proportional to whatever we are giving them up to pursue. Suppose, for example, that one of the general categories is "meaningful friendship." Presumably, most people would have a somewhat automatic claim to at least a few close friends, in the sense that it would generally be unreasonable to ask a person to

sacrifice *all* of her close friends in service of some moral project, such as feeding the poor. Thus it would generally not be the case that a person is *morally obligated* to feed the poor at the cost of all of her close friendships, even if doing so would be praiseworthy. However, if a person already has a rich and abundant social life, and feeding the poor will require only that she spend a bit less time with a few of her friends—but not require that she give up the good of friendship entirely—then she will have far less of a claim in the face of putative moral obligations. Proportionality entails that giving up time spent with a particular friend, for example, will be more or less of a sacrifice for different people depending on the background facts—more sacrifice for the person with just one close friend, less sacrifice for the person with abundant opportunities for social interaction.

We can now better appreciate the role of the intelligibility constraint. It helps us to measure urgency by requiring not only that we be able to understand unusual attachments as falling under broad consensus categories, but also that we be able to understand any given attachment, whether unusual or typical, as being *proportional* to its role within the category. Moreover, the categories themselves might be weighted relative to each other. So we determine the urgency of a particular good by evaluating it in light of a contextual, weighted interpretation of the objective list theory of wellbeing. But whether we can reasonably be *required* to give up a particular good will depend not just on the *urgency* of that good, but also on what is at stake, morally. Being healthy, for example, is a very urgent component of wellbeing. Yet clearly there are situations so dire, where so much is at stake, that one might be required to sacrifice one's health—for instance in various rescue scenarios, especially involving family members.

With something like Griffin's list in mind, then, I would like to offer a rough taxonomy of different respects in which the life of a moral saint might involve *sacrifice*. The taxonomy is not meant to represent an objective list theory in its own right, and represents only a subset of the goods one would find on such a list. It is rather a spelling out of how certain more common kinds of losses of well-being might group together in real life, especially in the lives of moral saints or those who resemble them in certain ways.

A Taxonomy of Sacrifice

The items in this taxonomy seem to fall roughly along two dimensions. In the first dimension we find more direct losses of well-being or instances of ill-being, including the categories I call "engagement with intrinsically unpleasant or aversive features" and "loss of creature comforts." The second dimension is higher-order: it includes ways in which a person's *capacity* to garner well-being is compromised. What I call "opportunity costs," "loss of liberty," and "personality changes" fall in this dimension. The division into two dimensions is rough and open to debate; certainly the higher-order sacrifices can also be thought of as direct losses of well-being of the sort found in the first dimension. Moreover, the precise boundaries of the five categories themselves are also open to interpretation. There is bound to be some overlap between them and some losses may not fit neatly into any of them.

First Dimension: Direct Losses of Well-Being or Instances of Ill-Being

Engagement with Intrinsically Unpleasant or Aversive Features. Some activities involve doing things that are simply awful. In the 2006 film *The Painted Veil* (based on

the Somerset Maugham novel of the same name) Edward Norton plays Dr. Walter Fane, a British infectious disease researcher who volunteers to travel to the interior of China in the 1920s to treat a cholera epidemic ravaging the rural villages. Though Fane's motives and character are mixed (he is surely not a moral saint), his action involves an enormous amount of sacrifice. Cholera is a positively foul diarrheal disease that causes patients to lose massive amounts of fluids very quickly and, without proper treatment, to die from dehydration. When Dr. Fane first arrives at the rural clinic, he is faced with an utterly disgusting atmosphere—heat, stench, and dangerous germs, all festering in close quarters. This is a case where doing one's job requires overcoming multiple intrinsically aversive features, including stifling the urge to vomit and watching other human beings die. The raw human suffering is ubiquitous and overwhelming. The real-life example of Paul Farmer bears many similarities; sometimes his job involves doing things that are not only difficult but disgusting, back-breaking, and heart-wrenching. The same goes for Susan Tom and the way she cares for Anthony, who suffers from Epidermolysis Bullosa. As I mentioned earlier, caring for Anthony involves cleansing his open sores with a bleach solution several times a week. This ritual is not just physically taxing but also intrinsically unpleasant: it involves seeing things one would rather not look at; it causes pain to a loved one; and it raises to salience in a particularly gruesome way the gravity of Anthony's condition.

Loss of Creature Comforts. Here I want to suggest that the losses of creature comforts associated with certain lifestyles are a kind of sacrifice distinct from the category above and distinct from opportunity costs, which I will describe below. In principle, this category could be broadened to include the loss of physical and mental

health at the most basic level. But here I will be concerned with more peripheral comforts, like hygiene, nutrition, and rest. In Haiti, Paul Farmer lived in a small hut with no hot water. He filled his suitcases with medicine and supplies from Boston, leaving room for only a few shirts, so he rarely had the comfort of clean clothes. The knowledge that people were dying and that only he could save them understandably constrained his ability to sleep, so he rarely had the comfort of a restful night's sleep. His travel schedule was so hectic that he offered the advice, "Traveler's tip number one thousand seventy-three. If you don't have time to eat, and there's no other food on the plane, a package of peanuts and Bloody Mary mix are six hundred calories" (191). Clearly, a healthy diet was just one of many things Farmer sacrificed in order to maintain his saintly lifestyle.

These sorts of comfort-based sacrifices don't seem to fit neatly within our first category, "engagement with intrinsically unpleasant or aversive features." It is not an essential feature of treating sick patients that you must live in a hut, or lose sleep, or survive on peanuts. (Of course, it might be an essential feature of treating patients in a *particular* location that it is, say, unbearably hot. If so, I would classify this under "engagement with intrinsically unpleasant…" But there are many cases where the lack of creature comforts comes apart from the features of the work itself.) When Paul Farmer forgoes creature comforts, it tends not to be because his work essentially requires it, but because he finds that the creature comforts get in the way of his doing the work as efficiently and wholeheartedly as possible. Moreover, these creature comforts seem insignificant in the face of the dire suffering he witnesses on a daily basis.[35]

[35] Indeed, hot water, clean clothes, and a varied diet are not basic necessities but *luxuries* when considered relative to the living conditions in some developing countries. Still, it seems uncontroversial to say that giving them up is a *sacrifice* for Farmer, given the

Nor does the loss of creature comforts fit neatly in the "opportunity costs" category that I will describe below. For the problem is not exactly that there are opportunities to do things that are closed off on account of the life Farmer has chosen, but rather that the life he has chosen puts pressure on him to do things a *certain way*. So, the fact that Farmer can't sleep in New York City while he's treating patients in Haiti is an opportunity cost, but the fact that he can't sleep in a comfortable bed is a lost creature comfort that arises from contingent facts about treating patients in Haiti: such a bed might not be available, or it might not make sense to own one when you've got so much work to do, or it might undermine your ability to do your job by making you seem pampered or picky. It could even be that forgoing certain creature comforts is not in any way necessary, but stems from a change in mindset that comes with choosing a certain lifestyle. Those who spend their days advocating on behalf of the poor may have trouble paying attention to things like their diet when doing so would seem self-indulgent. Or they may simply be *unable* to pay attention to such things because they are so occupied with helping others. This could be the case even if they would otherwise prefer to keep a healthy diet. In the extreme, certain saintly lifestyles might even induce full-fledged personality changes that could have consequences for one's ability to acquire or even appreciate various creature comforts. I'll say more about personality changes below.

---

economic conditions that exist in his home country and given the account of sacrifice I've put forward. Indeed, it seems uncontroversial that the lack of these comforts is a loss of well-being for those living without them no matter what the baseline conditions are. It seems we can only call the loss of these comforts a *sacrifice*, however, if they would otherwise be available but have been given up (or taken away) as a result of some kind of *choice*—perhaps even a coerced choice.

Second Dimension: Threats to One's Capacity to Garner Well-Being

Opportunity Costs. Perhaps the most frequently discussed variety of sacrifice that moral saints face is the *opportunity costs* associated with the lives they have chosen. Here I don't mean the technical term "opportunity cost" as it might be used in economics, for we are not interested in a notion of well-being that is understood in terms of preference satisfaction. Rather, as a first pass, opportunity costs could be defined counterfactually in the following way: *things a person could have done, if she weren't doing what she's doing instead*. Of course, there are an infinite number of things Paul Farmer could have done instead of working for Partners in Health: he could have been a professional wrestler, like his brother Jeff, who goes by the name "Lightning"; he could have been a radiologist; he could have been an armchair scholar. Yet we wouldn't want simply to add up all of these alternative careers and count them as losses. After all, nearly everyone has some set of alternative careers, but we don't think of everyone as having *sacrificed* all of those alternatives in choosing their actual-world career. So perhaps the notion of opportunity costs needs to be refined. Call opportunity costs *things a person could have done, and that she* would *have done, if she weren't doing what she's doing instead*.

Of course, we are not interested in just *anything* Paul Farmer would have done, only those things that would have garnered him well-being. Perhaps he would have wanted to be a loan shark, and perhaps he would have enjoyed this very much, but it's highly unlikely that this would have been good for him in the objective sense we are interested in. On the other hand, in choosing to be a doctor rather than a pro wrestler like his brother Jeff "Lightning" Farmer, Paul may certainly have been forgoing well-being—surely winning matches would count as "accomplishment," item number one on Griffin's

list. (That said, since this is *professional* wrestling, we might worry that it fares poorly on Griffin's requirement that one be "in touch with reality.")

Now, I have argued that we should count as opportunity costs only those foregone options that would generate well-being and that the person *would* have chosen, not all those that she *could* have chosen. Notice, though, that to a certain extent some of the options she *could* have chosen (but would not have) are the more interesting ones. Well-being as we are understanding it here is fundamentally normative. It's not about what a person *does* value but what is *worth* valuing—what we should want for the person insofar as we care for her. Susan Tom may not have chosen a personal life even if she could have; yet romantic relationships may still have been good for her. Insofar as such relationships constitute something that would have been good for her, but that her lifestyle makes impossible, it might seem that they should count as an opportunity cost.

But as I pointed out earlier, there are millions of things Susan Tom or Paul Farmer *could* have done, and so if we don't limit the opportunity costs to what they *would* have done, we risk piling on an unintuitively large set of costs. Notice, however, that while there may be a very large number of things, each of which it would have been good for Susan to be able to do, it does not follow that it would have been good for her to do *all* of those things put together. Her lost opportunities need not be aggregated in that way. Rather, it seems that determining the opportunity costs of a given lifestyle should be highly context-sensitive. Susan's lifestyle as an uber-mom rules out other opportunities in more and less salient ways. There are certain things she cannot do because she is an uber-mom in particular, and then there are many other things she cannot do in a weaker sense just because a person can generally only manage to have one career at a time. So, she

can't go to midnight movies because she has to be home for her kids—this, it seems, is a salient and relevant opportunity cost. On the other hand, she can't be a trapeze artist because she's already a full-time mom. This is ruled out in virtue of the fact that she can only have one career, rather than being ruled out in virtue of her *particular* career. The difference seems important. Furthermore, in evaluating opportunity costs, we will need to weigh not just the potential well-being that could have been garnered from certain alternate lifestyles, even when the person would not have *wanted* those lifestyles, but also the fact that perhaps a more fundamental source of well-being is the ability to choose one's own destiny on the basis of values one takes to be important. I'll say more about this below.

