

JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY

VOLUME XXI · NUMBER 1

January 2022

ARTICLES

What Is Group Well-Being?

Eric Wiland

The Case for Consent Pluralism

Jessica Keiser

Utilitarianism, Altruism, and Consent

Christopher J. G. Meacham

Forgiveness and the Significance of Wrongs

Stefan Riedener

Positive Rights: Two-Person Cases

David Alm

DISCUSSIONS

Should Libertarians Reject the Free Market?

On Olsaretti's Positive Answer

Peter Bornschein

Divine Commands Are Unnecessary for Moral
Obligation

Erik J. Wielenberg

JOURNAL *of* ETHICS
& SOCIAL PHILOSOPHY

VOLUME XXI · NUMBER 1

January 2022

ARTICLES

- 1 What Is Group Well-Being?
Eric Wiland
- 24 The Case for Consent Pluralism
Jessica Keiser
- 49 Utilitarianism, Altruism, and Consent
Christopher J. G. Meacham
- 82 Forgiveness and the Significance of Wrongs
Stefan Riedener
- 114 Positive Rights: Two-Person Cases
David Alm

DISCUSSIONS

- 132 Should Libertarians Reject the Free Market?
On Olsaretti's Positive Answer
Peter Bornschein
- 142 Divine Commands Are Unnecessary for
Moral Obligation
Erik J. Wielenberg

The *Journal of Ethics and Social Philosophy* (ISSN 1559-3061) is a peer-reviewed online journal in moral, social, political, and legal philosophy. The journal is founded on the principle of publisher-funded open access. There are no publication fees for authors, and public access to articles is free of charge and is available to all readers under the CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-NODERIVATIVES 4.0 license. Funding for the journal has been made possible through the generous commitment of the Gould School of Law and the Dornsife College of Letters, Arts, and Sciences at the University of Southern California.

The *Journal of Ethics and Social Philosophy* aspires to be the leading venue for the best new work in the fields that it covers, and it is governed by a correspondingly high editorial standard. The journal welcomes submissions of articles in any of these and related fields of research. The journal is interested in work in the history of ethics that bears directly on topics of contemporary interest, but does not consider articles of purely historical interest. It is the view of the associate editors that the journal's high standard does not preclude publishing work that is critical in nature, provided that it is constructive, well-argued, current, and of sufficiently general interest.

Executive Editor

Mark Schroeder

Associate Editors

Saba Bazargan-Forward	Hallie Liberto
Dale Dorsey	Errol Lord
James Dreier	Tristram McPherson
Julia Driver	Colleen Murphy
Anca Gheaus	Hille Paakkunainen
David Plunkett	

Discussion Notes Editor

Kimberley Brownlee

Editorial Board

Elizabeth Anderson	Philip Pettit
David Brink	Gerald Postema
John Broome	Joseph Raz
Joshua Cohen	Henry Richardson
Jonathan Dancy	Thomas M. Scanlon
John Finnis	Tamar Schapiro
John Gardner	David Schmidtz
Leslie Green	Russ Shafer-Landau
Karen Jones	Tommie Shelby
Frances Kamm	Sarah Stroud
Will Kymlicka	Valerie Tiberius
Matthew Liao	Peter Vallentyne
Kasper Lippert-Rasmussen	Gary Watson
Elinor Mason	Kit Wellman
Stephen Perry	Susan Wolf

Managing Editor

Rachel Keith

Copyeditor

Susan Wampler

Typesetting

Matthew Silverstein

WHAT IS GROUP WELL-BEING?

Eric Wiland

IN ORDINARY LIFE, we strive to benefit various groups of people. We worry about what is best for our families. Coaches aim to help their teams. Chairs hope to strengthen their departments. Those seeking reparations presuppose some notion of what is good and bad for oppressed groups of people. In *The Republic*, Socrates aims to establish the well-being of the entire *polis*, not the well-being of any individual. In her recent American Philosophical Association presidential address, Valerie Tiberius considers what is good not for individual philosophers but for the discipline of philosophy.¹

We can wisely determine how to benefit our families, teams, departments, cities, and disciplines, however, only if we understand what it is for a group to do well. And it is quite unclear how we should understand the well-being of a group. Economists, of course, have had various ideas about the notion of a social welfare function.² However, they tend to simply presume that the welfare of a society is some mathematical function of the welfare of the individuals in that society. Political scientists tend to be more imaginative—they understand that the common good may be distinct both from aggregative utility and from what economists call “public goods”—but they fail to spell out how the good of a group *does* relate to the welfare of the individuals who constitute the group in question.

We are not only benefactors, but also beneficiaries. If, as I will argue, your individual well-being depends upon the well-being of the groups to which you belong, this would be good for you to know, both for the purpose of deciding which groups to join and for the purpose of prioritizing their flourishing. In a world in which group identities are not fixed (Should you quit your book club? Should you leave your department? Should you move to/from Canada?), you would do well to know whether and how the quality of your own life depends upon the well-being of the groups you partially constitute.

A philosophical account of group well-being is thus needed, and forging one

1 Tiberius, “The Well-Being of Philosophy.”

2 For example, Arrow, *Social Choice and Individual Values*.

is what I begin to do here. One might expect that philosophers who explicitly work on the topic of well-being would already have developed views about the well-being of groups. There is, after all, a vast literature devoted to understanding the well-being of *individuals*, focusing upon whether the well-being of an individual is grounded in experiences, in the satisfaction of desires, in the fulfillment of values, in achievements, or in whether one's life bears objectively valuable features. But individuals are (or might be) only one thing that can be well or ill, and, unfortunately, philosophers who work extensively on well-being have been largely silent about the well-being of groups.

Given recent trends in philosophy more generally, this is very surprising. During the past decade philosophers have produced a tremendous amount of work about groups: there is now a cascade of exciting work on group agency, group ontology, group duties, group assertion, group attitudes, group responsibility, and group reasoning.³ These various debates zero in on what it takes for a group to harm (or to benefit) another, to have a duty not to harm (or to benefit) another, and to be responsible for so acting. But what is it for a group *itself* to be harmed or benefited? That is, what is group well-being?

Unfortunately, I will not have any grand conclusions about what constitutes group well-being; here I attempt to sketch some of the logical space of possible answers (or forms of possible answers), and nudge us to seriously consider certain widely overlooked options. More specifically, I will describe several importantly different ways the well-being of a group may be related to the well-being of the individuals who constitute the group. This, I hope, will be eye opening.

In section I, I describe the kinds of groups whose well-being I am investigating. In section II, I briefly introduce and taxonomize four competing conceptions of group well-being. In sections III and IV, I distinguish two different ways group well-being and individual well-being might be related to each other: bottom-up dependence and top-down dependence. In sections V and VI, I argue that top-down dependence might obtain: your individual well-being might be partially constituted by the well-being of the groups to which you belong. In section VII, I consider whether *both* forms of dependence could obtain at once.

3 On group agency see List and Pettit, *Group Agency*; Tollefsen, *Groups as Agents*. On group ontology see Epstein, *The Ant Trap*. On group duties see Collins, *Group Duties*. On group assertion see Lackey, "Group Assertion." On group attitudes see List, "Three Kinds of Collective Attitudes." On group responsibility see Bazargan-Forward and Tollefsen, *The Routledge Handbook of Collective Responsibility*. And on group reasoning see Karpus and Gold, "Team Reasoning."

I

Before discussing well-being, I will make three quick points about groups themselves. First, I assume groups exist. If groups *do not* exist, then nothing makes groups better or worse off—exploring the well-being of groups would then be no more important than exploring the well-being of ghosts. But even Margaret Thatcher, who famously said “There’s no such thing as society,” went on to concede that “There are families.” So here I will just assume that there are indeed *some* kinds of groups.

Second, I am thinking only about groups that are entirely constituted by individual people. So, if your household has a dog as one of its members, I am not considering the well-being of your household so understood for present purposes. Nor am I considering here groups of groups: the Association of American Universities or the European Union might be viewed as a meta group. The well-being of these groups is especially complex. So, for now, I am interested only in groups of individual human beings.

Third, and most importantly, there are likely different types of groups of human beings, and what constitutes the well-being of these groups might vary by type. Some groups are highly organized around a particular mission, a mission the fulfillment of which may have something to do with the group’s well-being. Other groups are constituted by individuals who share some important features, e.g., the physically disabled. Still others seem to be little more than mathematical sets of individuals. So, even if we figure out what constitutes the well-being of, say, a nation, we might not yet understand what constitutes the well-being of a particular friendship, or of the group of residents that lives in a particular apartment building. Thus, we should beware of overgeneralizing our results: even if we make progress in understanding the well-being of some types of groups, we might not thereby understand the nature of the well-being of other types of groups any better. This will prove relevant shortly.

II

What, then, *is* group well-being? One contemporary philosopher of well-being who *has* explicitly considered this question, if only briefly, is Ben Bradley. Bradley describes several different ways one might think of the “well-being of a group of people.”⁴ He writes, “If we can compare well-being levels between people, we

4 Bradley, *Well-Being*, 70.

can attempt to figure out how well off a population is on the basis of how well its members are.”⁵

One strategy is to *add up* the well-being of each member of the group in order to calculate the well-being of the group itself. The well-being of a group, then, would just be equal to the sum of the well-being of each of its members. But Bradley correctly notes that this proposal implausibly implies that a group of a billion extremely well-off individuals is not as well off as a much larger group of individuals, each of whom is barely well off. Intuitively, the well-being of a group does not (always?) depend solely upon how much well-being its members experience *in toto*. This first strategy inappropriately favors large groups. Summing the well-being of individual members to determine group well-being seems to be too crude.⁶

A better approach, one in line with what social psychologists do, is to average the well-being of the individuals in the group.⁷ This would avoid the unintuitive result that a group comprising an infinite number of barely well-off people is itself infinitely well off. But Bradley rightly worries about this view too:

Suppose that population P₃ has 100 moderately well-off people. P₄ has 99 unhappy people plus George, who is extraordinarily well-off. George is so well-off that when you calculate the average well-being levels of P₃ and P₄, P₄'s average is higher even though everyone in P₃ is better-off than all the non-George members of P₄. You might wonder whether that really means that P₄ is better-off than P₃.⁸

The fact that George is extremely well off surely counts for something, but it does not seem that it swamps all other facts about individual well-being in determining what the well-being of P₄ is. Although averaging individual well-being seems better than adding individual well-being, neither yields a satisfying answer to our investigation. The well-being of a group surely depends upon other factors too.

Nevertheless, Bradley eventually concludes that since the average well-being approach is superior to the other approaches he considers, it is the “most plausi-

5 Bradley, *Well-Being*, 78. He switches from talking about a “group” to a “population.” I will postpone worries about this.

6 Parfit, *Reasons and Persons*.

7 Di Tella and MacCulloch write that “a large fraction of the happiness literature in economics is based on comparing average happiness scores for large numbers of people” (“Some Uses of Happiness Data in Economics,” 29). This passage is approvingly cited by Angner, “Well-Being and Economics.”