Loss of Liberty. Loss of liberty is a kind of sacrifice that is related to, but distinct from, opportunity costs. While it would not be good for Susan Tom to live life as a homeless, itinerant menial laborer, it is certainly good for her to be free to choose the course of her own life, and being an itinerant laborer is one of the things she might choose. As a result of this freedom to choose the course of our lives, we often *volunteer* to place limits on our own liberty. By having children, for example, a parent limits her freedom to do certain things, like travel in the rugged backcountry or keep certain dangerous animals as pets. Nonetheless, there does seem to be a meaningful line between ordinary constraints on liberty and the extraordinary constraints that come with a life like Susan Tom's. Ordinary motherhood, for example, might come with certain limits built-in, like being forbidden to ingest certain substances while one is breast-feeding. But there is an entirely different degree of restriction that comes with raising a disabled child, let alone raising twelve disabled and ill children like Susan Tom does. The obligations

associated with being the primary care-giver for these children significantly limit the things Susan Tom can do, for example by limiting the places she can go with her family (only those that are wheelchair-accessible), the times she can freely determine her own schedule and routine (only when none of the kids are in the hospital), and the people she could, at least in theory, be romantically interested in (only those who aren't looking to be taken care of, since she has enough of that at home already, she says).

Loss of liberty as a type of *sacrifice* is tricky. After all, Susan Tom *freely chose* to adopt twelve children, knowing the special obligations that would result. Paul Farmer *freely chooses* to treat the world's poor and to live among them, knowing that because of this choice he cannot wake up one morning and satisfy a spontaneous urge to, say, go to McDonalds. It might seem odd that we should consider the minor restrictions on liberty associated with major life choices, which major choices *were themselves free choices*, to be a particular species of sacrifice *qua* loss of liberty. But while Susan's original choice to adopt sick children was freely made, and while it was supererogatory and hence morally optional, *once she made it* she incurred moral obligations that cannot be ignored.

Now, it might also seem that Susan Tom's inability to go on certain dates, and Paul Farmer's inability to go to McDonalds, if sacrifices at all, fall under *opportunity costs* as described above. Nevertheless, I think we are dealing with a unique phenomenon here. If Susan cannot go to the park because her son is in the hospital, this results in two kinds of sacrifice. First, she misses out on the wellbeing she could have gotten from a trip to the park—that's an *opportunity cost*. Second, she is in the position of having her options *constrained* in a way that the options of someone without a sick child would not be—this constraint is a *loss of liberty*, and it's burdensome in a way that is different from

the way opportunity costs are burdensome. The fact that she *chose* to adopt a sick child does not cancel the effect that the loss of liberty has on her well-being. We can choose to place limits on our own liberty and still bear the losses in well-being those limits entail. It is good for me to be free to make my own choices; that doesn't mean that whatever I choose will necessarily be good for me. (And even if choosing to adopt sick children is, on the *whole*, good for Susan, that does not weaken the fact that it comes with certain sacrifices. Sacrifice is not negated by *net benefit*, as I will discuss later on.)

Personality Changes. Thus far I've focused on types of sacrifice that, while they may affect the agent profoundly, are not essentially a matter of *changing* the agent. But there is a type of sacrifice that involves the agent's very constitution, and it can count as a sacrifice even if the agent herself thinks it is a change for the better. The type of change I'm referring to happens when, as a consequence of the sort of lifestyle led by people like Susan Tom and Paul Farmer, an agent's personality undergoes profound changes that it otherwise would not have. For example, as a consequence of being around poverty, disease, and the inaction of the privileged, Paul Farmer has developed a personality marked by cynicism and anger (though not dominated by them). As a mother who has now lost three children to terminal illnesses, Susan Tom could develop a disposition to be resigned to tragedy, or an inability to fully realize joy. A Holocaust rescuer who spends years engaging in deception in order to keep others safe might subsequently find it difficult to cultivate habits of truth-telling, or to take others on their word.

As I've argued elsewhere, I don't think that there is something about moral sainthood *itself* that constrains personality traits (with the exception, perhaps, of essentially moral traits like integrity or sympathetic concern). I disagree on this matter

with Susan Wolf (1982), who thinks that moral saints, necessarily, cannot enjoy certain kinds of humor. I don't think that, in order to *be* a moral saint, a person needs to have certain character traits and lack certain others. Nevertheless, I *do* think that the actions of a moral saint can, over time, serve to alter her personality. These personality changes will count as *sacrifices* if they result in losses of well-being, even if they also make possible other actions which bring in a net positive amount of well-being.

Sacrifice and Net Benefit

It would be unwise to spend so much time cataloging the varieties of sacrifice without also noting that moral saints undoubtedly reap great *rewards* from their actions. For Susan Tom, a houseful of fourteen children isn't a moral project—it's just her *family*, and as such it is the most important thing in her life. She makes great sacrifices for her family, but in return she gets a vibrant, grateful brood, always there to light up her day and to make use of her rare talents. The satisfaction she feels is revealed in her facial expressions when she presides over the backyard festivities on Halloween, including the "sawing in half" of one of the girls with no legs; when she gives a pretend haircut to her daughter, who has no hair because of burns sustained in infancy but giggles through the haircut nonetheless; and when she watches in amused pride as one of her wheelchair-bound middle-schoolers brings home her first "boyfriend." In a more serious vein, it's also evident that Susan Tom has an acute awareness of what life might have been like for her children had she not adopted them. She seems aware on some level of the immensely good deed she has done, and surely she finds fulfillment in that knowledge.

Paul Farmer is perhaps a bit different. One doesn't get the impression that his days are filled with joy, or that he sleeps well at night knowing he has done an immensely good deed. Instead, Farmer is driven compulsively to work. "I can't sleep," he says. "There's always somebody not getting treatment. I can't stand that" (24). While he acknowledges making great sacrifices, he claims they stem from a need to relieve the tension that comes with the feeling of not doing enough. Whereas other people might feel joy and satisfaction, Farmer merely feels *relief*, and only temporarily so. "If you're making sacrifices," he says, "[…] it stands to reason that you're trying to lessen some psychic discomfort" (24). He thinks his own sacrifices can "be regarded as a way to deal with ambivalence. I feel ambivalent about selling my services in a world where some can't buy them" (24). So while Farmer may not feel the phenomenological upside of a *net benefit* in well-being when his sacrifices are weighed against the rewards of his work, he certainly does not slump through his days burdened and bothered by the weight of the sacrifices. Sacrifice keeps him going.

Indeed, in *Welfare and Rational Care*, Darwall argues that his rational care theory of well-being is fittingly paired (though it need not be paired) with a normative theory of a person's good according to which "the best life for human beings is one of significant engagement in activities through which we come into appreciative rapport with agent-neutral values, such as aesthetic beauty, knowledge and understanding, and the worth of living beings" (75). On this criterion, which is similar to our objective list except insofar as it emphasizes the "appreciative rapport," one can't help but think that the lives of Susan Tom and Paul Farmer are stunningly good ones, with unusually high levels of well-being, despite the sacrifices these lives involve. How better to come into

"appreciative rapport with agent-neutral values" than to give a home to children who might not otherwise have one, and to heal the sick when no one else will? Indeed, as Liam Murphy has suggested, the level of sacrifice involved in certain moral projects may change over time as the agent comes to *identify* with the goals of the project, to take its ends as her own (2000, 19). It could be that the more Paul Farmer identifies with the project of healing the sick, the more those who care about him ought to want a life *just like his life*, for him, for his own sake. We might still want him to get a good night's sleep and a healthy diet, but the well-being garnered from his self-sacrificial lifestyle could increasingly outweigh the well-being he could have garnered from the creature comforts he forgoes.

That said, even if it is the case that lives like that of Tom and Farmer turn out to have a *net benefit* of well-being, this need not threaten our initial judgment that their actions, on the whole, are beyond obligation, and that part of what makes them beyond obligation is the level of sacrifice involved. For there are many ways to live one's life and garner a net benefit of well-being without making deep sacrifices, and there could be many lives that have both a much greater net level of well-being and a much lesser degree of sacrifice than the life of a moral saint. Great sacrifice, especially at the margins, seems to be something that mitigates obligation even if the sacrifices enable person to perform actions that greatly enhance their well-being.

A Worry About the Priority of Well-Being

By grounding a theory of reasonable sacrifice in well-being, I have tried to reconcile two basic facts: on the one hand, if sacrifices can relieve people of moral

obligations, then we don't want people to be able to claim just anything they want as a sacrifice; on the other hand, we want to leave room for the rich variety of human interests, and we want a theory that is *humane*—that gives people the benefit of the doubt when they take something to be a burden, and that treats their own judgment as having some legitimacy. A theory of sacrifice that cannot accommodate both of these facts would generate implausible results. If a theory couldn't accommodate the first fact, it would say that a person is not obligated to save a drowning toddler if doing so would involve missing his afternoon nap, given how much the nap means to him. And if it could not accommodate the second fact, it would seem to say that a person is required to sacrifice her life's single passion—which is to play a beautiful but somewhat expensive violin—to donate money to feed the poor, since no musical instrument could be as important as the need for food, no matter its role in someone's life.

By trying sacrifice to an objective, agent-neutral theory of well-being, we ensure that sacrifices rise to the level of legitimate claims. Feeling something *as* a sacrifice doesn't make it so, because not just any old preference or feeling is sufficiently urgent to be able to justify demands on others. Still, the theory of sacrifice I have offered gives agents the benefit of the doubt with regard to their own lives. It does, of course, require that losses fall under general categories about which there is some consensus—the categories we find in our objective list—and as such it only legitimizes those losses that are intelligible to other agents. But the intelligibility constraint is flexible enough, and the objective list is broad enough, that is it highly unlikely that anything that really *matters* to a person's life would be barred from counting as a sacrifice. It's just that whether something matters is not essentially tied to the fact that she *thinks* it matters; this theory

leaves open the possibility that agents can be mistaken about the urgency of their own commitments.

One worry about tying sacrifice to well-being in this way is that it is unreasonably paternalistic, insofar as it defines sacrifice independently of the agent's preferences.[36] Why should we tie sacrifice to *well-being* rather than, say, to people's standing as autonomous agents? After all, there are certain losses that would seem to be sacrifices even if they do not lower someone's well-being. For instance, surely it is not a loss of well-being to quit smoking, but requiring someone to quit smoking certainly compromises his ability to choose the course of his own life—surely if he prefers to smoke, he is making a sacrifice in giving it up. (Of course, cases where the state *forces* or *coerces* people to do things like quit smoking—paradigmatic cases of paternalism—are not directly relevant to my arguments, which concern how sacrifice bears on the boundary between obligation and supererogation. However, the objection still has some pull, for in saying that some losses serve to defeat moral obligations while others don't, we do seem to be effectively restricting what people can do—not restricting it by force or coercion, but by imposing *moral* restrictions. And this might seem paternalistic in at least a weak sense.)