8 Bradley, *Well-Being*, 78.

ble view about the well-being of a population.”⁹ I instead suggest that we need to survey more options before embracing that—or any—conclusion. The options Bradley *does* consider are all alike insofar as they treat the well-being of a group as depending solely upon the well-being of the individuals in the group. In each case, one is to determine how well off a group of people is by first determining how well off each individual in the group is, and then somehow calculate the well-being of the group from these determinations. The well-being of a group is thus assumed to be some mathematical function of the well-being of the individuals constituting the group. Suppose group G_1 has three individuals: I_1 , I_2 , I_3 . If we know the relevant function, then, once we know the well-being of I_1 , I_2 , and I_3 , we can calculate the well-being of G_1 . Nothing else matters. Call this a *reductive* conception of group well-being.¹⁰

We should not be quick to assume, however, that *any* reductive account of group well-being is correct. Even if every individual in some group is doing well, it seems possible that the group itself is not doing well. After all, some wholes do not have properties that all their proper parts have: to assume otherwise is to commit the fallacy of composition. So, if we do accept a reductive view, it would be nice to have an argument for it. Or, at a minimum, we should note and rule out the alternatives.

What, then, *are* the alternatives? Here is one. According to what I will call the *independent view* of group well-being, the well-being of a group of individuals and the well-being of the individuals in the group have nothing directly to do with one another. The fact that all the individuals in some group are doing well entails nothing about the well-being of the group itself, and whatever makes a group well off has nothing essentially to do with whether the individuals in the group are well off. What makes a group well off is something other than the well-being of its members.

We have to be careful about how to articulate this view. Even if the independent view is true, the well-being of a group and the well-being of the individuals in a group may often *happen* to depend upon the same thing, even though they do not *essentially* depend upon the same thing (or each other). To illustrate this possibility, suppose that some objective list theory of individual well-being is true, according to which an individual is well off to the extent they have three

9 Bradley, *Well-Being*, 90. Along the way, he also considers and wisely rejects the view that the well-being of a group is determined by the *median* level of well-being in the group (79).

10 In the philosophy of groups more generally, reductive or summative views are often one popular theoretical option. For discussion, see Tollefsen, *Groups as Agents*. For an instance of a reductive or summative conception of group intention, see Searle, “Collective Intentions and Actions.”

traits: pleasure, knowledge, and health. An individual who learns a lot during their pleasant, healthy life, then, is well off. Suppose also that the well-being of some particular group of individuals (say, a manufacturing firm) essentially depends only upon whether the group produces a lot of widgets: the more widgets it produces, the better off the group is. The well-being of the individuals and the well-being of this group, then, would not *essentially* depend upon each other, nor upon some common factor. But this is compatible with the fact that some of the things that are good or bad for each of the individuals will also *happen* to be things that are good or bad for the group. For example, if a viral plague strikes the land, then this will be bad for many of the individuals (the plague is painful, it damages health, etc.), and this will also be bad for the group (fewer widgets are produced when many individuals are sick). If a new vaccine makes many individuals' lives more pleasant (such that they are better off), this may *cause* them to work more productively, thereby making more widgets (such that the group is better off). So even if the independent view of group well-being is correct, there may be some *merely causal* links between the two. We thus should not confuse *constitutive* dependence with mere causal dependence. The proponent of the independent view denies that the well-being of the group constitutively depends upon the well-being of the individuals who make up the group.

The independent view is also compatible with the possibility that the well-being of a group constitutively depends upon *other* features of the individuals in the group. To see this, imagine a theory of group well-being according to which the well-being of the group simply depends upon the *size* of the group: the bigger the group, the better off the group is. (This theory is obviously highly implausible, but it cleanly illustrates the present point.) Suppose Susan joins the group. Now, the group is better off in virtue of Susan's membership. If she were to exit the group, the group would suffer. The well-being of the group constitutively depends upon her membership in the group. This, however, is perfectly consistent with the independent view. The independent view denies only that the well-being of the group constitutively depends upon the *well-being* of the individuals in the group: it does not deny that it so depends upon any *other* feature of the individuals in the group.¹¹

Is the independent view plausible? For *some* sort of groups, I think it might be. It may be true that the well-being of a for-profit corporation is in no way constituted by the well-being of the individuals who make up the corporation. It may instead depend exclusively upon things like its profitability, market share, or

11 Compare to a view that whether a collective is morally responsible for some fact in no way depends upon whether any of the individuals in the collective are so responsible. See Copp, "The Collective Moral Autonomy Thesis."

success in achieving its mission. The independent view of well-being will seem especially plausible, then, if we think that the well-being of something substantively depends upon the *kind* of thing it is, and if individuals and (certain kinds of) groups of individuals are very different kinds of things.

So far, I have sketched two views of group well-being: the reductive view and the independent view. These two views starkly oppose each other. The reductive view says that group well-being is a *mathematical function* of the well-being of the individuals in the group. Group well-being depends upon *nothing* else. The independent view, by contrast, says that group well-being in no way constitutively depends upon the well-being of the individuals in the group; it depends *entirely* on something else. But these are not the only options in logical space. Perhaps the well-being of a group *partially* constitutively depends upon the well-being of the individuals in the group. This would be so if the well-being of a group constitutively depends upon both (a) the well-being of the individuals in the group, and, say, (b) its success in achieving its stated mission (which need not concern the well-being of its individuals). Two groups, then, might be equally well off, even though one group's members are individually better off, while the second group better achieves its stated mission.

Intermediate positions such as this, then, *partially* relate the well-being of individuals and that of groups. On these partial conceptions of group well-being, there is *some* constitutive dependence relation between the well-being of the individuals in the group and the well-being of the group, but knowing only the well-being of the individuals does not let you calculate the well-being of the group, because the well-being of groups also partially depends upon factors that do not play the same role in determining the well-being of individuals. Call this view *group well-being partialism*—or, for short, *partialism*. And we can now call both partialism and the independent view *non-reductive* views of group well-being.

Are these all the options? I can think of at least one more: although groups themselves are indeed real, there is no such thing as the well-being of a group, and so we do not need to worry about *how* the well-being of a group relates to that of the individuals in the group. Maybe *only* individuals can be well or ill. Call this view *group well-being eliminativism* (or, for short, *eliminativism*).

While eliminativist views in philosophy sound radical, this one need not be so edgy. Although economists and other social scientists often talk about the well-being of groups or populations, perhaps this is only a *façon de parler*. When we want some way to speak about the well-being of many individuals at once, we will speak as if the group of individuals can be well or ill, but perhaps this is

at least sometimes merely a handy locution.¹² Perhaps groups just *cannot* be the sort of thing that can be well or ill.¹³

While I will not evaluate group well-being eliminativism here, I do find it an initially plausible account of *some* groups. Imagine a group of people waiting in front of Walmart for it to open. I do not think that this *group* could be better or worse off than it is. Each of the individuals in the group could have been better off or worse off, to be sure. And we may want some simple way to talk about this. To talk about this concisely, we might pretend that the group could have been better off or worse off. But perhaps this is not literally true; much seems to depend upon how loose a set of individuals can be related to one another and still constitute a group of some sort. With groups that seem to be *merely* populations, eliminativism can seem plausible.

There are, then, at least four ways to understand the relation between the well-being of a group and the well-being of the individuals in the group. Either (1) eliminativism is true, or the relation between the well-being of a group and the well-being of the individuals in the group is (2) reductive, (3) partial, or (4) independent. Further, it is not immediately obvious which of the four options is superior, and it is plausible that the correct account for some kinds of groups differs from the correct account for groups of other kinds.

III

Suppose that a non-reductive view of group well-being is correct for at least some groups. Here is a further question: What is the *direction* of constitutive dependence? In what way do individual well-being and group well-being depend upon each another?

So far, we have been explicitly wondering only whether the well-being of the group constitutively depends upon the well-being of the individuals in the group. But perhaps the dependence can run the other way, such that the well-being of individuals in some group constitutively depends upon the well-being of the groups they constitute. When a group to which you belong does well, do you yourself ever *thereby* benefit?

- 12 Sumner brusquely writes that “talk ... of the welfare of groups, if it is not merely metaphorical, must be interpreted as referring to the aggregate or collective well-being of their members. Collectivities have no interests to be promoted beyond those of individuals,” thereby straddling a reductive view and an eliminativist view of group well-being (Sumner, *Welfare, Happiness, and Ethics*, 180).
- 13 Compare to atomism about collective responsibility the view that only individuals can be morally responsible. See Lewis, “Collective Responsibility”; Sverdlik, “Collective Responsibility”; Narveson, “Collective Responsibility.”

Although this way of conceiving the relation between individual well-being and group well-being is completely at odds with the way it is commonly theoretically conceived, I believe it has some merit.¹⁴ Perhaps you are better off simply in virtue of the fact that your department is thriving; perhaps you are worse off simply in virtue of the fact that your nation is deteriorating. We can label these as two ways that individual well-being and group well-being might constitutively depend upon each other: if the well-being of a group depends upon the well-being of the individuals who constitute the group, then we have a case of *bottom-up* dependence; if, however, the well-being of some individual constitutively depends upon the well-being of the group or groups to which she belongs, we have a case of *top-down* dependence. If top-down dependence about well-being obtains, then when a group is doing well, its individual members thereby do better than they otherwise would be doing—a phenomenon we would miss if instead we calculated the well-being of each individual in isolation from the ways group well-being might partially determine it.

I suspect, however, that most philosophers and social scientists assume, if only implicitly, that *only* bottom-up dependence obtains. Philosophers who embrace standard hedonistic or desire-satisfaction views of individual well-being will probably be inclined to reject top-down dependence. They are already hostile to the thought that an individual's well-being could directly depend upon something so apparently external. Others, however, might be interested in thinking more about top-down dependence.

Although I am ultimately friendly to the idea of top-down dependence, and will try to argue for it, I first will discuss some methodological *obstacles* toward establishing it. For if we think about this issue wrongly, it can look too *easy* to establish top-down dependence. When I talk to some nonphilosophers about this topic, they think it is just obvious that an individual's well-being depends upon the well-being of the groups to which she belongs. But we have to be careful here. Top-down dependence is actually more radical than it might initially appear. So, I will first show why it is *not* so easy to establish top-down dependence before ultimately arguing that in some cases it is plausible.

Let us think about the relation between an individual's well-being and the well-being of an academic department that person partially constitutes. One might concede that the well-being of your department affects your well-being,

14 Note, however, that the analogous question about group and individual responsibility has been widely discussed. Philosophers differ over whether the fact that a group is responsible for some wrongdoing entails or implies that its members are individually responsible for that wrongdoing. See, for instance, French, *Individual and Collective Responsibility*; May, *Sharing Responsibility*; Raikka, "On Disassociating Oneself from Collective Responsibility"; and Bazargan-Forward and Tollefsen, *The Routledge Handbook of Collective Responsibility*.

but insist that it does so *entirely through* some individual feature about yourself. For example, suppose Pat is a member of your department. Suppose some fact improves your department: perhaps a toxic administrator leaves.¹⁵ The fact that the department is now better off could then enhance Pat's well-being, but entirely because of some individual feature of Pat. For example, if Pat's well-being improves simply in virtue of the fact that Pat *cares about* her department, then this would not really be a case where Pat's well-being *constitutively* depends upon the well-being of her department. It would not be a genuine case of top-down dependence. It would instead run entirely through her *caring*, a feature of her as an individual. Pat would indeed be better off, because the thing that she cares about is doing well, and her personal well-being depends (on this internalist view) only on things in this way.