While I think this is a significant worry, I do think there are satisfactory ways to respond to it. The first response is simply to point out that objective list theories of well-being normally give a prominent place to liberty and autonomy (Griffin's certainly does), and so to a certain extent people's standing to lead their lives as autonomous agents is already *built in* to the theory. Anything that restricts that liberty or autonomy will count

[36] I thank Stephen Darwall for raising this worry.

as detracting from well-being. We saw this quite vividly in the taxonomy of sacrifices, where a lifestyle like Susan Tom's was shown to be sacrificial in virtue of the restrictions it placed on her liberty.

Of course, on the view I have proposed, liberty and autonomy are just two concerns among many. Sacrifice is not *grounded* in liberty or autonomy, but rather in a notion of well-being that trades on many values. There will be cases where some of these values conflict. For example, demanding that someone quit smoking might threaten his autonomy, but it might at the same time promote his health and enable him to engage in more meaningful relationships.[37] In terms of the net cost to his well-being, autonomy gets no special place; the other values might prove to be more significant. Yet we ought to recall our earlier discussion of the relation between sacrifice and net benefit. Just because an action results in a net positive amount of well-being does not mean it imposed no sacrifices. Quitting smoking might entail no *net* loss of well-being but still entail *some* losses. Indeed, it is perfectly compatible with the view I have proposed that the person who is compelled to quit smoking makes *sacrifices* (due to his lost autonomy) even though, *overall*, it is "for his own good"—that is, a net benefit in well-being. So it is actually mistaken to claim that my view leaves no place for sacrifice when there is no loss of well-being.

Nevertheless, one still might worry about granting *priority* to well-being rather than to the standing of autonomous agents to lead their lives. I'm not sure I have much to

[37] The quitting smoking case is actually an interesting one, since there is perhaps a sense in which smoking (assuming one is addicted) *itself* restricts one's autonomy (liberty?). An addiction can be coercive in its ability to interfere with a person's ability to make certain choices. So perhaps the autonomy/liberty concerns actually point in the same direction as the other values in this case.

say in response to this worry other than it seems to me that the well-being view best captures what we desire out of a theory of sacrifice in the ways I outlined earlier—it makes sense of urgency in an agent-neutral way, while still giving the benefit of the doubt to agents. And it certainly does not *neglect* autonomy and liberty, since these factor into well-being. Insofar as the well-being view risks paternalism, it seems to do so in only the weakest sense—if demanding that people sometimes sacrifice their own interests in order to fulfill their moral obligations is paternalistic, then it seems that morality *itself* is paternalistic. *Forcing* people to fulfill obligations (say, by state-imposed coercion or incentives) may be paternalistic in a stronger sense, but my argument about sacrifice and moral obligation does not have any particular implications about whether states should coerce citizens in this way.

Returning to The Puzzling Data

Recall that when I first introduced the notion of sacrifice, I posed the following problem: if feeling something *as* a sacrifice *makes* it a sacrifice, then people could claim it was a sacrifice not to have a brownie sundae every hour, and this seems absurd—especially if sacrifice is thought to mitigate moral obligation. On the other hand, if something is a sacrifice regardless of whether it is *felt* as a sacrifice, then people for whom it is just no trouble to, say, sleep in a hut with no hot water, could get immense amounts of moral merit for doing things that come easily to them. After offering an account of sacrifice that is indexed to an objective, agent-neutral theory of well-being, I've come to believe that the first absurdity is much more significant than the second one. If some things that meet the definition of a sacrifice turn out not to be all that *irksome* to

the agent who bears them, that's OK. What is good for an agent needs to be able to come apart from what she takes to be good for her, or what makes her feel good, because well-being should be understood independently of the agent's point of view. This needs to be the case so that we can understand the idea of self-sacrifice.

Of course, the fact that similar levels of sacrifice produce different levels of *irksomeness* in different agents is not something we want completely to ignore. It might play an important role in a psychological explanation for how the actions of people like Susan Tom and Paul Farmer come to be. And it might explain the "puzzling data" we began this section with—the data that agents who perform intuitively supererogatory actions sometimes describe their actions as obligatory. Perhaps the action seemed obligatory to the agent because the *felt* aspect of the sacrifice was far milder than the actual loss in objective well-being. Indeed, perhaps this is how moral boundaries are extended: by demonstrating to the rest of us just how much sacrifice can be endured before an agent becomes burned out or self-defeating, moral saints influence our intuitive judgments about where to draw the line between the obligatory and the supererogatory. We might still think that levels of sacrifice, indexed to an objective theory of wellbeing, have a mitigating effect on obligation, but think that the line where this mitigation goes into effect (the line between obligation and supererogation) is moving. This is what I want to explore in the next chapter.

CHAPTER 5

THE RATCHETING-UP EFFECT

In this chapter I explore how the boundary between obligation and supererogation might be flexible, and how the behavior of moral saints affects where the boundary lies for the rest of us.[38] We tend to think that there is some amount of hardship or sacrifice that we just cannot expect people to bear for the sake of morality. James Fishkin (1982) calls this the "cutoff for heroism" (5). But what exactly is that cutoff, and could it change? After all, if certain psychological and sociological facts were different, the boundary could easily have ended up much lower or higher than it is. If getting your shoes wet were a grave hardship, then perhaps we would be calling those willing to step in puddles to rescue someone's dropped wallet "moral saints." Indeed, we might think what passes for sacrifice and hardship among well-off members of wealthy countries is already severely distorted. There is an amusing display of this phenomenon in a recent Roz Chast cartoon in *The New Yorker*.[39] The cartoon is titled "Latter-Day Saints" and the picture shows several people out and about on the street in New York City, each with a

[38] Not everyone agrees that it makes sense to think in terms of this boundary. In "Above and Below the Line of Duty" (1986), Susan Wolf claims that the concept of obligation (or duty) has had too much influence on our thinking about morality. "There is a line of duty," she writes, "but it is, necessarily, a dotted line" (32). In recent work on "Deontic Pluralism" (in progress), Margaret Little argues that we need many more concepts than just the classic obligatory, forbidden, permissible, and supererogatory.

[39] Roz Chast, "Latter Day Saints." *The New Yorker*, October 16, 2006.

little bubble describing the "sacrifices" they are making—not necessarily for the sake of morality, but perhaps for the sake of their own self-improvement or the supposed welfare of their families. "Eats however many grams of fiber you're supposed to" reads the bubble next to a woman, and next to her young son it says, "Is putting up with a 56K modem." One of the men in the street "does all the grocery shopping." The other people in the picture bear similarly weighty burdens: "Has commuted from New Canaan every day for 32 years"; "Jogs ten miles a day, no matter what"; "Buys her shampoo in an ordinary drugstore"; "Doesn't own a TV"; "Forcing self to read 'Beowulf,' although she hates it."

One reason Chast's cartoon is funny is that, though we can sympathize with the tedium of commuting for 32 years, we don't *really* think it makes you a saint. But we *are* rather easily impressed. The cartoon pokes fun at our tendency to valorize people for doing pretty much anything that they would rather not do. It's not that we think "putting up with a 56K modem" involves a particularly significant loss of objective well-being; it's that we think putting up with it is irksome—it *feels* like a sacrifice in light of our high expectations. Perhaps that is why we are so awed by people like Paul Farmer and Susan Tom: for those of us who can't handle a slow internet collection, what *they* do is almost *unimaginable*. We are shocked by their lifestyles, and this shock could send us in either of two directions. We might think that they are just fundamentally better people than we are and we could never do what they do; there is no hope for us. On the other hand, we might see their lifestyles as evidence that we could do *more*, and that it might not be as hard as we thought. I shall argue that the latter is the more likely and fitting reaction, and that this has implications for what we are required to do. This isn't just an

epistemological claim about how moral saints influence our beliefs about our obligations. It's actually a bit more ambitious than that—it's a claim that the epistemology of the moral boundary could actually change the *metaphysics* of the boundary. What we're obligated to do depends in a subtle way on what we *believe* about what we're obligated to do. I spell this out in more detail below.

When Moral Obligation Is (and Isn't) Affected by What Others are Doing

Moral saints fill an odd niche in the moral ecosystem. In a way, they are like explorers or record-setters, defining the boundaries of moral behavior—both in terms of the impact one person can make through sheer effort, and in terms of the burden one person can bear without crumbling under the weight. So it might seem perfectly uninteresting to show that their extraordinary lives could influence our own obligations. However, it is actually rather common to think that what others are (or are not) doing cannot have any effect whatsoever on what I *ought* to do (as opposed to, say, what I am *likely* to do). It is often thought that our obligations are what they are, regardless of whether other people are behaving well or badly.

David Estlund (2007) thinks moral philosophers differ from political philosophers with respect to this point. "Moral philosophers know that people are likely to lie more than they morally should," he writes, "but this doesn't move many theorists to revise their views about when lying is wrong. Things are often different in political philosophy" (12). The moral philosophers Estlund mentions are no doubt correct about a certain class of moral judgments. If, for instance, it is wrong to keep the extra $10 a cashier gives you by mistake, then surely it is wrong regardless of whether 87% of people would keep the

money. And if you really are morally obligated to get a bike helmet for your child, surely you are obligated even if only a few of the other parents are doing the same.

So to a certain extent, it seems that our considered judgments about whether an action is on one side of a moral boundary or the other are independent of sociological facts about what most people believe or how most people behave. And this is how it *should* be, assuming there is a fact of the matter—at least in certain fixed contexts—about what one ought to do in those cases. Our judgments should track the fact of the matter even if that means not tracking what most people believe or do. Nevertheless, in some domains things are more complicated. The examples above were perhaps misleadingly simple. The consequences of *my* not returning the extra $10 are unaffected by whether anyone else does the same when they are given extra money. If I don't give the money back, the cash register will come up short, and no less short on account of the fact that I'm doing something fairly common. The cashier may be blamed for the shortage, may even be thought to be a thief. And, consequences aside, I am still *taking something that is not mine* and *being deceptive and disingenuous* in not revealing the mistake—my behavior has these morally troublesome features regardless of what other people are doing. The same goes for the bike helmet. If I don't get a helmet for my child, I'm putting her at risk, and it's the same risk regardless of whether other kids are wearing helmets (unless we're concerned about head-to-head collisions!).