To see this point more clearly, compare the above case to a case where Pat cares about something that involves neither any group to which she belongs nor to her at all: suppose, for example, she cares deeply about whether most scientists deem Pluto to be a planet. Whether most scientists deem Pluto to be a planet, then, could plausibly affect Pat's well-being—if they do, it would satisfy a desire of hers—but that is no way shows that Pat's well-being constitutively depends upon whether Pluto is deemed to be a planet. Rather, it so depends upon whether some particular attitude of hers is satisfied, an attitude that *happens* to be about Pluto and planets. What matters to her well-being is not the way the world is, but whether the world is the way she wants it to be. Likewise, if the *only* way that the well-being of her department affects Pat's well-being is in virtue of the fact that Pat *cares about* her department, then her personal well-being does not *directly* depend upon the well-being of her department. It just so happens that the thing she cares about (*viz.*, her department) *happens to be* a group of which she is a member. But this is not importantly different from the way that some fact completely external to her (e.g., whether scientists deem Pluto to be a planet) can affect her individual well-being. In both cases, her well-being is affected simply because she cares about or wants something, and the world lines up aptly with the content of her desires.

Because human beings are social creatures, it is obvious that individual well-being positively co-varies with the well-being of the groups to which they belong. But this fact alone does not show that there is top-down dependence about well-being. If individuals are better off only in virtue of the fact that they *care about* these groups, or have *better experiences* when their groups thrive, then there is no top-down dependence. Genuine top-down dependence obtains only

15 All names, characters, and incidents portrayed in this example are fictitious. No identification with actual persons (living or deceased) or places is intended or should be inferred.

if an individual like Pat is better off in virtue of some improvement in a group to which she belongs, such as her department, where this is *not* merely mediated by her experiences, desires, values, or any other *individual* feature of hers. Thus, it is far from obvious that top-down dependence about group well-being obtains.

IV

Perhaps we can vindicate top-down dependence by showing that individual well-being is in fact rooted in something *irreducibly* social. Larry May is after something like this idea in his discussion of what he calls group-based harm. He says that harms “are group-based when there is something about the structure, or perceived structure, of a given group that makes all of the members of the group at least indirectly or vicariously harmed whenever one of the members is directly harmed.”¹⁶ For some groups, then, “harm of some members also harms all group members.”¹⁷ Group-based harm, then, is “not reducible to the aggregate harms and claims of the individual members of these groups.”¹⁸

How does this happen? May describes several possibilities. First, all members of a group can be harmed when there is “strong empathy by other members for those who are directly harmed.”¹⁹ If another member of my department is harmed, and I strongly empathize with her, then I too am harmed.²⁰

This sounds correct, but one need not be a member of the relevant group in order to strongly empathize with people who are directly harmed. Someone who is not a member of my department might empathize with the person directly harmed just as strongly as I do, and thus be harmed in the same way I am. You do not need to share membership in a group in order to fully empathize with another. So, empathy alone does not explain what if anything is special about group-based harms.

Second, May notes that every member of a group is harmed when “directly harmful treatment of some members could just as easily have been directed at those who were not directly affected.”²¹ According to this thought, you are harmed when someone else in your group was harmed on the basis of being a member of a group, and that harm could instead have befallen a different member of the group, such as yourself. A full explanation of the harm that befalls

16 May, *The Morality of Groups*, 116.

17 May, *The Morality of Groups*, 115.

18 May, *The Morality of Groups*, 116.

19 May, *The Morality of Groups*, 115.

20 Simon (“Group Harm”) also argues along these lines.

21 May, *The Morality of Groups*, 115.

every member, then, cannot be made without some reference to a social group or to one of its features, such as its structure.

It is unclear to me, however, how my being counterfactually harmed entails that I am actually harmed. Suppose some tourist is murdered simply because she was an American. Since I too am an American who sometimes travels overseas, I could have been murdered instead. But I *have not* been murdered, and, indeed, it sounds strange to say I have been harmed at all. Rather, I merely could have been harmed. Perhaps if murdering Americans is sufficiently common and systematic, I would somehow be harmed by these murders vicariously. But as a general statement, mere counterfactual possibility is not enough to ground actual harm.

There is a deeper problem, however, with employing May's notion of a group-based harm (which might be fine as it stands) as a solution to *our* task of understanding group well-being and the possibility of top-down dependence. Let us grant that an individual's well-being can essentially depend upon *socially constituted* features of herself, such as group membership. For example, as May soundly argues, there is no way to understand the nature of the harms suffered by many people in South Africa in the 1980s without referring to the groups to which they belong. Does that fact establish that individuals are harmed because the groups to which they belong are harmed—that there is top-down dependence? No.

To see why not, let us examine a toy example: suppose that an individual's well-being (partially) depends upon whether she plays baseball. Whether she plays baseball is, in a way, a fact about her as an individual. Nevertheless, whether her various bodily movements add up to instances of *playing baseball* depend upon whether she is a member of a baseball team.²² You cannot understand what it is for her to play baseball (a team sport) without reference to the group to which she belongs (a team). Whether an individual person is stealing a base depends upon facts beyond what is going on with the person psychologically. It depends also upon whether the person is a part of (participating in) the practice

22 See Rawls, "Two Concepts of Rules": "Many of the actions one performs in a game of baseball one can do by oneself or with others whether there is the game or not. For example, one can throw a ball, run, or swing a peculiarly shaped piece of wood. But one cannot steal base, or strike out, or draw a walk, or make an error, or balk; although one can do certain things which appear to resemble these actions such as sliding into a bag, missing a grounder and so on. Striking out, stealing a base, balking, etc., are all actions which can only happen in a game. No matter what a person did, what he did would not be described as stealing a base or striking out or drawing a walk unless he could also be described as playing baseball, and for him to be doing this presupposes the rule-like practice which constitutes the game. The practice is logically prior to particular cases: unless there is the practice the terms referring to actions specified by it lack a sense" (25).

of baseball. So, *we* cannot tell whether her various movements constitute playing baseball without considering the wider context. Even to describe her acts accurately, then, we need to look not only inside her head, or at what her body is doing, but at the bigger picture.

Now, *if* an individual's well-being were to essentially depend upon whether she is playing baseball, then to characterize her well-being we would need to know some social facts. These facts might include group-level facts about baseball teams. To know whether Pat played baseball, struck out, or made an error (as opposed to fiddled with some pieces of wood and spheres), we *do* need to know whether she was indeed on a baseball team, a group. That is a fact *about* a group. But this would not vindicate top-down dependence about well-being, for nothing said so far implies that we must know anything about group *well-being*. To know that she plays baseball, we do not need to know anything about *how well off* her team is. According to the toy view we are now exploring, it matters not to Pat's well-being whether her baseball team thrives or wilts, only *that* she plays baseball.

What does this all mean? It is very plausible that *some* determinants of individual well-being concern group-level facts—probably not facts about whether one plays baseball in particular, but perhaps facts about whether one participates in some civic organization or other, or, now thinking of May's concerns, facts about whether it is a socially dominant or oppressed group. But *even if* we establish that all this is so, we would not have thereby established top-down dependence about group well-being. For top-down dependence of this sort obtains only if facts about group *well-being* determine facts about individual well-being. And in the toy baseball example, although some facts about *groups* determine facts about individual well-being, it is not facts about *group well-being* that do so. We should not confuse this weaker kind of phenomenon with genuine top-down dependence about group well-being.

Distinguishing top-down dependence about group well-being from things it merely resembles, then, reveals the difficulty in arguing that the well-being of individuals indeed directly depends upon the well-being of the groups to which individuals belong. The skeptic about top-down dependence can concede that people are better off when they are in social groups of various sorts: when they have friends and family and colleagues and are members of labor unions, and so on. The need for some sociality is completely compatible with the absence of top-down dependence.

V

To establish top-down dependence about group well-being, we need to show that an individual can be better off *directly* in virtue of the fact that a group to which she belongs is better off. How can we do that?

Here is a case to isolate our intuitions. Suppose that some college cross-country running teams are better off than others are. Plausibly, *how* well off they are depends at least in part upon whether they win their competitions. In the United States, teams are typically constituted by seven runners, and a team's score is a function of the times of its fastest five runners in that competition. The times of the sixth and seventh fastest runners do not affect the team's success; in the absence of a tie, these times might as well be infinite.

Imagine Arkansas is competing against BYU in the national championship, and that Arkansas's fastest five runners run more quickly than BYU's fastest five runners. Arkansas thus wins the competition, thereby plausibly making the Arkansas team better off than it would otherwise be, and the BYU team worse off than *it* would otherwise be.

Now imagine further that each team's sixth-fastest runner runs exactly the same time. Individually, then, their achievements are identical. If Arkansas-6 (as I will call her) is indeed better off than BYU-6 (as I will call *her*) is, it is *not* because of the former's individual achievement. After all, their individual achievements are identical. Of course, it is highly plausible that Arkansas-6 is better off than BYU-6 is, because she *wants* her team to win, and she *enjoys* winning, and so on. We do not have to talk about team well-being in order to make some sense of the thought that Arkansas-6 is doing better than BYU-6 is.

Next, let us compare Arkansas-6 to someone I will call Arkansas-Fan. Let us stipulate that Arkansas-6 and Arkansas-Fan equally enjoy the fact that Arkansas won, equally want Arkansas to win, and are alike in all their relevant psychological attitudes toward the event of Arkansas's victory. Even so, it seems plausible that Arkansas-6 benefits in a way that Arkansas-Fan does not. After all, Arkansas-6 achieved something that Arkansas-Fan did not, even though her individual time did not affect the outcome of the race. Even though Arkansas-6's time might have been terrible, it was arguably still *some* sort of achievement. We do not have to talk about team well-being in order to make some sense of the thought that, when the Arkansas team wins, Arkansas-6 is doing better than Arkansas-Fan is.

But let us combine the two scenarios. Suppose BYU-6 just *is* Arkansas-Fan. Perhaps before running for BYU's team, BYU-6 grew up in Arkansas, has followed Arkansas's sports programs for years, is happy when Arkansas wins, etc. More specifically, let us assume that her psychological attitudes toward the Arkansas

team's success are essentially the same as Arkansas-6's, and that her individual achievement in the race is the same as Arkansas-6's. Thus, the relevant individual features of the two people are identical. They are equally pleased, their desires are equally satisfied, their individual achievements are identical, and so on. (If we need to suppose further details to make this plausible—Arkansas-6 grew up in Utah, etc.—let us do so.) And yet Arkansas-6 *is* a member of the team that is doing well, while BYU-6/Arkansas-Fan is a member of the team that is not doing so well.