But there are many situations where what other people are doing might directly affect what morality demands of each person. These include aid scenarios and collective action problems, where the number of people helping might determine how much each person needs to contribute, or where the noncompliance of others might render

meaningless the compliance of some (say, in walking-on-the-grass scenarios). In fact, philosophers have taken very seriously the idea that the compliance or noncompliance of other people might affect what a single person is obligated to do. The difficulty is in figuring out whether others' noncompliance means that I ought to do *more* or that I am permitted to do *less*. Both positions have some inherent plausibility. Some might think, "Since no one else is donating to famine relief, surely *I* don't have to—why should I hold myself to a higher standard? At most I ought to give only what my fair share would be if everyone were donating." However, others might think, "Since no one else is donating to famine relief, surely I ought to donate *as much as I can*, because everyone else has left me with a greater need to fill. The less they give, the more I must."[40]

Recently, Tamar Schapiro (2003) has considered this question from a deontological perspective. She argues that there are certain kinds of activities in which "our actions depend *constitutively* on others' compliance with the practice, so that beyond a certain threshold, others' wrongdoing can alter the character of what we ourselves are doing […]" (333). In such a case, the fact that others are acting wrongly has "the effect of mitigating the stringency with which moral principles apply to us" (329). At the risk of oversimplifying Schapiro's subtle argument, the upshot is that sometimes morality

[40] See Singer (1972) and Unger (1996) for examples of the view that one should give as much as one can. See Schapiro (2003) for the view that we can sometimes do less, and Murphy (2000) for the view that we should only do what would be our fair share under full compliance. Recently, Michael Ridge has articulated a compromise position in "Fairness and Non-Compliance" (forthcoming). Ridge claims that we overemphasize the unfairness to the compliers in situations of partial compliance and neglect the unfairness to the persons who are on the *receiving* end of the aid, who are forced to bear an unfair share of the burden. He argues that the burdens should be distributed fairly amongst the compliers, and between the compliers and the recipients, which means the compliers may have to give more than they would in a situation of full compliance, but less than they would if their obligation was completely unconstrained. Thus Ridge demands less than Singer and Unger but more than Murphy.

requires *less* of us on account of the fact that others are unscrupulous. Thus our "pragmatic" intuition, that sticking with our "high-minded standards" would be "naïve at best," wins out over the "purist" intuition that we ought not "stoop to their level" because "two wrongs…just cannot make a right" (329).

Liam Murphy (2000) considers this same question from a consequentialist perspective. When it comes to a principle like *optimizing beneficence*—which requires that we keep giving to those in need until the point of marginal utility—it seems that noncompliance poses a particularly tough problem. Given that so many people are not complying, it would seem that the burden on those who *do* comply becomes shockingly high when there is great need. This is known as the problem of "overdemandingness," and it is thought to be a good objection to the principle of optimizing beneficence. Murphy doesn't actually think the traditional overdemandingness objection, understood as the charge that a theory makes "extreme" or "excessive" demands, is coherent (39). But he nevertheless thinks that optimizing beneficence fails for a closely related reason, namely that it "imposes its demands *unfairly* in situations of partial compliance" (7, emphasis mine). Ultimately, he instead endorses a principle of *collective* beneficence, where what we are obligated to give under conditions of noncompliance is just whatever we would be obligated to give if everyone else were complying. In other words, one must give only one's "fair share."

Schapiro and Murphy both take on the question of how other's behavior might affect our obligations, and, again at the risk of oversimplification, both of them conclude something like this: the less-than-scrupulous behavior of others can have the effect of *lessening* what you are required to do (Shapiro), or at least not increasing it (Murphy).

The argument that I shall offer differs from Murphy and Schapiro in several respects. They each approach the question from the perspective of a particular moral theory; in contrast, I am interested in the line between obligation and supererogation as it appears in common-sense morality. Moreover, both of them are interested in cases where most of the people who are subject to a *particular* moral demand fail to comply with it. I am interested, rather, in a more general phenomenon—that is, what happens when a small number of people do a lot more than the rest of us, not in any particular domain or with respect to any particular duty. Indeed, my argument focuses not on *noncompliance*, but on *over*compliance. For we are no longer talking about people failing or succeeding in doing their duty, but rather about people *exceeding* their duties. And I am concerned not with specific practices (say, famine relief or voting or walking on the grass) but with life more generally. Finally, the most significant difference between my argument and those of Murphy and Schapiro is that I conclude that the behavior of moral saints (who are in a way the complement of the noncompliers) serves to *increase* our moral obligations, not lessen them.

Before laying out my argument for this claim in detail, it's worth taking a brief detour to see what Kant has to say on this subject. While Kant is a big fan of the idea of appealing to moral role models, he seems to think moral saints would be the *worst* role models. We need to address his worries before proceeding.

Kant on Moral Education

Kant has some very interesting things to say about moral education in "The Doctrine of the Method of Pure Practical Reason" in the *Critique of Practical Reason*

(1997). He claims that we ought to use case studies as part of moral education, but when it comes to just how exactly we ought to pick the case studies, and what lessons we should learn from them, I think he gets it precisely backward. He first points out that it can be great fun, even for the least scholarly of us, to argue about the moral character of various real or fictional persons:

> If one attends to the course of conversation in mixed companies consisting not merely of scholars and subtle reasoners but also of business people or women, one notices that their entertainment includes, besides storytelling and jesting, arguing; […] Now, of all arguments there are none that more excite the participation of persons who are otherwise soon bored with subtle reasoning and that bring a certain liveliness into the company than arguments about the *moral worth* of this or that action by which the character of some person is to be made out (126).

Given these sociological facts (which I believe are still true today minus the 18th-Century chauvinism and condescension), Kant argues that educators should make use of vignettes about moral character in order to facilitate the moral education of children. The kind of moral education in question is not the philosophical study of moral theory, but rather education concerning what things are in fact right and wrong—education designed to transform the pupils into virtuous moral agents. This can be done by way of

> […] a game of judgment in which children can compete with one another, yet will leave behind a lasting impression of esteem on the one hand and disgust on the other, which by mere habituation, repeatedly looking on such actions as deserving approval or censure, would make a good foundation for uprightness in the future conduct of life (127).

Apart from the fact that this sort of moral education sounds rather much like indoctrination, I have little quarrel with Kant so far. My own view begins to depart from his in the course of what he says next.

On the question of just what kinds of "biographies" we ought to use for this moral education, Kant says the following:

> But I do wish that educators would spare their pupils examples of so-called *noble* (supermeritorious) actions, with which our sentimental writings so abound, and would expose them all only to duty and to the worth that a human being can and must give himself in his own eyes by consciousness of not having transgressed it; for, whatever runs up into empty wishes and longings for inaccessible perfection produces mere heroes of romance who while they pride themselves on their feeling for extravagant greatness, release themselves in return from the observance of common and everyday obligation, which then seems to them insignificant and petty (127-128).

If noble or "supermeritorious" actions are roughly equivalent to the kinds of actions *moral saints* perform, then Kant thinks we should not use moral saints as examples. His worry is a familiar one even today: that we ought not "let the perfect be the enemy of the good." If we give children examples of moral perfection, the reasoning goes, they will become "heroes of romance." Lofty ideals will distort their thinking so much that they will take themselves to be exempt from mundane duties. Kant later repeats the point by saying that showing children examples of moral saints is "contrapurposive," because it will make them "fantasizers" (129). It sounds like Kant doesn't want kids to get what we sometimes refer to as "big ideas."

Of course, it's not clear just why exactly a pupil would, on the basis of having adopted certain lofty ideals about morality, decide that he was no longer interested in, or in fact no longer *bound* by, everyday moral duties. If we could get children to be

committed to "greatness" and to admire great moral figures, wouldn't this raise morality to salience in a way that would make them more cognizant of even the most banal duties of daily life? The problem here is that in condemning the use of "noble" figures in education, Kant is not just making a point about the *magnitude* of their moral worth—that they are *too good* for emulation; rather, he is making a point about the *nature* of their moral worth—they are not good in the right *way*. For Kant is using "noble," "meritorious" (or "supermeritorious"), and "magnanimous" to refer to a kind of touchy-feely disposition that is precisely *not* an orientation toward *duty*. Since moral education should focus primarily on inculcating a sense of duty, such education should ignore these touchy-feely noble agents found in the "sentimental" literature. Indeed, in a footnote Kant concedes that while we should praise "actions in which a great, unselfish, sympathetic disposition or humanity is manifested," in the course of educating children we should emphasize the "*subjection of the heart* to *duty*" rather than the "*elevation of soul*" (128).

Indeed, Kant's point becomes clearer if we think in terms of the distinction Susan Wolf (1982) makes between the "rational saint" and the "loving saint." The rational saint acts out of a sense of duty; he "sacrifices his own interests to the interests of others, and feels the sacrifices as such" (420). The loving saint helps others because his happiness "would truly lie in the happiness of others, and so he would devote himself to others gladly, with whole and open heart" (420). Not surprisingly, it seems that the kind of "noble" and "supermeritorious" person Kant thinks would be a bad role model is more like the loving saint. To use a loving saint as an exemplar is to send a mixed message, he might say, because we would be preaching respect for the moral law for its own sake while drawing attention to agents who might seem to be driven by happiness instead.

Kant thinks that morality will "have more power over the human heart the more purely it is presented" (129). Presenting morality "purely" means not allowing pupils to see any *incentive* in acting morally, especially not the incentive of happiness. In fact, Kant thinks the best way to present "the law of morals and the image of holiness" is "*in suffering*" (129, emphasis added).

Kant is making a very important point here. When it comes to moral education, we surely do not want children to infer from the fact that acting morally might make you happy to the fact that the reason one *ought* to act morally is (solely) that it will make you happy. So showing them examples of super-happy noble agents might be misleading. That said, perhaps things are more complicated than Kant lets on. For one thing, though the idea of "rational saints" and "loving saints" (on which Kant seems to be implicitly relying) may be theoretically useful, I do not think there is anything so neat as this distinction in real life. Consider Paul Farmer and Susan Tom. It may be that Farmer is a bit closer to the "rational saint" pole, and Tom is closer to the "loving saint" pole, but I don't think you would know this by observing them. Both of them have love for the people they help. Both feel morally obligated (if not *de dicto*, certainly *de re*) to do what they do. Both get happiness from their work, but in both cases it also brings them great pain. Both make great sacrifices, but in both cases the *felt sacrifice* is probably considerably lower than the actual loss of well-being. They seem like great role models of moral worth, and yet they don't quite display the "suffering" that Kant desires in his examples. They give up quite a bit to do what they do, and there is occasional suffering to be sure, but they also reap great rewards. As I argued in Part I, the fact that they reap

rewards does not diminish the force of their sacrifice, but it does make their lives appear less "pure" than the exemplars of duty that Kant seeks.

And what should we make of Kant's worries about the dangers of seeking perfection? Is the great really the enemy of the good? For one thing, as I have argued elsewhere, I don't think the right way to think about moral saints, even in theory, is as moral perfectionists. And of course the people we find in real life who are moral saints, or near enough, are not morally perfect nor do they seek to be. It's unlikely that any of the "biographies" Kant's moral educators actually have at their disposal depict real moral perfection. But setting aside questions about perfection, Kant could still argue that we ought not take people quite so extraordinary as Paul Farmer as our role models. There are really two worries here. First, when faced with someone like Paul Farmer, it's easy to get disillusioned and throw your hands up—I can never be like him, why bother trying! Or, as Kant worries, you might get so seduced by the idea of saving the world that lesser moral obligations seem trivial in comparison and no longer command any attention, or you give yourself a "pass" on these smaller duties because you think your bigger pursuits make you a good enough person overall.