I find it at least plausible that Arkansas-6 benefits from the fact that the Arkansas team benefits *in a way* that BYU-6/Arkansas-Fan does not. When the Arkansas team thrives, its players thereby do better, and this seems to me not to be wholly explainable by focusing on the psychological states and the achievements of the individuals. The fact that the Arkansas team just *is* Arkansas-6 (and some other individuals) seems to be some grounds for thinking that an individual's well-being can be partially constituted by the well-being of a group that *she herself* partially constitutes.

VI

Of course, this case might strike you differently. So, instead, I will play a little defense. Sometimes, the best way to defend a view is to dispel objections to it. Consider, then, the following thought:

My welfare must not be alien to me, a value that floats down from some Platonic realm and, remora-like, affixes itself to me with little regard to the particulars of my constitution. . . . What counts toward my well-being must not depend on what any other individual, *or group or class* of individuals—actual or hypothetical—is like. It must be possible to specify the ultimate or fundamental conditions for my well-being without making essential reference to other individuals, or to classes or groups of individuals.²³

The thought here is that if something is *my* welfare, then this cannot depend upon what anything *other* than me is like. Call this claim the *non-alienation claim*.²⁴ I think there is something right about the non-alienation claim. My welfare must not be completely alien to me. And I will not take issue with the claim that “it must be possible to specify the ultimate or fundamental conditions for my well-being without making essential reference to *other* individuals,

23 Haybron, *The Pursuit of Unhappiness*, 157, emphasis added.

24 Railton (“Facts and Values”) is perhaps the *locus classicus* of this non-alienation claim.

or to [*other*] classes or groups of individuals” (emphasis and brackets added), although May’s remarks discussed earlier suggest otherwise.

Instead I want to question whether groups that I partially constitute *are* indeed alien to me. For these groups do *not* exist in a “realm” different from the realm I inhabit. Nor are they different from me in the way that a remora is different from me. And some groups to which I belong *do* have some regard for the particulars of my constitution.

Rather, when I am part of a “We,” this “We” is *not* something I am separate from, or other than, or alienated from. Put another way, *both* “We” and “I” are first-person pronouns. Just as my finger is not alien from my hand, nor my hand from my body, so too I am not alien from at least some groups that I partially constitute. (After all, we *do* call fingers and hands our *members*.) And, since I am not alien from groups I partially constitute, it still might be true that I do well in virtue of these groups doing well—at least, we should not reject this claim merely due to worries about alienation. Top-down dependence about group well-being, then, need not violate the spirit of the non-alienation claim. In fact, once you appreciate the fact that you can use a first-personal pronoun (“We”) to refer to some such groups, perhaps you can see your way to grasping why your individual well-being can indeed directly depend upon the well-being of groups you partially constitute.

How so? Since the experience of being a member of a baseball team is alien to most of my readers, I will return to the example we all understand better. Imagine again an academic department, a department with various members, one of whom is you. Many things might be good for your department. There are at least two ways your well-being and your department’s well-being can be tethered together.

First, some things are good for your department in virtue of the fact that they are good for you, or for some other departmental member. For example, if you publish an article, that is good for you, and plausibly what is good for you is also good for your department. This exhibits bottom-up dependence, and it is not very controversial. Second, something that is good for your department might be good for you simply in virtue of caring about your department and wanting your department to do well. Here what benefits your department benefits you, but only in virtue of the fact that you care that your department does well. That is not very controversial either.

But can you also benefit from your department’s success not *simply* because your department is in the content of some attitude of yours (e.g., wanting your department to do well), nor because of any other individualistic feature. It might be tempting to think:

Were I *not* to care for my department, its benefitting would *not* benefit me. So, what is good for my department benefits me individually *only because* I care about my department, and thus there is no genuine top-down dependence.

But I do not think we should draw this conclusion. We first need to disambiguate two claims that are easy to conflate:

1. What is good for my department is good for me merely *in virtue of the fact that* I care about my department's well-being.
2. What is good for my department is good for me *only on the condition* I care about my department's well-being.

According to the first claim, what is good for my department is good for me simply because my department is one of the things I happen to care about, and so when my department thrives, one of my desires is thus satisfied, which, in turn, improves my life. And that is not genuine top-down dependence about well-being.

According to the second claim, caring about my department is a *necessary condition* for it to be true that what is good for my department is thus good for me. If I were to care *nothing* about my department, then perhaps I would not benefit when my department benefits. Perhaps my caring is a necessary piece of what must obtain in order for the well-being of my department and my personal well-being to be aptly linked up. Even so, this would be no barrier to there being genuine top-down dependence about well-being between my department and me. Even if my caring is a necessary condition for the well-being of me and my department to link up, it is still possible for me to benefit *in virtue of* my department benefitting.

Here is an analogy to see how this could operate. It is plausible that having an excellent espresso machine is good for me only on the condition that I am not being tortured. For if I am being tortured, my excellent espresso machine does me no good. Nevertheless, it is false that having an excellent espresso machine is good for me *in virtue of* the fact that I am not being tortured. Rather, it is good for me in virtue of other facts. Nor is there some instrumental relation between the two facts. So, we should not confuse a thought of the form—*p* only if *q*—with a thought of the form—*p* in virtue of *q*.

So, similarly, perhaps what is good for your department is good for you only on the condition that you care about your department, but it may still be false that what is good for your department is good for you only and entirely in virtue of the fact that you care about your department. Perhaps once you *do* care about

your department, your department's success then benefits you *directly*. The simple fact that caring about your department plays *some enabling role* in explaining why your personal well-being depends on the well-being of your department is no real threat to top-down dependence about group well-being. Top-down dependence about group well-being is thus still possible.

Stepping back, perhaps you can directly benefit by your group's doing well even if you *do not* happen to care about the group to which you belong. You might not happen to care about the welfare of the working class; but if you are a member of the working class, an injury to the working class might thereby be an injury to you. You might not be very interested in politics, but the erosion of the rights of the *demos* might immediately disable your own freedoms and capabilities, even if you never notice that you have thereby been harmed. If there is top-down dependence, then *your* well-being essentially depends upon the well-being of the groups to which you belong. As such, it is in your interest to understand whether and how the ups and downs of the groups to which you belong affect the quality of your own life. Beware not only the company you keep, but also the company you constitute.

VII

So far, we have explored the idea that the well-being of a group at least partially depends upon the well-being of the individuals in the group, *viz.*, bottom-up dependence. We have also explored the idea that the well-being of individuals partially depends upon the well-being of the group(s) to which they constitute, *viz.*, top-down dependence. I now want to ask whether it is possible that *both* forms of dependence obtain in one and the same group. How *could* *A* constitutively depend upon *B* and *B* constitutively depend upon *A*? Is this viciously circular?

First, note that each form of dependence is only partial. Your well-being might *partially* depend upon the well-being of, say, your department, but also partially depend upon other things. The well-being of your department might likewise *partially* depend upon the well-being of the members of your department, but also partially depend upon other things. If each of those forms of dependence was total rather than partial, then perhaps what I have described would indeed be incoherent.²⁵ But partial interdependence does not seem to

25 For an energetic defense of the coherence of their identity, see Ford and Laurence, "The Parts and Whole of Plato's *Republic*": "For he [Socrates] denies that there is any meaningful conception of a happy human being that is not simultaneously a conception of a happy community, or any conception of a happy community that is not a conception of just human beings. In that case, justice and happiness, both individual and communal, enter the scene at once, or else do not enter at all."

be blatantly incoherent. This combination of top-down partial dependence and bottom-up partial dependence I will call *bidirectional dependence*.

Bidirectional dependence is an exciting idea, but it prompts a worry. It is easiest to explain this worry by illustration. Suppose bidirectional dependence obtains between your department and its members. Now, suppose you individually benefit somehow—perhaps you discover something important, and this redounds to your individual well-being. Given partial bottom-up dependence, your department also benefits in virtue of your increased well-being. So far, so good. But if partial top-down dependence also obtains, then when your department thus benefits, that is good for you ... again. It would seem that you thus benefit *twice*: once when you discover something important, and then again when your department benefits in virtue of the fact that you benefited upon discovering something important.

That seems weird. Moreover, the weirdness does not end there. For if, as described above, you benefit a second time, then this too would seem to benefit your department a second time, if bidirectional dependence obtains. Not only do *you* benefit twice, but your department benefits twice. And there is no reason to stop there. The benefit keeps bouncing up and down. When you discover something important, *that* is good for you, which is good for your department, which is good for you, which is good for your department, which is good for you, *ad infinitum*.

This looks to be a problem. We intuitively know that your personal well-being does *not* rocket to infinity when you discover something important. (Alas!) But it might seem that if bidirectional dependence were to obtain, your personal well-being *would* rocket to infinity when you discover something important. So, we might understandably conclude that bidirectional dependence does *not* obtain. The implications of bidirectional dependence might seem too counter-intuitive to embrace.²⁶

Nevertheless, I think this problem can be sidestepped, and we can rescue the notion of bidirectional dependence from absurdity. We get the race to infinity if personal and group well-being depend on each other *heavily*. But let us see what happens if they depend upon each other modestly. Let us hyper-artificially assign some numbers to these notions. Suppose your personal well-being begins at 50, and your department's well-being begins at 500. Suppose that 50 percent of your department's well-being depends on the well-being of its five members (10 percent each), while the other 50 percent depends upon other things. Suppose

26 Winter ("On the Possibilities of Group Injury") voices a similar worry about double counting. But it is not *just* double!

further that 10 percent of your personal well-being depends upon the well-being of your department.

Imagine, then, that when you discover something important, your individual well-being goes from 50 to 60, thus increasing by 20 percent. But this also benefits your department. Indeed, 10 percent of your department's well-being depends upon your personal well-being, and so when your personal well-being increases by 20 percent, the proportion of your department's well-being that depends upon your well-being increases by 20 percent. This pushes your department's total well-being from 500 up to 510, an overall 2 percent increase. But this 2 percent increase in your departmental well-being then benefits you individually. After all, 10 percent of your personal well-being depends upon the well-being of your department. This pushes your personal well-being from 60 up to about 60.1, a 0.2 percent overall increase. This 0.2 percent increase in your personal well-being then benefits your department again. It will push your department's well-being up from 510 to about 510.1. But all further reverberations quickly approach zero. Neither your well-being nor your department's well-being skyrockets to infinity. Rather, your personal well-being approaches a number less than 61, while your department's well-being approaches a number less than 511. Bidirectional dependence means that each measure is a *little* higher than it would be if there instead were no such dependence, but as long as the dependence is not heavy, things do not explode in an absurd fashion.

Moreover, there is something *plausible* about this way of conceptualizing well-being. It does seem better for you (1) to be a member of a department whose well-being is rather sensitive to your own individual well-being than (2) to be a member of a department whose well-being is completely insensitive to your own individual well-being. Are you not better off when your department thrives as a constitutive upshot of your thriving—better off than you would be if your department were to do *no* better when you do better? Does it not matter that you matter to your group?