Ought we to take these worries seriously? Surely it is an empirical question whether people really would have the reactions Kant fears, and I have no data one way or the other. But it seems that reactions of that sort would be dwarfed by another kind of reaction, namely, *wow, if Paul Farmer can do that, maybe I can too, perhaps not as well as he does, but better than I am doing now*. After all, in almost any other realm—sports, art, science—when we want to know how to do something better, we ask an expert. The fact that some experts (Michael Jordan, Leonardo DaVinci, Einstein) seem to have

"natural talent" doesn't deter us from thinking that, by studying and emulating them, we might get better ourselves.

But we have gone slightly astray. For while Kant's focus here is on explicit "moral education" of the sort involving students and teachers and the direct aim of improving moral character, what I am interested in is somewhat different. While I do think that moral saints have an important place as role models in moral education, what I'm interested in here is not how the behavior of moral saints might affect how other people *do* act, but rather how the behavior of moral saints might actually effect how other people *ought* to act. These are related, though. Both involve what people *believe*. Whether I am obligated to perform a given action depends partly on whether it is reasonable for me to believe that I ought to do it. Moral saints, I argue, provide us with crucial *evidence* in light of which our beliefs, and consequently our obligations, might change.

The Argument for Ratcheting

Part of what differentiates the *good and required* actions from the *good but not required* seems to be that the latter involve a greater degree of sacrifice. As I argued in Part I, the kind of sacrifice that's relevant here is agent-neutral and objective; an agent can be mistaken about how much of a sacrifice it would be to give something up, and thus mistaken about whether she is obligated to give it up. Thus an agent's *beliefs* about how bad it would be for her to give something up do not play a *direct* role in determining whether she is in fact obligated to give it up, because sacrifice is not, so to speak, in the eye of the beholder. Nevertheless, I now wish to argue that the agent's beliefs about sacrifice do play an *indirect* role in determining what her obligations are, because

obligations are constrained by what we might call a "knowledge condition." The knowledge condition can be formulated roughly as follows:

Knowledge Condition (K). Someone can be obligated to ϕ only if she could know (or it would be reasonable for her to believe) that she is obligated to ϕ.

Throughout this chapter, we have also been assuming that obligation is subject to another constraint, a constraint on reasonable sacrifice. Let's now call this the "sacrifice condition." It can be formulated as follows:

Sacrifice Condition (S). Someone can be obligated to ϕ only if ϕ-ing does not involve an unreasonable sacrifice.

If the Knowledge Condition is correct, then moral obligation is dependent in an important way on our evidence. This suggests a refinement of the Sacrifice Condition. Call this refinement K-S, the Knowledge of Sacrifice Condition.

Knowledge of Sacrifice Condition (K-S). Someone can be obligated to ϕ only if she could know (or reasonably believe) that ϕ-ing does not involve an unreasonable sacrifice.

Admittedly, K-S does not follow straightforwardly from K and S alone. One would presumably need to add an intermediate principle along the lines of "someone can *know* that she is obligated to ϕ only if she could know (or it would be reasonable for her to believe) that ϕ-ing does not involve an unreasonable sacrifice." Without getting too bogged down in the constraints on knowledge—what is really of interest to us here is

constraints on *obligation*—I think this intermediate principle is as plausible as the sacrifice condition is.

We can begin to see how the Ratcheting-Up Effect would work: if K-S is true, our obligations are constrained by what we could reasonably believe about how much of a sacrifice it would be to perform a given action. And what we can reasonably believe about this level of sacrifice is surely a function of what sort of *evidence* is available to us. When the lives of moral saints are publicized, our pool of evidence about sacrifice changes. To be sure, the behavior of *everyone*—not just moral saints—contributes to our stock of evidence. But moral saints play a unique role, both because they are rare and because they behave at the margins of obligation, taking on unusual sacrifices. In short, the argument looks like this:

(1) Moral obligation is at its core about what we can reasonably demand of each other.

(2) We cannot reasonably demand that someone do something that it would be reasonable for her to believe to be too much of a sacrifice (the K-S Condition).

(3) Exposure to moral saints can change what it would be reasonable to believe about how much of a sacrifice it would be to take on certain actions or patterns of behavior.[41]

(4) So, exposure to moral saints could change what a person is obligated to do by way of removing a defeater of obligations.

Much of the argument rests on the Knowledge Condition (and its derivative K-S). We ought to look more closely at this condition, since it is not uncontroversial.

[41] As I will explain below, it makes no difference if the moral saint's action is in fact supererogatory, because there will be some lesser action that involves a proportionately lesser amount of sacrifice, which itself is not supererogatory, and which the observer will come to be obligated to perform.

The Knowledge Condition says that, if it is reasonable for someone to believe that she is not obligated to perform an action, then she is not obligated to perform the action. At first glance, this may seem too permissive. Ignorance about morally relevant facts is fairly common, so one might worry that the knowledge condition releases people from too many obligations. Yet this worry is misguided. For the knowledge condition only governs cases of what we might call *justified ignorance*; it does not govern just *any* case of ignorance. It trades on what it would be reasonable for us to believe, not on what we actually believe. Some people might believe that beating their children "builds character," and that parents are not obligated to refrain from beating their kids. But surely it is not *reasonable* to believe either of these things, at least not in the modern-day epistemic environment.

According to this way of thinking about moral obligation, *what I am obligated to do* simply amounts to the same thing as *what I am obligated to do in light of my justified beliefs*.[42] We can see the motivation for this principle if we look at a puzzle about utilitarianism. According to act utilitarianism, the right action is the one with the highest utility compared to the alternatives. But is the agent morally obligated to perform the action with the highest *actual* utility, or the action with the highest *expected* utility? Suppose that John can donate to Charity A or Charity B, and the best available evidence

[42] In "Justified Wrongdoing" (1997), Sarah Buss argues that "an agent's blameworthiness is a function of what he can reasonably be expected to know" (338). According to Buss's account, justifiable ignorance can result in blameless wrongdoing. Whereas Buss's account focuses on *wrongdoing*, my account focuses on *obligation*. I argue that a certain type of justifiable ignorance does not merely free a person from blame, but in fact frees him from obligation. His action may remain "wrong" in some sense. But it is *obligation*, and not wrongness, that best captures the type of normative expectation of behavior that members of the moral community can hold each other accountable for.

indicates Charity A will produce more utility. It turns out, however, that Charity A is actually going through an unpublicized crisis and John's money will go further at Charity B. Despite this fact, it seems that the most plausible theory of moral obligation is the one that says John should maximize *expected* utility, and thus he was obligated to donate to A. Indeed, it would not even make sense for the moral community to demand that John maximize *actual* utility, since the only way to do so would be to aim for it via *expected* utility. If obligation were determined by the actual utility of the consequences regardless of the agent's beliefs and expectations at the time of action, then we would only be able to determine the agent's obligations either retroactively or from an omniscient perspective. Of course, we can still say that given how the things turned out, John "ought" to have chosen B rather than A, meaning that it would have been better if he had. But the "ought" here is not the same sense of moral obligation that is grounded in the demands we justifiably make of one another. For how could we be justified in making demands on others that they will be systematically incapable of living up to on account of their limited information?

Indeed, this would seem to be so unfair that philosophers have proposed simply separating the "criterion of rightness" in utilitarianism from the "decision procedure."[43] The idea here is that what is "right" is a matter of the *actual* consequences, but what an agent "ought" to do is follow the most rational decision procedure—such as maximize expected utility. So John did the "wrong" thing, but given his limited information, we cannot blame him, and since he is morally obligated to act using a particular decision procedure, he lived up to his moral obligation. Notice, however, that when we choose to

[43] See, e.g., Eugene Bayles (1971).

separate what's "right" from what's obligatory in this way, "rightness" is no longer particularly connected to what we can justifiably demand of one another. And *that*—what we can justifiably demand of one another—seems to remain the key moral concept governing how we can live together in a moral community. So even if we cordon off a concept of "rightness" that is not subject the Knowledge Condition, it seems that the concept of moral *obligation* still is.

Now, worries remain about the Knowledge Condition. Even granting that the kind of moral obligation in question is a matter of what we can justifiably demand of one another, there is still room to deny that the metaphysics of obligation is dependent on the epistemology of obligation in the way the Knowledge Condition lays out. In particular, one might worry that ignorance never alters what we are in fact *obligated* to do, it only alters what sort of behavior we can justify, or what sort of behavior we judge praiseworthy or blameworthy. The idea here is that our evidence has *practical* consequences about what we can hold others responsible for, but it has no *metaphysical* consequences about what our obligations are. Just as, in science, we might distinguish the true theory from the theory we are justified in believing given our evidence, here we might distinguish what is genuinely the obligatory action from what is the justifiable action given our evidence.

But how can what a person is "justified" in doing come apart from what she is obligated to do, if moral obligation really is a matter of what we can reasonably demand of one another? It seems to me that what we can reasonably demand of you *just is* whatever you can be justified in doing in light of your information. A stronger interpretation of reasonable demands—according to which we could reasonably demand

things of you even if you could not know that you ought to do them—would seem to strain the credibility of the term "reasonable."

Consider an example: Presumably there was a time—if not in the 20th century, then surely in, say, the 17th century—when it was entirely reasonable to believe that second-hand smoke had no negative health consequences. There simply was no available evidence to the contrary. Surely it would be incorrect to say that people in 1650 were *obligated* not to smoke around their kids, but not blameworthy for failing to meet this obligation. Instead, we ought to say that there was no obligation in the first place.

Another way to frame the "no metaphysical consequences" worry is to make a distinction between a *justification* and an *excuse*. As before, I will use "justification" to mean a consideration that actually *releases* an agent from a given obligation. In this sense, justifications genuinely affect the metaphysics of obligation, whether by eliminating or nullifying an obligation. An "excuse," on the other hand, is simply a consideration in light of which the agent is not blameworthy for failing to heed the obligation. When an excuse is present, the metaphysics of the obligation remain unchanged; all that changes is whether or not we can blame the agent. For example, if I'm obligated to meet you for lunch, but on the way to meet you I encounter a severely injured person who needs my help, we might say that I am *justified* in missing lunch—I am no longer obligated to be there, or at least my obligation does not have its original force. I have done nothing wrong. However, if I simply forget about our lunch, not from carelessness but because my mind has been preoccupied with something important lately, we might say I am just as obligated to be there as ever, but since it was an innocent mistake, I have an *excuse*—I am not blameworthy. With this terminology in place, we

might formulate the objection to the Knowledge Condition this way: ignorance of our moral obligations is never a justification, but it may sometimes be an excuse. In other words, there is no Knowledge Condition on moral obligation, only on moral blameworthiness. Agents living in 1650 *were* morally obligated to protect their kids from second-hand smoke; they just didn't realize it. Their behavior was not *justified*, but they did have an *excuse*. While the obligation was there all along, the agents were not blameworthy for failing to heed it.