If bidirectional dependence can obtain, we can make better sense of various odd features of social life. For instance, the motto of the Wobblies is that “an injury to one is an injury to all.” If there is genuine bottom-up dependence, then an injury to one worker might really be an injury to the union itself. And if there is *also* genuine top-down dependence, an injury to the union might then really be an injury to each union member. So, if bidirectional dependence obtains, the Wobblies' motto is more plausible than it might have initially seemed. Perhaps for some types of groups—ones that are tightly unified and unionized—what helps or harms one member really does help or harm all.

VIII

Which view of group well-being is correct? I see no reason to quickly rule out *any* of the options canvassed above. Perhaps for some kinds of groups, bidirectional dependence obtains; for other groups, only bottom-up dependence obtains; for still other groups, only top-down dependence obtains; and for yet other groups neither form of dependence obtains (that is, the independent conception holds). Each of these logically possible options may be actual in some cases, given what else we know about human sociality. Stepping back further, I also would not be surprised that for some groups, eliminativism about group well-being is correct—perhaps not all groups can be well or ill. The take-home lesson should be that there are quite a few different ways to think about the well-being of groups, and we should not prejudice which of these ways characterizes any particular group.

Can we say more than this? Yes, but not here. I close merely with a suggestion rather than an argument: the nature of group well-being may depend essentially upon the *aims* of the particular group in question. Philosophers tend to think that highly *organized* groups (e.g., corporations, courts) have richer agency, mentality, and responsibility than loosely organized ones. Collins thinks the same way about group duties: whether a group has duties and, if so, how that duty relates to the duties of its members, depends upon the organizational structure of the group in question.²⁷ But I *do not* think the nature of group well-being closely tracks how organized a group is; the well-being of both highly organized groups (Apple) *and* highly unorganized groups (the Washington Square Park chess players) seem to be rather independent from the well-being of their members.

By contrast, bidirectional dependence most likely obtains for groups (1) whose *function* it is to look out for the well-being of their members (families, schools, unions, religious bodies) and (2) whose members tend to *identify* with the groups. Call such groups *mirror groups*, for the well-being of the group and of their members partially reflect each other. If there are such groups, it is better *for you* both to join flourishing mirror groups (rather than non-mirror groups that are flourishing, or struggling mirror groups, or no groups) and to benefit the mirror groups to which you belong (rather than the non-mirror groups to which you belong, or groups to which you do not belong). This is what true community looks like.²⁸

University of Missouri–St. Louis
wiland@umsl.edu

27 Collins, *Group Duties*.

28 Many thanks to the members of audiences at the Kansas Workshop on Well-Being, St. Louis

REFERENCES

- Angner, Erik. "Well-Being and Economics." In *The Routledge Handbook of Philosophy of Well-Being*, edited by Guy Fletcher, 492–503. New York: Routledge, 2015.
- Arrow, Kenneth. *Social Choice and Individual Values*. New Haven, CT: Yale University Press, 1963.
- Bazargan-Forward, Saba, and Deborah Tollefsen, eds. *The Routledge Handbook of Collective Responsibility*. New York: Routledge, 2020.
- Bradley, Ben. *Well-Being*. Cambridge: Polity Press, 2015.
- Collins, Stephanie. *Group Duties: Their Existence and Their Implications for Individuals*. Oxford: Oxford University Press, 2019.
- Copp, David. "The Collective Moral Autonomy Thesis." *Journal of Social Philosophy* 38, no. 3 (Fall 2007): 369–88.
- Di Tella, Rafael, and Robert MacCulloch. "Some Uses of Happiness Data in Economics." *Journal of Economic Perspectives* 20, no. 1 (Winter 2006): 25–46.
- Epstein, Brian. *The Ant Trap: Rebuilding the Foundations of the Social Sciences*. Oxford: Oxford University Press, 2015.
- French, Peter, ed. *Individual and Collective Responsibility*. Cambridge, MA: Schenkman, 1998.
- Ford, Anton, and Benjamin Laurence. "The Parts and Whole of Plato's *Republic*." In *Practical Reason: Historical and Contemporary Perspectives*, edited by James Conant and Sebastian Rödl. Berlin: De Gruyter, forthcoming.
- Haybron, Daniel M. *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*. Oxford: Oxford University Press, 2008.
- Karpus, Jurgis, and Natalie Gold. "Team Reasoning: Theory and Evidence." In *The Routledge Handbook of Philosophy of the Social Mind*, edited by Julian Kiverstein, 400–17. New York: Routledge, 2017.
- Lackey, Jennifer. "Group Assertion." *Erkenntnis* 83, no. 1 (February 2018): 21–42.
- Lewis, H. D. "Collective Responsibility." *Philosophy* 23, no. 84 (January 1948): 3–18.
- List, Christian. "Three Kinds of Collective Attitudes." *Erkenntnis* 79, no. 59 (October 2014): 1601–22.
- List, Christian, and Philip Pettit. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford: Oxford University Press, 2011.

Ethics Group, Rocky Mountain Ethics Conference, University of Memphis, and Colgate University for their comments and enthusiasm. I thank the editors and referees of this journal for their critical feedback. In addition, I thank Kathryn Lindeman, Sam Filby, Peter Zuk, and William Bell for their detailed written comments on earlier drafts.

- May, Larry. *The Morality of Groups: The Nature of Social Groups*. Notre Dame, IN: University of Notre Dame Press, 1987.
- . *Sharing Responsibility*. Chicago: University of Chicago Press, 1992.
- Narveson, Jan. "Collective Responsibility." *Journal of Ethics* 6, no. 2 (June 2002): 179–98.
- Parfit, Derek. *Reasons and Persons*. Oxford: Oxford University Press, 1984.
- Raikka, Juha. "On Disassociating Oneself from Collective Responsibility." *Social Theory and Practice* 23, no. 1 (Spring 1997): 93–108.
- Railton, Peter. "Facts and Values." *Philosophical Topics* 14, no. 2 (Fall 1986): 5–31.
- Rawls, John. "Two Concepts of Rules." *Philosophical Review* 64, no. 1 (January 1955): 3–32.
- Searle, John. "Collective Intentions and Actions." In *Intentions in Communication*, edited by Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, 401–15. Cambridge, MA: Bradford Books, 1990.
- Simon, Thomas W. "Group Harm." *Journal of Social Philosophy* 26, no. 3 (1995): 123–38.
- Sumner, L. W. *Welfare, Happiness, and Ethics*. Oxford: Oxford University Press, 1996.
- Sverdlik, Stephen. "Collective Responsibility." *Philosophical Studies* 51, no. 1 (January 1987): 61–76.
- Tiberius, Valerie. "The Well-Being of Philosophy." *Proceedings and Addresses of the American Philosophical Association* 91 (2017): 65–86.
- Tollefsen, Deborah Perron. *Groups as Agents*. Cambridge: Polity, 2015.
- Winter, Stephen. "On the Possibilities of Group Injury." *Metaphilosophy* 37, nos. 3–4 (July 2006): 393–413.

THE CASE FOR CONSENT PLURALISM

Jessica Keiser

ON THE CREST of the #MeToo movement, a campaign aimed at spreading awareness about the global culture of sexual harassment and assault, website Babe.net published a story in which an anonymous woman detailed her date with comedian and self-proclaimed feminist ally Aziz Ansari. Under the pseudonym “Grace,” she described an evening in which Ansari disregarded her repeated attempts to communicate discomfort with his sexual advances, calling it “the worst night of [her] life” as well as an experience she has come to recognize as “sexual assault.” The flurry of public response involved a number of high-profile op-eds in journals, including *The Atlantic* and *The New York Times*, that decried what they saw as the passive self-victimization of Grace and the overreach of the #MeToo movement. Others, in contrast, argued that Grace’s all-too-familiar experience highlights the need for radical change in our conventions and conversations involving consent.

Some of the key issues emerging within public discourse in the wake of the Ansari scandal reflect a long-standing philosophical debate about the nature of consent—in particular, whether consent is best construed as a type of mental state or a form of behavior. For instance, some disputes over the status of Ansari’s actions seem to hinge on whether Grace’s preferences and attitudes are taken to determine her consent, or rather her behavioral cues about these mental states—and, if the latter, which *types* of cues are taken to be relevant. The last few decades have seen a deepening divide over this question among philosophers and legal theorists, which is now unfolding in the broader public sphere—but until there is some sort of convergence on what *kind* of thing consent is, we cannot hope to understand the mechanism behind its peculiar power to alter the normative landscape, nor to have a principled basis on which to form moral decisions and social policies.

I argue that the impasse in the debate over the ontology of consent is owed to the fact that prominent accounts on all sides have gotten something crucially right while sharing a common flawed methodological assumption. They are right in that they each identify something that plays consent’s normative role; they are

wrong, however, in their shared assumption that there is one thing that does so uniquely or essentially. Rather, this role is played by both mental states and behavior in *independent* and *context-sensitive* ways.¹ The upshot is that, insofar as we seek to understand consent-based moral transformations, we ought to adopt a pluralistic approach to consent that gives independent weight to the moral contributions of facts about mental states and facts about behavior relative to a context.²

Here is the road map: In section 1, I outline consent's *normative role*—its distinctive power to alter the normative landscape. In section 2, I introduce two competing conceptions of consent—*mental consent* and *behavioral consent*—showing that each plays this role and thus neither plays it *uniquely*. In section 3, I argue that neither mental consent nor behavioral consent plays this role *essentially*, given that (1) there are contexts in which mental consent fails to introduce permissions in the absence of behavioral consent, and (2) there are contexts in which behavioral consent fails to introduce permissions in the absence of mental consent. In section 4, I address potential responses. In section 5, I argue that moving to a hybrid account (in which consent requires *both* behavior and mental states) provides an inadequate solution, as it collapses the complex normative structure arising from two distinct mechanisms for introducing permissions. Section 6 concludes that there is no single type of thing that uniquely or essentially displays consent's peculiar power to alter the normative landscape, and that an adequate theory of consent must adopt a pluralistic approach.

1. THE NORMATIVE ROLE OF CONSENT

I will begin by considering what a theory of consent aims to do. As in any other domain, theorists may approach this topic from divergent intellectual traditions, and with an eye to varying theoretical questions and practical concerns.³ Con-

- 1 Westen makes a similar point in relation to legal consent; however, he is concerned with *legal* rather than *moral* power—specifically, the role of different conceptions of consent within criminal law (*The Logic of Consent*).
- 2 Since this paper was written, Dougherty has published an article arguing that violation of mental and behavioral consent ought to be treated as independent offenses for the purposes of *legal policy*, though he is explicitly neutral on questions of *ontology* (“Affirmative Consent and Due Diligence”). In contrast, this paper argues that they ought to be treated as independent moral mechanisms for the purpose of developing an *ontology of consent*; however, because these issues are related, some of the considerations provided in support of this thesis are congenial to Dougherty's arguments about policy making.
- 3 For instance, some theorists have been interested in consent in relation to political authority (see, e.g., Hobbes, *Leviathan*; Locke, *The Two Treatises of Civil Government*; Beran, *The Consent Theory of Political Obligation*; Estlund, *Democratic Authority*), others in relation to rape law (see, e.g., MacKinnon, *Toward a Feminist Theory of the State*; Archard, *Sexual Consent*;

sequently, there may be a variety of perspectives regarding the *desiderata* for a theory of consent. So while I will not presume to identify a unique or privileged characterization of its goals, it will be helpful to say something about how I am approaching the issue.