But I don't think that this language of "*justifications"* and "e*xcuses"* makes the objection any more compelling. Intuitions may differ on whether reasonable ignorance of one's purported obligations is more like the case of the preoccupied person who simply forgets her lunch date, or more like the person who encounters an emergency along the way. But on the account I've given, moral obligation is fundamentally about what we can reasonably demand of one another. Just as it would be unreasonable for us to demand that you show up for lunch dates when an emergency arises, and just as it would have been unreasonable to demand that 17th-century parents worry about second-hand smoke, it seems that it would be unreasonable to demand that members of the moral community perform actions even when their evidence suggests that such actions are not obligatory.

Yet while I do not find the "No Metaphysical Consequences" objection compelling, it turns out that this objection would not be devastating to the Ratcheting-Up argument even if it were true. After all, the Ratcheting-Up Effect could exist even if only *blameworthiness* were subject to the knowledge condition. What we can be blamed for would be *ratcheted-up* by the behavior of moral saints. This in itself would be a normatively significant result, even if it did not constitute a change in the underlying

metaphysics of obligation. Still, the notion of moral obligation I've put forward is closely tied to blameworthiness; on my account, we could not ratchet-up what we could *reasonably blame* people for without also ratcheting-up what we can reasonably *demand of* people, that is, without ratcheting up the obligations themselves.

Evidence and Ratcheting

It remains to be seen just how the Ratcheting-Up Effect works. For it seems that moral saints provide us with two kinds of evidence: evidence about *felt* sacrifice and evidence about sacrifice proper (that is, gross losses of well-being). Both kinds of evidence are imperfect. What we can learn about felt sacrifice by reading about Paul Farmer or even meeting him is only as useful as *anything* we can learn about another person's subjective experiences. But we can get clues from his expressions, his behavior, and from what he tells us. We know that he can't sleep, that he would rather see more of his family, and that he would rather pack more than three shirts in his suitcase (if only he didn't need room for all the gifts he is asked to carry). We know that his work makes him angry and sad, and that he feels guilty for being associated with a wealthy nation that does not do nearly enough to help the poor. We know that often he's uncomfortable, hot, jetlagged, hungry, and overworked—and occasionally ill with an exotic bug. But we also know that most of these worries probably do not surface on a moment by moment basis. He's too busy.

We know something about the psychic *rewards* of Farmer's work, too. Call these *felt* rewards, the counterpoint to felt sacrifices. We know he enjoys seeing his patients, especially when they are getting better, as they often do. We know he likes the

intellectual challenges of his job. We know he likes the flexibility of his unique lifestyle: being able to spend a few months working with experts in Boston, and then a few months in a rural area treating patients, and then a few months traveling the world giving lectures.

What we know about Farmer's *actual* sacrifices is imperfect as well, for a proper accounting of how his various losses stack up against the items in our objective list theory, including assigning relative weights to the various types of losses, would surely require knowing him pretty well. (Recall that even though our list is objective and agent-neutral, it is responsive to the details of individual situations in that we interpret various losses as falling under the categories on the list in light of their roles in a person's life. Access to TV isn't on the objective list, but Raymond's TV in *Rain Man* surely contributes to his well-being. Knowing this requires knowing Raymond relatively well.) Still, we can see quite clearly that Farmer incurs several gross losses of well-being as were cataloged earlier under "opportunity costs," "personality changes," etc.

The question before us is how this evidence from a moral saint—of felt sacrifice and sacrifice proper, both gross and net—bears on observers' beliefs about their *own* obligations. As I argued in Part I, observers are frequently better positioned to judge *objective* levels of sacrifice than agents, who tend to be preoccupied with *felt* sacrifice. (Moral saints might be willing to overachieve, in part, because it doesn't *feel* that bad to give up what they have given up, and moral underachievers might be failing to live up to their obligations, in part, because sometimes it *feels* really bad to give up things that are not, in fact, significant sources of well-being, objectively considered.) I also argued in Part I that it is *objective* levels of sacrifice that bear on obligation; whether an action is

obligatory or supererogatory is directly a function of the gross losses of well-being, considered objectively, involved in performing it. In Part II, however, I have argued that there is another route by which sacrifice influences obligation, this one an *indirect* route: what we reasonably *believe* about how much of a sacrifice it would be to perform a given action bears on whether we are obligated, due to the Knowledge-of-Sacrifice Condition on obligation.

Now, presumably when a non-philosopher contemplates whether a given level of sacrifice is a *reasonable* or *bearable* amount, she does this via some combination of the notions of *felt* sacrifice and sacrifice proper. Certainly, when considering Paul Farmer's lifestyle, the non-philosopher does not ask (at least not *de dicto*) "Is a lifestyle like his what those who care about me would be rational to want for me for my own sake? How does his lifestyle stack up against the objective list?" Nevertheless, I think it's clear that observers regularly assess objective losses of wellbeing in addition to felt sacrifice—how else would they come to conclude that Paul Farmer and the Holocaust rescuer are incorrect when they judge their own actions to be obligatory rather than supererogatory? Indeed, it seems that an observer will be mostly likely to look for gross losses of objective well-being when there is an intuitive gap between what the agent seems willing to endure and what the observer herself would be willing to endure. In such cases, it becomes clear that one cannot read off how much the agent has given up from how bad it *seems* to the agent, because there is a discrepancy between the agent's felt sacrifice and his actual sacrifice. For instance, Paul Farmer might seem to be less bothered than the many Americans would be by some of his lost creature comforts. As Kidder writes, "Like all his siblings, he emerged from the bayou's waters with what he called a 'very

compliant GI system,' and from dinners of hot dog-bean soup without much fussiness about food, and from years of cramped quarters with the ability to concentrate anywhere. He could sleep in a dentist's chair, as he did at night for most of one summer…" (54). Reading this, an observer who would himself find it difficult to sleep in a dentist's chair all summer might find it necessary to think harder about what has been given up, since Farmer's level of felt sacrifice is perhaps a bit idiosyncratic.

There is a trace of Kant's worry here; if the moral saint appears too unusual, he becomes less useful as a role model ("but *he's* willing to sleep in a *dentist's chair*!"). If this worry is correct, a moral saint who appears to have an unusually high tolerance for hardship will be less likely to contribute to the Ratcheting-Up Effect. In practice, I don't think this would be a problem for Paul Farmer. In reading about him, one is more often surprised at how *normal* he appears, given what he accomplishes, than at how strange he is. (He curses; he makes fun of people. He likes bad action movies. He cares more about his own daughter than his patients.) It would be even less of a problem with Susan Tom. As it turns out, though, this worry may have little traction in theory as well as in practice. For the mechanism underlying the Ratcheting-Up Effect need not be "Paul Farmer sleeps in a dentist's chair, so I ought to sleep in one as well" but rather "Paul Farmer sleeps in a dentist's chair—surely I can go without purchasing this luxurious new bedding." Even if the observer concluded that sleeping in a dentist's chair was an unreasonable sacrifice (and hence any action that required it would be supererogatory), the fact that Paul Farmer was willing and able to bear this sacrifice would be evidence that it was at the very least not an *unbearable* sacrifice. By testing the boundary of what sorts of sacrifices are

bearable, someone like Paul Farmer challenges the rest of us to recalibrate our notions of bearability and reasonability.

To sum up, the Ratcheting-Up argument goes like this: (1) Moral obligation is at its core about what we can reasonably demand of each other. (2) We cannot reasonably demand that someone do something that it would be reasonable for her to believe to be too much of a sacrifice (the K-S Condition). (3) Exposure to moral saints can change what a person believes about how much of a sacrifice it would be to take on certain actions or patterns of behavior. (4) So, exposure to moral saints could change what a person is obligated to do by way of removing a defeater of obligations.

It remains somewhat unclear just how (4) is supposed to follow from (1)-(3). By what mechanism does information about the behavior of a moral saint—which by stipulation is, in general, *supererogatory* behavior—become evidence that is relevant to our obligations? How does information about the sacrifices involved in *saintly* behavior bear on what the rest of us *ought* to do, since presumably it is not the case that we *ought* to be moral saints? In response to this worry I think it might be helpful to consider two possible ways in which the behavior of moral saints might constitute *evidence* that could plausibly be useful for the rest of us. Both ways depend on the general principle that saintly behavior can be relevant to required behavior in the following way. Put informally, the observer thinks, "Wow, I thought doing as much as Paul Farmer does, which is clearly not required, would be *terribly* burdensome. But he gets along pretty well; his life is not as miserable as I would have thought. If *his* level of moral action involves only *that* level of sacrifice, then surely a much more modest level of action would involve even *less* sacrifice. Maybe I can do more than I realized." It is by this sort

of reasoning that evidence about the sacrifice levels involved in supererogatory behavior might be thought to bear on the sacrifice levels even of potentially *obligatory* behavior—that is, behavior far down the scale from what Farmer and Tom are doing.

Below are two possible ways of understanding how this ratcheting mechanism might work.

CASCADING RATCHET. Suppose that I am giving $100 a year to global public health organizations, while Paul Farmer is working twenty-hour days and sleeping in a hut to treat as many patients as he can. It might seem that Farmer's actions are too different from my own, and too alien to my lifestyle, to be particularly good evidence about what it would be like for me to do a little more. If this is right, then one way to understand ratcheting is as a cascading effect: Farmer ratchets up the people who are doing just a little bit less than him, and they ratchet up the people just below them, and so on, all the way down, so that Farmer's efforts have an *indirect* influence on me and the other $100 donors. The appeal of the cascading ratchet is that it seems to require fewer leaps of faith when it comes to seeing the sacrifices someone else makes as bearing on the sacrifices one might make oneself. Knowing that daylong hikes in the mountains to make housecalls can actually be quite refreshing, or that sleeping in a hut is not so bad after all—knowing these things would seem not to be of much use to the person who is thinking of going from a $100/year contribution to $500/year, or thinking of volunteering at a local blood drive. However, this knowledge might be relevant to someone who already works in Farmer's field and is considering moving to a remote location to take on more work. Then *that* person's sacrifices could be relevant to people one level further down, who want to spend their two-week vacation at a rural clinic. And so on.

HIGH-TO-LOW RATCHET. I think that the Ratcheting-Up Effect actually works more directly than the Cascading Ratchet suggests. In fact, I would propose that evidence about sacrifice is actually *better* when it comes from someone *further* away on the moral continuum. Call this High-to-Low Ratchet. On this view, it is not particularly useful to look at the sacrifices being made by your nearest moral neighbor. So, for example, it would not be useful for the $100/year person to study the sacrifices made by someone who gives $500/year. She should already have a pretty good idea of this herself, after all. She knows what that $500 could buy her and what she would have to give up. She is unlikely to gain any remarkable new justified beliefs about sacrifice by looking at the $500 person, and so it is unlikely that any defeaters of moral obligation will have been removed by exposing her to this new information.

Indeed, perhaps the force of the ratcheting-up effect lies precisely in the *distance* between agents and observers. A lifestyle like Paul Farmer's is utterly foreign to the agent who gives just $100 per year. Until she is exposed to someone like Paul Farmer or Susan Tom, she has every reason to believe that lifestyles like theirs would be unbearable, perhaps even impossible. Learning what their lives are really like changes her stock of evidence in a meaningful way. To be sure, lives like Paul Farmer's and Susan Tom's may still be more self-sacrificial than could be reasonably thought to be *obligatory*. But the ordinary person can extrapolate that a life involving a bit less of a commitment than Farmer's life would involve a bit less sacrifice, until a point is reached where the sacrifice is no longer thought to be unreasonable, and then we reach an obligation. In addition extrapolating downward from the top in this way, the ordinary person can also extrapolate upward from the bottom, by way of a kind of shaming.

Compared to Paul Farmer's lifestyle, giving $100 is a joke, and even $200 would be rather laughable, and $1000 might not even be so bad, and even $5000 might be bearable. Ordinary people can thus extrapolate in two ways—top-down and bottom-up—to triangulate toward a middle ground. As we incorporate new reasonable beliefs into our stock of evidence, we find that morality demands more of us than before.

Of course, the preceding discussion has made it sound like ratcheting is some sort of conscious effort of self-improvement on the part of the observers. But nothing like that is necessary for the Ratcheting-Up Effect to take place. It is enough that people be exposed to the lifestyles of extraordinary moral agents, and through that exposure gain knowledge about burdens and benefits. One need not do any conscious deliberation about where exactly in between one's current level of commitment and the level of a moral saint one should aim to insert oneself. That can come unconsciously or through the gradual shift in social norms governing self-sacrificial behavior. What is most interesting about ratcheting is not that it makes an empirical prediction that moral saints will cause others to *do* more, but that it has a normative entailment: that exposure to moral saints will have the effect of increasing what people can be *required* to do.

The Ratcheting-Up Effect illustrates one role moral saints play in the moral ecosystem. To be sure, moral saints are exceptional people living highly unusual lives. But one way of seeing them is simply as otherwise ordinary people acting as test cases for the rest of us, like marathon runners breaking from the pack. They push the boundaries of moral behavior, discovering what sorts of sacrifices one can realistically bear. When sacrifices are easier to bear than expected, the moral saint takes on even more demanding actions, disseminating evidence to the rest of us. Through a feedback-loop

involving perceptions and predictions about sacrifice, this information serves to change what we can reasonably demand of each other. Thus we can think of moral demands, like legal systems, as flexible over time and influenced by precedent. Moral saints set precedents and landmarks that affect the rest of us.

This brings us back to the "Puzzling Data". Recall that it seemed puzzling that moral saints so frequently declare their extraordinary actions to be unremarkable, even obligatory. I argued that part of what is going on in these cases is that the moral saints and the observers have different perspectives on how much sacrifice the actions involve. Sacrifice, I argued, is properly measured by gross losses of well-being, where well-being is understood objectively and agent-neutrally. The moral saints may be giving up more than they realize, because we can be mistaken about how bad things are for us. And yet we do not immediately dismiss the moral saints' claims that they are "only doing their duty." We take these claims seriously. We find it disconcerting that the moral saint is living up to such a high moral standard; we are shaken from complacency. In learning more about the lives of moral saints, we might begin to see them as hypothetical interlocutors in a moral conversation, as a potential audience for our excuses and complaints. As G.A. Cohen has argued, normative arguments can have more or less force depending on the speaker and the audience (1995, 341). For instance, my excuses for not giving more to charity sound "feeble" when my interlocutor already gives far more than I do (342). With moral saints in the mix, we may find that the standards of what we can reasonably demand from each other are much more flexible than we thought.

BIBLIOGRAPHY

Adams, Robert M. (1984) "Saints." *Journal of Philosophy*, 81, 392-400.

Arpaly, Nomy. (2003) *Unprincipled Virtue: An Inquiry Into Moral Agency*. (Oxford: Oxford UP.

Attfield, R. (1979) "Supererogation and Double Standards." *Mind*, 88, 481-499.

Baron, Marcia. (1987) "Kantian Ethics and Supererogation." *The Journal of Philosophy*, 84:5, 237-262.

Baron, Marcia. (1995) *Kantian Ethics Almost Without Apology*. Ithaca: Cornell University Press.

Baier, Annette. (1995) *Moral Prejudices*. Cambridge: Harvard University Press.

Baier, Annette. (1995) "What Do Women Want in a Moral Theory?" *Moral Prejudices*. Cambridge: Harvard University Press, 1-17.

Bayles, Eugene. (1971) "Act-Utilitarianism: Account of Right-Making Characteristics or Decision-Making Procedure?" *American Philosophical Quarterly*, 8:3, 257-265.

Benbaji, H. and David Heyd. (2001) "The Charitable Perspective: Forgiveness and Toleration as Supererogatory." *Canadian Journal of Philosophy*, 31, 567-585.

Bernstein, Mark. (1986) "Moral and Epistemic Saints." *Metaphilosophy*, 17, 102-108.

Besser-Jones, Lorraine. (2008) "Social Psychology, Moral Character, and Moral Fallibility." *Philosophy and Phenomenological Research*, 76:2, 310-332.

Brink, David O. (1997) "Moral Motivation." *Ethics*, 108:1, 4-32.

Burchill, Lorenne M. (1965) "In Defence of Saints and Heroes." *Philosophy*, 40, 152-157.

Buss, Sarah. (1997) "Justified Wrongdoing." *Noûs*, 31:3, 337-369.

Buss, Sarah. (1999) "Appearing Respectful: The Moral Significance of Manners." *Ethics*, 109:4, 795-826.

Carbonell, Vanessa. (2009) "What Moral Saints Look Like." *Canadian Journal of Philosophy*, forthcoming.

Chisholm, Roderick. (1964) "The Ethics of Requirement." *American Philosophical Quarterly*, 1, 147-153.

Chisholm, Roderick M. (1968) "Supererogation and Offence: A Conceptual Scheme for Ethics." Judith Thomson and Gerald Dwokin, Eds. *Ethics*. New York: Harper and Row, 412-429.

Chopra, Yogendra. (1963) "Professor Urmson on Saints and Heroes." *Philosophy*, 38, 160-166.

Clark, Michael. (1978-79) "The Meritorious and the Mandatory." *Proceedings of the Aristotelian Society*, 79, 22-33.

Clinton, Bill. (2007) *Giving: How Each of us Can Change the World*. (New York: Knopf.)

Cohen, G.A. (1995) "Incentives, Inequality, and Community." Darwall, Stephen, Ed. *Equal Freedom: Selected Tanner Lectures on Human Values*. Ann Arbor, MI: University of Michigan Press.

Cohon, Rachel. (1997) "Hume's Difficulty with the Virtue of Honesty." *Hume Studies*, 27:1, 91-112.

Copp, David. (1997) "Belief, Reason, and Motivation: Michael Smith's 'The Moral Problem.'" *Ethics*, 108:1, 33-54.

Dancy, Jonathan. (1998) "Supererogation and Moral Realism." Jonathan Dancy, J.M.E. Moravcskik, and C.C.W. Taylor, Eds, *Human Agency: Language, Duty, and Value*. Stanford: Stanford University Press.

Dancy, Jonathan. (2000) "Should We Pass the Buck?" in Anthony O'Hear, Ed., *Philosophy: The Good, the True, and the Beautiful*. *Philosophy* Vol. 47 Supplement, 159-173.

Darwall, Stephen. (1996) "Smith's Moral Problem." *The Philosophical Quarterly*, 46:185, 508-515.

Darwall, Stephen. (2002) *Welfare and Rational Care*. Princeton: Princeton University Press.

Darwall, Stephen. (2006) *The Second-Person Standpoint*. Cambridge: Harvard University Press.

Doris, John. (1998) "Persons, Situations, and Virtue Ethics." *Nous*, 32:4, 504-530.

Doris, John. (2002) *Lack of Character: Personality and Moral Behavior*. New York: Cambridge University Press.

Dreier, James. (2000) "Dispositions and Fetishes: Externalist Models of Moral Motivation." *Philosophy and Phenomenological Research*, 61:3, 619-639.

Driver, Julia. (1989) "The Virtues of Ignorance." *The Journal of Philosophy*, 86:7, 373-384.

Driver, Julia. (1992) "The Suberogatory." *Australasian Journal of Philosophy*, 70:3, 286-295.

Enoch, David. (2005) "Why Idealize?" *Ethics*, 115, 759-787.

Estlund, David. (2008) *Democratic Authority*. (Princeton: Princeton University Press.)

Fishkin, James S. (1982) *The Limits of Obligation*. New Haven: Yale University Press.

Feinberg, Joel. (1961) "Supererogation and Rules." *Ethics*, 71:4, 276-288.

Feldman, Fred. (1986) *Doing the Best We Can: An Essay in Informal Deontic Logic*. Boston: D. Reidel.

Flanagan, Owen. (1991) *Varieties of Moral Personality*. Cambridge, MA: Harvard University Press.

Flescher, Andrew Michael. (2003) Heroes, Saints, and Ordinary Morality. Washington, D.C.: Georgetown University Press.

Forrester, Mary. (1975) "Some Remarks on Obligation, Permission, and Supererogation." *Ethics*, 85:3, 219-226.

Fogelman, Eva. (1994) *Conscience and Courage: Rescuers of Jews During the Holocaust*. New York: Doubleday.

Gauthier, David. (1967) "Morality and Advantage." *Philosophical Review,* 76, 460-475.

Gilbert, Martin. (2003) *The Righteous: The Unsung Heroes of the Holocaust*. (New York: Henry Holt.)

Glannon, W. and L.F. Ross. (2002) "Are Doctors Altruistic?" *Journal of Medical Ethics*, 28, 68-69.

Griffin, James. (1986) *Well-being: Its Meaning, Measurement and Moral Importance*. (Oxford: Clarendon Press.)

Hale, Susan. (1991) "Against Supererogation." *American Philosophical Quarterly*, 273-285.

Harman, Gilbert. (1999) "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error." *Proceedings of the Aristotelian Society*, New Series, 99, 315-331.

Harman, Gilber. (2000) "The Nonexistence of Character Traits." *Proceedings of the Aristotelian Society*, New Series, 100, 223-226.

Haybron, Daniel. (2002) "Moral Monsters and Saints." *The Monist*, 85:2, 260-284.

Herman, Barbara. (2000) "Morality and Everyday Life." *Proceedings and Addresses of the American Philosophical Association*, 74:2, 29-45.

Herman, Barbara. (2001) "The Scope of the Moral Requirement." *Philosophy and Public Affairs*, 30:3, 227-256.

Herman, Barbara. (1984) "Rules, Motives, and Helping Actions." *Philosophical Studies*, 45:3, 369-377.