I adopt a starting assumption that consent displays an identifying characteristic feature: the power to alter the normative landscape. Of course, characterized so broadly, this power is not unique to consent; normative facts are responsive to myriad kinds of occurrences in the world. But consent functions to alter the normative landscape in a distinctive and somewhat puzzling way. First, consent is something that agents *do* in some important sense, and this is often described in terms of being a manifestation of *agency*.⁴ We must tread carefully here, however: there are open and controversial issues in the philosophy of agency that are often breezed over when this term is co-opted in the literature on consent. To pack too much into this notion of *being a manifestation of agency* (stipulating that it must be an action, or that it must be intentional, etc.) would beg the question against certain accounts from the outset; therefore, I assume only that consent is an action/mental state of the consentor, which manifests agency in the sense of being responsive to her reasons and settling certain facts in the world.⁵ Second, consent is *intentional* in the sense of bearing content—it is about, or directed toward, the act consented to.⁶ Third, consent is *other-regarding*—by consenting, the consentor alters the duties of another moral agent. Fourth, this alteration involves *rendering permissible* an action that would have otherwise been impermissible.⁷ The distinctive nature of consent is often contrasted with that of promising.⁸ While

Ferzan, “Consent, Culpability, and the Law of Rape”), and others in relation to medical practice (see, e.g., Faden and Beauchamp, *A History and Theory of Informed Consent*; Manson and O’Neill, *Rethinking Informed Consent in Bioethics*).

- 4 Cf. Pateman, “Women and Consent”; Schulhofer, “The Feminist Challenge in Criminal Law”; Hurd, “The Moral Magic of Consent”; Kleinig, “The Nature of Consent”; West, “Sex, Law, and Consent”; Anderson, “All-American Rape”; Dougherty, “Sex, Lies, and Consent”; Manson, “Permissive Consent.”
- 5 Responsiveness to reasons is standardly taken to be characteristic of moral and epistemic agency. For discussion about how manifestations of agency settle facts in the world, see Steward, *A Metaphysics for Freedom*.
- 6 Cf. Hurd, “The Moral Magic of Consent”; Archard, *Sexual Consent*; Wertheimer, *Consent to Sexual Relations*; Owens, “The Possibility of Consent”; Dougherty, “Sex, Lies, and Consent”; Alexander, “The Ontology of Consent”; Manson, “Permissive Consent.”
- 7 Cf. Hurd, “The Moral Magic of Consent”; Archard, *Sexual Consent*; Wertheimer, *Consent to Sexual Relations*; Westen, *The Logic of Consent*; Owens, “The Possibility of Consent”; Alexander, “The Ontology of Consent”; Dougherty, “Yes Means Yes”; and Alexander, Hurd, and Westen, “Consent Does Not Require Communication.”
- 8 Cf. Shiffrin, “Promising, Intimate Relationships, and Conventionalism”; Owens, “The Pos-

both are content-bearing exercises of individual agency, promising introduces an *obligation* for its *performer*, while consent introduces a *permission* for its *receiver*. Fifth, consent alters the normative landscape *via* altering the normative relationship between its giver and receiver; it relieves the consentee of a *directed* duty to refrain from performing a certain action, which is owed to the consenter. Finally, consent removes certain—but not necessarily all—barriers to permission.

There are broadly two classes of cases in which consent may fail to introduce an *all-things-considered* permission.

1. When there are independent, overriding sources of impermissibility.

How to non-circularly characterize what qualifies as “independent” is a vexed issue that we will return to in section 4. To get a grip on things for now, we can simply think of *independent* prohibitions against an agent’s performing an action as those grounded in reasons *other than* the fact that the action would violate the would-be consenter’s consent.

2. When consent violation is not prohibited to begin with—either because it does not wrong the would-be consenter (I do not wrong you by failing to obtain your consent to water my houseplants), or because the wrongdoing is justified by independent moral considerations (I am permitted to borrow your car without consent to drive a sick child to the hospital, even if this consent violation wrongs you).

Let us call this second class of cases *trivial cases*. I will for the most part ignore trivial cases, and the characterization below can be understood as implicitly bracketing them off. I will return to the first class of cases in section 4.⁹

We can think of the property outlined above as consent’s normative role:

Normative Role of Consent: Being a content-bearing act/mental state that manifests the consenter’s agency and introduces a permission for the consentee (absent independent, overriding sources of impermissibility), thus altering the normative relationship between them.¹⁰

Though this dialectic is not usually made explicit, debates over the ontology of consent often take the normative role of consent as an anchor or starting point; fit with this role is used to adjudicate between competing accounts. Ontological

sibility of Consent”; and Dougherty, “Fickle Consent” and “Yes Means Yes.”

9 Theorists sometimes appeal to the distinction between *valid* consent—in which consent succeeds in altering the normative landscape—from *invalid* consent, in which it fails.

10 This characterization is not intended to be exhaustive; it is designed to include uncontroversial features that do not stack the deck on either side of the ontological debate.

theories of consent typically take for granted that the normative role of consent is its identifying feature and aim to characterize the *kind* of thing that plays this role; they are justified at least partially to the extent that it succeeds in this task, and criticized at least partially to the extent that they fail to do so.¹¹

My approach follows this dialectic in assuming that a primary desideratum of an ontology of consent is to provide an account of *kind of thing* that plays its normative role. However, I reject a crucial assumption underlying this dialect, which is that there is something that does so uniquely and essentially.¹² I argue that both mental consent and behavioral consent play this role, though contingently and for different reasons; insofar as a theory of consent aims to elucidate its normative power, it must take a pluralistic approach that recognizes *both* mental and behavioral consent as independent and context-sensitive mechanisms for changing the normative landscape.

2. NON-UNIQUENESS OF NORMATIVE ROLE: TWO CONCEPTIONS OF CONSENT

In this section I introduce two competing conceptions of consent: *mental consent**, which takes consent to be a kind of mental state, and *behavioral consent**, which takes consent to be a form of behavior. I include a star to avoid the misleading impression of begging the question by presupposing that either of these conceptions *is*, in fact, consent—rather, we will treat them as *candidates* to be evaluated. The central claim of this section is that the normative role of consent does not lie uniquely within the domain of either of these candidates, but is filled by both. I support this claim by showing that there are at least some kinds of mental states and some forms of behavior that play consent’s normative role—thus, neither mental states nor behavior do so uniquely.

2.1. *Mental Consent**

According to one prominent position, consent is a kind of attitude or mental state—I will call this *mental consent**. Adherents of this position will adopt something like the following schema, offering various substitutions for ψ :

- 11 This approach has much in common with a methodology championed by philosophers, including David Lewis and Frank Jackson, known as the “Canberra Plan,” which proceeds in roughly two steps. The first step can be thought of as conceptual analysis: here we identify the theoretical role of the object of inquiry. The metaphysics comes in at the second step: here we seek to identify what kind of thing plays that theoretical role.
- 12 These additional assumptions also part ways with the Canberra Plan, in many of its forms: it was typically assumed that different kinds of things would play the same theoretical role across worlds.

A mentally consents* to B performing ϕ just in case A has mental state ψ .

Hurd, for instance, characterizes ψ as the intention for B to perform ϕ , while Alexander suggests that it is A's waiving their right to complain about a moral or legal boundary crossing resulting from B's performing ϕ .¹³ Westen describes it as state of mind of acquiescence to B's performing ϕ .¹⁴

My aim in this section is to establish that there is *at least some* precisification of this schema (that is, some value of ψ) such that mental consent* plays consent's normative role. I remain neutral on how best to characterize this mental state (or states); however, I will adopt a simplified *toy theory* for illustration. This theory is intended to facilitate clarity in the following discussion, rather than as a serious proposal for how to characterize mental consent*. However, if I can show that the mental state used in the toy theory plays the normative role of consent, I will have established that there is *at least one* mental state that does so. The reader may substitute her preferred characterization of ψ ; if she agrees that there is *some* way of filling out the schema such that mental consent* plays consent's normative role—regardless of whether she agrees that the toy theory meets this condition—I will have achieved my aim. I will have failed if the reader doubts that there is *any* way of filling out the schema that does so.

Here is the toy theory I will adopt:

*Mental Consent**: A mentally consents* to B performing ϕ just in case A is not opposed to B performing ϕ .

According to this theory, mental consent* marks a threshold in an agent's preference ordering. Above the threshold are states of affairs to which she consents, including those for which she has a positive preference, as well as those toward which she is merely indifferent. Below are states of affairs to which her preference ordering assigns a sufficiently low value—i.e., those to which she is opposed. One way in which this toy theory may be too simple to serve as a serious characterization of mental consent* is that we might want to allow that agents can prefer things to which they do not consent, and vice versa. Recall, however, that my aim is not to offer a proposal for how mental consent* ought to be defined, but rather to establish that there is *at least one mental state* that plays consent's normative role.

As noted earlier, we will bracket off trivial cases in which consent fails to introduce permissions because the relevant action is not morally prohibited to begin with.¹⁵ On this conception of mental consent*, nontrivial cases are those

13 Hurd, "The Moral Magic of Consent"; and Alexander, "The Moral Magic of Consent (II)."

14 Westen, *The Logic of Consent*.

15 For instance, I may prefer that a stranger not sit in the seat next to me on the bus, but they

in which performing an act in opposition to another's preferences wrongs them, and the prohibition against this wrongdoing is not overridden by independent moral considerations. I will remain neutral about what *grounds* the fact that acting in opposition to another agent's preferences wrongs her in some contexts—because such preferences track her welfare, or because doing so would constitute a rights violation, etc.—but I take it as uncontroversial that this is so. Tellingly, examples can be found in paradigmatic contexts where consent is morally significant, such as sexual activity. West, for instance, suggests a number of potential sources of the “distinctive harm” of unwanted sex:

Such sex is likely to be alienating, and in something like the original sense of that word: it alienates a girl or woman from her own desires and pleasures, and from that sense of unified identity that comes from acting in the world on the basis of one's own desires and pleasures. . . . And—if it becomes a central part of a life that ties her existence, survival, and hence her interests to that of another—if unwanted sex is the *raison d'être* for a way of life that limits her mobility, her ambition, and the development of her talents or remunerative skills—it constitutes a threat to her autonomy, likewise.¹⁶

I will use unwanted sex as a paradigm case of wrongdoing by violating mental consent*; however, readers may substitute a different case if they find it contentious.¹⁷ The arguments that follow rest only on the assumption that there is at least one mental state ψ and at least some action ϕ , such that *B* wrongs *A* by performing ϕ if *A* fails to have mental state ψ . I take it that this holds true when the relevant action is sexual activity with *A* and the relevant mental state is failure to oppose it, and will proceed with this example in the remainder of the paper—the reader, however, may substitute any mental state ψ and action ϕ that she finds convincing.

Because engaging in sexual activity with a person in opposition to their preferences wrongs them (I will set aside issues of blameworthiness and return to them later), *A* has a directed duty toward *B* to refrain from engaging in sexual activity with her without her mental consent*. If *A* gives her mental consent*, she relieves *B* of this duty, thereby introducing a permission absent independent and overriding sources of prohibition.

are nonetheless permitted to do so. It is notable that mental consent* still plays an important social function even in many trivial cases; it is common for a stranger on the bus to ask “May I?” before sitting down, and there is a general expectation that others will honor our preferences if it comes at no cost to themselves.

16 West, “Sex, Law, and Consent,” 27.

17 Cf. Dougherty, “Affirmative Consent and Due Diligence.”

If the foregoing is correct, mental consent* plays the normative role of consent: it is a manifestation of the consenter's agency in that it is a mental state that is responsive to reasons and settles certain facts in the world—in this case, facts about what is morally permissible for others. (Again, readers preferring a more stringent account of agency can feel free to substitute a different account of mental consent; my argument only depends on there being some mental state that satisfies these conditions.) It is *intentional* in the sense of bearing content—the relevant preferences are directed toward the act being consented to. It is *other regarding* in that an agent's mental consent introduces a permission for its receiver absent independent, overriding sources of impermissibility. It alters the normative relationship between the consenter and consentee by relieving the latter of a *directed* duty owed to the former: the duty to refrain from violating their preferences.

2.2. Behavioral Consent*

An alternative view is one that takes consent to be some form of behavior—I will call this *behavioral consent**. (Again, I add a star to highlight that we are not presupposing that this conception *is*, in fact, consent.) While few adopt a *purely* behaviorist view, consideration of this position will serve to map out the logical space before turning to a hybrid account in section 5. Adherents of this position will adopt something like the following schema, offering various substitutions for ψ :

A behaviorally consents* to B performing ϕ just in case A performs behavior ψ .

Here I argue that there is *at least some* precisification of this schema (that is, some value of ψ) such that behavioral consent plays consent's normative role. Again, I will remain neutral on how best to characterize the particular behavior (or behaviors) that constitutes behavioral consent*, but will adopt a working theory for clarity. As before, the reader can feel free to substitute her preferred characterization of ψ in the discussion that follows; as long as the reader agrees that there is *some* way of filling out the schema such that behavioral consent* plays consent's normative role, I will have achieved my aim.

I will adopt the following toy theory:

*Behavioral consent**: A behaviorally consents* to B performing ϕ just in case A performs some action that signals to B that she mentally consents* to B performing ϕ .

We will understand signaling in a rough and ready way, leaving it deliberately unspecific in order to accommodate a range of approaches. This theory characterizes behavioral consent* as an action that functions to provide the consen-

tee with evidence about the consentor's mental states. We will assume for now that signaling need not be factive; in providing evidence about her mental state, the consentor may be disingenuous or inaccurate. (We will consider hybrid accounts, which may introduce factivity, in due course.)

The last section claimed that there is at least one mental state ψ and at least some action ϕ , such that B wrongs A by performing ϕ if A fails to have mental state ψ . If this is so, then performing such actions without sufficient evidence of such mental states puts the would-be consentor at *risk* of being wronged.¹⁸ It is commonly acknowledged that placing someone at risk of being wronged—regardless of whether that wrong comes to fruition—is morally prohibited when this risk is sufficiently high and there are not overriding moral reasons in favor of performing the action. For instance, it is morally prohibited to randomly fire shots into a crowd regardless of whether anyone gets hit, and it is morally prohibited to blow up a building without knowing whether it is occupied, regardless of whether it turns out to be empty; many, including Jackson and Zimmerman, have emphasized that an adequate moral theory must accommodate such facts.¹⁹ As Zimmerman points out, one may do so using a variety of moral frameworks, whether they take moral permissibility to be determined by “objective” or “subjective” facts. For instance, on the former approach one might take facts about permissibility to be determined by objective value, while assigning objective disvalue to the running of risks.²⁰ On the second approach, one might take facts about permissibility to be determined by the agent's beliefs—or the evidence available to her—regarding objective value.²¹ So there are a range of moral theories able to accommodate the impermissibility of risking wrongdoing—but however this is cashed out, I take Zimmerman to be correct in claiming that it is unconscionable to take certain risks of wronging others.

In certain types of cases, the evidence required for reducing the level of risk to a degree sufficient for eliminating this source of moral prohibition may need to take a certain form. Consider, again, the case of unwanted sex: if engaging in sexual activity absent mental consent* is morally prohibited, then doing so without sufficient evidence of a partner's mental consent* puts her at risk of such

18 For extensive discussion of the significance of risk to our practices of consent, see Bolinger, “Moral Risk and Communicating Consent.”

19 Jackson, “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection”; and Zimmerman, “Is Moral Obligation Objective or Subjective?”

20 Sosa, “Consequences of Consequentialism.” For discussion, see Zimmerman, “Is Moral Obligation Objective or Subjective?”

21 See, e.g., Jackson, “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection”; and Zimmerman, “Is Moral Obligation Objective or Subjective?”

wrong and thus is in itself morally prohibited. Plausibly, there will be cases in which such evidence must come in the form of behavioral consent*. Consider again the date between “Grace” and Aziz Ansari: given that they were relative strangers lacking the sort of history and background that would put them in a position to make reliable judgments about each other’s mental states, the kind of evidence required to avoid running a prohibitive risk of mental consent* violation would need to come from the other’s behavior—that is, their *behavioral consent**.²² Leaving aside the question of exactly what kind of behavior is required, given the assumption that behavioral consent* *in some form* can provide the requisite evidence of mental consent*, it can introduce a permission by eliminating an unjustified risk of wrongdoing.

Therefore, like mental consent*, behavioral consent* exhibits the characteristic normative power of consent. It involves *agency*, manifested through the behavior of the consenter. This behavior is responsive to the consenter’s reasons and settled facts in the world—in particular, moral facts, facts about risk, and facts about evidence available to the consentee. It is *intentional* in the sense of bearing content—the agent’s behavior constitutes a signal whose content involves the act behaviorally consented* to.²³ It is *other regarding* in that an agent’s behavioral consent* alters the duties of another by introducing a permission absent independent overriding reasons. It alters the normative relationship between the consenter and consentee by relieving the latter of a directed duty owed to the former: the duty to refrain from running the risk of wronging them.

I have argued that both mental consent* and behavioral consent* can play the normative role of consent, which entails that this role is not filled *uniquely* by either. Though I used concrete theories of mental and behavioral consent* for illustration, my argument did not rely on these particular theories. The crucial assumption was, rather, that there is at least one mental state ψ and at least some action ϕ , such that *B* wrongs *A* by performing ϕ if *A* fails to have mental state ψ . If this mental state has the additional features of being directed toward ϕ , it follows that it plays the normative role of consent; adopting this mental state is a content-bearing action/mental state that introduces a permission for its receiver absent independent, overriding sources of impermissibility, thus altering the normative relationship between consenter and consentee. My argument that

22 Some have argued that the requisite evidence in such a case need always be behavioral. See, e.g., Dougherty, “Yes Means Yes” and “Affirmative Consent and Due Diligence”; and Guerero, “The Epistemology of Consent.”

23 Cf. Hurd, “The Moral Magic of Consent”; Archard, *Sexual Consent*; Wertheimer, *Consent to Sexual Relations*; Owens, “The Possibility of Consent”; Dougherty, “Sex, Lies, and Consent”; and Alexander, “The Ontology of Consent.”

there is at least some behavior that plays the normative role of consent crucially relied on one additional premise: that running an unjustified risk of wronging someone wrongs them. This entails that *B* wrongs *A* by performing ϕ if she runs an unjustifiable risk that *A* fails to have mental state ψ . On the plausible assumption that *A* can perform some behavior that sufficiently lowers this risk, it follows that there is some form of behavior that plays the normative role of consent by introducing a permission in the relevant way.

3. NON-ESSENTIALITY OF NORMATIVE ROLE

In the previous section I argued that both mental consent* and behavioral consent* can play the normative role of consent, thus this role is not filled *uniquely* by any one kind of thing. Here I argue that neither mental consent* nor behavioral consent* plays this normative role *essentially*. I demonstrate this by showing that there are contexts in which each of them fails to introduce permissions in the relevant way. Again, I will rely on my working theories for clarity, but the main claim does not rely on these particular conceptions of consent. I consider potential responses in section 4.

3.1. *Mental Consent* without Behavioral Consent**

While behavioral consent* is importantly related to mental consent* it is a mistake to suppose—as often happens in the literature on consent—that it is simply an “analogue” of mental consent*, in the sense of being an imperfect stand-in for what is *truly* morally significant. This is because behavioral consent* transforms the normative landscape through a mechanism that functions *independently* from mental consent*: the wrong does not involve acting in opposition to an agent’s preferences, but rather running an unjustified *risk* of doing so. So mental consent and behavioral consent* are related in that the latter is *conceptually* and *explanatorily* derivative of the former. Behavioral consent* is conceptually derivative of mental consent* in the sense that its analysis is given in terms of mental consent*—specifically, running the risk of its violation. It is explanatorily derivative in the sense that its normative features are explained in terms of the normative features of mental consent*.²⁴ However, their normative roles are *functionally* and *metaphysically* independent in the sense that the wrong of an instance of behavioral consent* violation need not depend upon the wrong of an instance of mental consent* violation; i.e., an action can be wrong on account of violating behavioral consent* without violating mental consent*. Again, consider the context of a sexual encounter between relative strangers: as noted above, because

24 Thanks to David Enoch for this point about explanation.

participants may lack the sort of history and background that would put them in a position to make reliable judgments about one another's mental states, the kind of evidence required to avoid running a prohibitive risk of mental consent* violation would need to come from the other's *behavioral consent**.²⁵ In such cases, violating behavioral consent* is morally prohibited, *even when mental consent* is not violated*. Its moral significance is tied to the impermissibility of running certain kinds of risks, and is not always vitiated when such risks fail to eventuate.

If this is the case, then mental consent* does not play the normative role of consent *essentially*; there will be cases in which it fails to introduce permissions due to lack of behavioral consent*. Again, this is not contingent upon the particular toy theory of mental consent* I have offered here—it only relies on the assumptions detailed at the end of the last section: as long as (1) there is at least one mental state ψ and at least some action ϕ , such that B wrongs A by performing ϕ if A fails to have mental state ψ and (2) B wrongs A by performing ϕ if she runs an unjustifiable risk that A fails to have mental state ψ , then mental consent cannot perform the normative role of consent essentially. There will always be the possibility of cases in which A has the requisite mental state, but because A has not signaled with her behavior that she has this mental state, in performing ϕ B wrongs her by running the unjustifiable risk that she does not. In such cases, mental consent* fails to introduce a permission in virtue of the absence of behavioral consent*.