Herman, Barbara. (1981) "On the Value of Acting from the Motive of Duty." *The Philosophical Review*, 90:3, 359-382.

Heyd, David. (1994) "Supererogation and Ethical Methodology: A Reply to Mellema." *Philosophia*, 24, 183-189.

Hill, Thomas E. (1971) "Kant on Imperfect Duty and Supererogation." *Kant-Studien*, 55-76.

Johnson, Conrad D. (1991) *Moral Legislation: A Legal-Political Model for Indirect Consequentialist Reasoning*. (Cambridge: Cambridge University Press.)

Kagan, Shelly. (1989) *The Limits of Morality*. New York: Oxford University Press.

Kahn, Peter H. Jr. (1992) "Children's Obligatory and Discretionary Moral Judgments." *Child Development*, 63:2, 416-430.

Kamm, Fancis. (1985) "Supererogation and Obligation." *Journal of Philosophy*, 82, 118-138.

Kamtekar, Rachana. (2004) “Situationism and Virtue Ethics on the Content of Our Character.” *Ethics*, 114, 458-491.

Kant, Immanuel. (1997) *Critique of Practical Reason*. Trans. Mary Gregor. Cambridge: Cambridge University Press.

Karsh, Jonathan. Director. (2003) *My Flesh and Blood*. (DVD. HBO Documentary Films.)

Kidder, Tracy. (2003) *Mountains Beyond Mountains*. (New York: Random House)

Ladd, John. (1957) *The Structure of a Moral Code*. Eugene, OR: Wipf and Stock.

Lawry, Edward. (2002) “In Praise of Moral Saints.” *Southwest Philosophy Review*, *18*, 1-11.

Levinson, Barry. Director. (1988) *Rain Man*. (DVD. MGM Studios.)

Lillehammer, Hallvard. (1997) “Smith on Moral Fetishism.” *Analysis*, 57:3, 187-195.

MacNamara, P. (1996) “Making Room for Going Beyond the Call.” *Mind*, 105, 415-450.

McDowell, John. (1979) “Virtue and Reason.” *Monist*, 62: 231-250.

McGoldrick, Patricia. (1984) “Saints and Heroes: A Plea for the Supererogatory.” *Philosophy*, 59, 523-528.

McKay, A.C. (2002) “Supererogation and the Profession of Medicine.” *Journal of Medical Ethics*, 28, 70-73.

McNaughton, David and Piers Rawling. (2003) “Can Scanlon Avoid Redundancy by Passing the Buck?” *Analysis*, 63:4, 328-331.

Mellema, Gregory. (1991) *Beyond the Call of Duty: Supererogation, Obligation, and Offence*. Albany: SUNY Press.

Mellema, Gregory. (1991) “Supererogation and the Fulfillment of Duty.” *Journal of Value Inquir*y, 167-175.

Mellema, Gregory. (1994) “Supererogation, Blame, and the Limits of Obligation.” *Philosophia*, 24, 171-182.

Mellema, Gregory. (1996) “Is it Bad to Omit an Act of Supererogation?” *Journal of Philosophical Research*, 21, 405-416.

Milgram, Stanley. (1974) *Obedience to Authority*. New York: Harper Perrenial.

Miller, Alexander. (1996) "An Objection to Smith's Argument for Internalism." *Analysis*, 56:3, 169-174.

Monroe, Kristin R. (2004) *The Hand of Compassion: Portraits of Moral Choice During the Holocaust*. Princeton: Princeton University Press.

Montague, P. (1989) "Acts, Agents, and Supererogation." *American Philosophical Quarterly*, 26, 100-111.

Moore, G.E. (1903) *Principia Ethica*.

Murphy, Liam B. (2000) *Moral Demands in Nonideal Theory*. Oxford: Oxford UP.

New, Christopher. (1974) "Saints, Heroes and Utilitarians." *Philosophy*, 49, 179-189.

Nussbaum, Martha. (1999) *Sex and Social Justice*. (Oxford: Oxford University Press.)

Oliner, Samuel P. and Oliner, Pearl M. (1988) *The Altruistic Personality: Rescuers of Jews in Nazi Europe*. New York: Collier MacMillan.

Olson, Jonas. (2002) "Are Desires *De Dicto* Fetishistic?" *Inquiry*, 45, 89-96.

Orwell, George. (19??) "Reflections on Gandhi" in *A Collection of Essays*. NY: Harcourt Brace.

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

Pojman, Louis J. "In Defense of Moral Saints." in Pojman, Ed. (2002) *Ethical Theory: Classical and Contemporary Readings*, Fourth Ed. (Belmont, CA: Wadsworth, 2002), 388-396.

Portmore, Douglas W. (2003) "Position-Relative Consequentialism, Agent-Centered Options, and Supererogation." *Ethics*, 113, 303-332.

Pybus, Elizabeth M. (1982) "Saints and Heroes." *Philosophy*, 57, 193-200.

Pybus, Elizabeth M. (1986) "A Plea for the Supererogatory: A Reply." *Philosophy*, 61, 526-531.

Raz, Joseph. (1975) "Permissions and Supererogation." *American Philosophical Quarterly*, 12, 161-168.

Ridge, Michael. (2000) "Modesty as a Virtue." *American Philosophical Quarterly*. 37:3, 269-283.

Ridge, Michael. (2003) "Contractualism and the New and Improved Redundancy Objection." *Analysis*, 63:4, 337-342.

Ridge, Michael. (2009) "Fairness and Non-Compliance." In *Impartiality and Partiality in Ethics*, ed. Brian Feltham, John Cottingham and Philip Stratton-Lake. (Oxford: Oxford University Press.).

Scanlon, T.M. (1975) "Preference and Urgency." *The Journal of Philosophy*, 72:19, 655-669.

Scanlon, T.M. (1998) *What We Owe to Each Other*. (Cambridge: Harvard UP)

Scanlon, T.M. (2003a) "Precis of *What We Owe to Each Other.*" *Philosophy and Phenomenological Research* 66:1, 159-161.

Scanlon, T.M. (2003b) "Reply to Gauthier and Gibbard." *Philosophy and Phenomenological Research* 66:1, 176-189.

Scanlon, T.M. (2007) "Wrongness and Reasons: A Reexamination," In Shafer-Landau, Ed., *Oxford Studies in Metaethics*, Vol. 2 (Oxford: Oxford University Press)

Schapiro, Tamar. (2003) "Compliance, Complicity, and the Nature of Nonideal Conditions." *Journal of Philosophy*, 100:7, 329-355.

Scheffler, Samuel, Ed. (1988) *Consequentialism and its Critics*. Oxford: Oxford University Press.

Scheffler, Samuel. (1992) *Human Morality*. New York: Oxford University Press.

Sen, Amartya. (1993) "Capability and Well-Being." Martha Nussbaum and Amartya Sen, Eds., *The Quality of Life*. New York: Oxford University Press, 30-53.

Sherman, Nancy. (1988) "Common Sense and Uncommon Virtue." *Midwest Studies in Philosophy*, 13, 97-114.

Shiffrin, Seana. (2000) "Paternalism, Unconscionability Doctrine, and Accommodation." *Philosophy and Public Affairs*, 29:3, 205-250.

Singer, Peter. (1972) "Famine, Affluence, and Morality." *Philosophy and Public Affairs*, 1:1, 229-243.

Slote, Michael. (1989) *Beyond Optimizing: A Study of Rational Choice*. Cambridge, MA: Harvard University Press.

Smith, Michael. (1994) *The Moral Problem*. (Oxford: Blackwell)

Smith, Michael. (1996) "The Argument for Internalism: Reply to Miller." *Analysis*, 56:3, 175-184.

Smith, Michael. (1996) "Normative Reasons and Full Rationality: Reply to Christine Swanton." *Analysis*, 56:3, 160-168.

Smith, T.V. (1948) "Saints: Secular and Sacerdotal—James Madison and Mahatma Gandhi." *Ethics*, 59:1, 49-60.

Sobel, John Howard. (2006) *Good and Gold: A Judgmental History of Metaethics from G.E. Moore to J.L. Mackie*. Chapter X: Moral Psychology. Manuscript. Accessed from www.utsc.utoronto.ca/~sobel/Gd_Gld/ on Aug. 28, 2006.

Sreenivasan, Gopal. (2002) "Errors about Errors: Virtue Theory and Trait Attribution." *Mind*, 111: 441, 47-68.

Stanlick, Nancy. (1999) "The Nature and Value of Supererogatory Actions." *Journal of Social Philosophy*, 30:1, 209-222.

Stocker, Michael. (1968) "Supererogation and Duties." In Nicholas Rescher, Ed., *Studies in Moral Philosophy*. Oxford University Press, 53-63.

Stocker, Michael. (1976) "The Schizophrenia of Modern Ethical Theories." *The Journal of Philosophy*, 73:14, 453-466.

Stocker, Michael. (1990) *Plural and Conflicting Values*. (Oxford: Oxford University Press).

Stratton-Lake, Philip. (2003) "Scanlon's Contractualism and the Redundancy Objection." *Analysis*, 63:1, 70-76.

Stratton-Lake, Philip. (2003) "Scanlon, Permissions, and Redundancy: Response to McNaughton and Rawling." *Analysis*, 63:4, 332-337.

Svarvarsdottir, Sigrun. (1999) "Moral Cognitivism and Motivation." *The Philosophical Review*, 108:2, 161-219.

Swanton, Chrisine. (2003) *Virtue Ethics: A Pluralistic Approach*. Oxford: Oxford University Press.

Swanton, Christine. (1996) "Is the Moral Problem Solved?" *Analysis*, 56:3, 155-160.

Trianosky, Gregory W. (1986) "Supererogation, Wrongdoing, and Vice: On the Autonomy of the Ethics of Virtue." *Journal of Philosophy*, 83:1, 26-40.

Unger, Peter. (1996) *Living High and Letting Die*. (New York: Oxford University Press.

Urmson, J.O. (1958) "Saints and Heroes." in A. I. Melden, Ed. *Essays in Moral Philosophy*. (Seattle: University of Washington Press), 198-216.

Vranas, Peter B.M. (2005) The Indeterminacy Paradox: Character Evaluations and Human Psychology." *Nous*, 39, 1-42.

Vranas, Peter B.M. (2007) "Against Moral Character Evaluations: The Undectability of Virtue and Vice." Unpublished/Forthcoming.

Wallace, R. Jay. (2006) "Moral Reasons and Moral Fetishes: Rationalists and Anti-Rationalists on Moral Motivation" in *Normativity and the Will*. Oxford: Oxford UP.

Wolf, Susan. (1982) "Moral Saints." *Journal of Philosophy*, 79:8, 419-439.

Wolf, Susan. (1986) "Above and Below the Line of Duty." *Philosophical Topics*, 14:2, 131-148.

Zimmerman, Michael. (1993) "Supererogation and Doing the Best One Can." *American Philosophical Quarterly*, 30:4, 373-380.

Zimmerman, Michael J. (1996) *The Concept of Moral Obligation*. Cambridge: Cambridge University Press.