3.2. Behavioral Consent* without Mental Consent*

Their metaphysical and functional independence similarly entails that behavioral consent* may fail to introduce permissions in cases where mental consent* is absent. If sexual activity can wrong someone when it occurs in opposition to their preferences, then it can do so in spite of misleading *evidence* regarding those preferences. Recall that we are dealing with two distinct prohibitions: one against violating an agent's mental consent*, the other against running an unjustified risk of doing so. In the case that an agent behaviorally consents*—i.e., signals mental consent* through her outward behavior—then insofar as risk of violating mental consent* is measured by her discernible actions, her partner will not be guilty of unjustifiably running that risk.²⁶ But even low risks can eventuate: in the case that an agent fails to mentally consent* to sexual activity, her partner will be guilty of violating her mental consent* even while he has not unjustifiably run

25 One might argue that the requisite evidence need always be behavioral. I am making the weaker claim that, at least in some cases, acting without behavioral consent is prohibited.

26 Though of course there may be other evidence that there is a risk involved, such as a problematic power dynamic.

the risk of doing so.²⁷ Depending on the situation, he may not be blameworthy, but he will nevertheless have done wrong (more on this below).

Note, again, that this result does not depend on our working theories of mental and behavioral consent*, but rather on the assumption that there is at least one mental state ψ and at least some action ϕ such that B wrongs A by performing ϕ if A fails to have mental state ψ —*even if B does not run an unjustifiable risk of doing so*. Nor does it matter how restrictive we make an account of behavioral consent*; as long as it does not entail mental consent there will be cases in which the two come apart.²⁸ And as long as some of those cases are such that performing an action without the person's mental consent* wrongs her, the action will be prohibited regardless of the presence of behavioral consent*. Therefore, behavioral consent* does not play the normative role of consent essentially; there will be cases in which it fails to introduce permissions due to lack of mental consent*.

4. RESPONSES

In this section I survey a number of responses to the arguments given in sections 2 and 3. First I consider how the mental consent* theorist might deny the claim that behavioral consent* plays the normative role of consent—thus enabling her to reject the corollary claims that this role is not unique or essential to mental consent*. I argue that this move commits the mental consent* theorist to an unpalatable moral theory. Next I consider how the behavioral consent* theorist might deny the claim that mental consent* plays the normative role of consent—thus enabling her to reject the corollary claims that this role is not unique or essential to behavioral consent. I argue that this move leads to an unpalatable theory of consent. Finally, I consider a potential response on behalf of each position: recall that the normative role of consent is being a content-bearing exercise of individual agency that introduces a permission for its receiver *absent independent, overriding sources of impermissibility*. This qualification leaves room for both the mental consent* theorist and the behavioral consent theorist to file the normative role of the other as one of the “independent” considerations to be bracketed off. I argue that this response is question begging in light of mutually accepted theoretical constraints governing the ontology of consent.

4.1. Denying the Normative Role of Behavioral Consent*

The argument that the normative role of consent is neither unique nor essential

27 Following Nagel in “Moral Luck,” we might classify this as a case of bad resultant moral luck.

28 For examples and discussion of various “infelicitous” ways that mental consent* and behavioral consent can come apart, see Sherwin, “Infelicitous Sex.”

to mental consent* hinged on the claim that behavioral consent* also plays this normative role; it follows that (1) this role is not *unique* to mental consent*, and (2) because a behavioral consent* violation can block permissions that would otherwise have been introduced by mental consent*, this role is not *essential* to mental consent*. The mental consent* theorist can resist these conclusions by denying that violating behavioral consent* can constitute a wrong. In this case, behavioral consent* would not play the normative role of consent; it would fail to introduce permissions for the trivial reason that its violation was not prohibited to begin with. And if violation of behavioral consent is not prohibited, then there will not be cases in which such violations block permissions that would otherwise be introduced by mental consent*. Indeed, proponents of the mental consent* view standardly claim that we are incorrect in judging risk-running acts of behavioral consent* violation to be wrong—our sense of wrongdoing is actually tracking the fact that they are *culpably performed*.²⁹ The distinction between wrongdoing and culpability, they argue, is crucial to both legal and moral theorizing. What determines *wrongdoing* is whether an agent's action violates the dictates of morality; in contrast, certain facts about her mental states in relation to that action—whether she intended to do wrong, could foresee the possibility of doing wrong, etc.—determine her *culpability*. With this distinction in hand, mental consent* theorists argue that in cases where behavioral consent* is violated but mental consent* has been obtained, agents are “merely culpable” but have done nothing morally impermissible. Considering a hypothetical case of sexual activity between Sam and Sue, Alexander, Hurd, and Westen claim that

if Sam believed there was a significant risk that [Sue] was not [mentally] consenting, then Sam acted culpably. And indeed, such culpability might well be sufficient to justify his being criminally punished, e.g., for attempt. But Sam did not wrong Sue.³⁰

The problem with this response is that one may accept the distinction between culpability and wrongdoing without endorsing the verdict that such cases are morally permissible. This is because *culpability* with respect to one action may imply *wrongdoing* with respect to another. If the foregoing arguments are correct, there are two related but distinct potential sources of wrongdoing to be considered in the case of sexual activity: (1) wronging someone by violating their mental consent*, and (2) wronging someone by unjustifiably running the risk of doing so. We can accept the culpability/wrongdoing distinction and concede that in such examples Sam is culpable but has done no wrong with respect to (1). How-

29 See Alexander, Hurd, and Westen, “Consent Does Not Require Communication.”

30 Alexander, Hurd, and Dougherty, “Consent Does Not Require Communication,” 656.

ever, if it is indeed the case that unjustifiably putting people at certain kinds of risks is itself wrong, then Sam is not merely culpable, but has violated the dictates of morality with respect to (2). A certain kind of culpability with respect to the former entails wrongdoing with respect to the latter.³¹ Thus, merely invoking the culpability/wrongdoing distinction does not help the mental consent* theorist here; rather, she must deny that placing others at significant and unjustified risk of being wronged is morally prohibited. This is a big bullet to bite; indeed, as discussed in section 2.2, people find it to be so unacceptable that there has been substantial work in tailoring moral theories to be able to accommodate such facts.³²

Perhaps there is another line of response for the mental consent* theorist, however; she may want to grant that unjustifiably running the risk of wronging someone is morally impermissible, but that it does not constitute a *directed* wrong. That is, she may want to claim that *B* acts impermissibly by running an unjustifiable risk of wronging *A*—but that in so doing, *B* does not *wrong A*.³³ Because the normative role of consent involves removing a barrier to permission by relieving the consentee of a directed duty to the consentor, it follows that violating this directed duty is not merely impermissible—it is a wrong that is *directed* toward the would-be consentor. Thus, if the act of running unjustifiable risks of wronging others does not constitute a *directed* wrong, then behavioral consent fails to play the normative role of consent—in which case my argument that mental consent* fails to play the normative role of consent uniquely and essentially does not go through.

I want to argue, however, that there are at least some cases in which such risk running violates a *directed* moral duty, and thus constitutes a directed wrong. I will not take a stand on the “direction problem,” i.e., the issue of what *grounds* the fact that a duty is directed; such a commitment would be beyond the scope of this paper, and the broader argument of this paper should be applicable to a variety of approaches to this issue. Instead, I will appeal to the widely accepted view (perhaps most familiar from Darwall) that directed duties standardly exhibit a certain feature that can be used as a diagnostic test: namely, that when a directed duty toward *B* has been violated by *A*, *B* can legitimately blame *A* and hold *A*

31 Just like mere culpability with respect to the crime of murder entails guilt with respect to the crime of attempted murder, we might think that negligence with respect to one action may entail wrongdoing with respect to another.

32 See, e.g., Jackson, “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection”; and Sosa, “Consequences of Consequentialism.” For discussion, see Zimmerman, “Is Moral Obligation Objective or Subjective?”

33 Thanks to an anonymous reviewer for suggesting this line of response.

accountable.³⁴ This feature is present in cases in which *A* runs an unjustifiable risk of wronging *B*; for instance, if *A* wrongs *B* by running an unjustifiable risk of shooting her in the head (say, by firing a partially loaded gun at her head without knowing which chamber the bullet is in), then *B* can legitimately blame *A* and hold her accountable. This feature is also present in the case of sexual activity: *B* can legitimately blame *A* and hold her accountable if *A* unjustifiably runs the risk of violating her mental consent to sex. It is not merely that *A* violates a general moral duty by running such risks; she violates a *directed* duty toward *B*. If the moral consent* theorist accepts that running the risk of doing wrong is at least in some cases a violation of a directed duty, then she must also accept that at least in some cases mental consent* fails to secure permissions; thus, she is forced to deny that mental consent* plays the normative role of consent uniquely and essentially. On the other hand, if she wants to hold on to the commitment to the view that mental consent* plays the normative role of consent uniquely and essentially, she will find herself in the uncomfortable position of being forced to adopt a very strong and controversial broader moral claim about risk.

4.2. Denying the Normative Role of Mental Consent*

The argument that the normative role of consent is neither unique nor essential to behavioral consent* worked in roughly the same way; the fact that mental consent* plays this role established that it is not *unique* to behavioral consent*, and because failure to obtain mental consent* can block permissions that would otherwise have been introduced by behavioral consent*, this role is not *essential* to behavioral consent*. Again, the behavioral consent* theorist can resist these conclusions by denying that violating mental consent* can constitute a wrong. In this case, mental consent* would not play the normative role of consent; it would fail to introduce permissions for the trivial reason that its violation was not prohibited to begin with. Moreover, if violation of mental consent* is not prohibited, then there will not be cases in which such violations block permissions that would otherwise be introduced by behavioral consent*.

The most plausible way to deny that violating mental consent* constitutes a wrongdoing is to adopt an agent-relative moral theory. One might claim with Frank Jackson, for instance, that facts about permissibility are determined by the mental states of the agent performing the act in question (here, the would-be consentee, rather than consentor)—such as her subjective probabilities about moral value. To illustrate, let us consider the case of Grace and Ansari: someone with this

34 Darwall, *The Second-Person Standpoint*. See also Mill, *Utilitarianism*, ch. 5; Brandt, *A Theory of the Good and the Right*, 163–76; Gibbard, “Wise Choices, Apt Feelings,” 41; and Wallace, “Reasons, Relations, and Commands,” 29.

view on permissibility might claim that though violating Grace's preferences regarding their sexual activity had moral disvalue, in doing so Ansari did not wrong her (assuming—implausibly, in my view—that his subjective probability that he was violating her mental consent* was sufficiently low). This is because the facts about permissibility are not simply determined by facts about moral value, but rather by facts about moral agents' *subjective probabilities* about such value. Thus, whether or not Ansari wronged Grace is determined by his subjective probability concerning whether she mentally consented*, not whether she in fact did so.

This kind of approach to moral permissibility is subject to objections that issues of consent bring out particularly well: there are different ways of understanding the claim that permissibility is determined by an agent's subjective probabilities regarding the moral value of her actions. If we take her subjective probabilities to be determined by her *beliefs*, it follows that any case of unwanted sex is permissible so long as the agent believes that the other party prefers it, regardless of how irrational or unjustified this belief might be. Moreover, contexts of sexual activity raise the particular problem that participants may fail to take heed of evidence in forming their beliefs because of bigoted attitudes, or because they would benefit from ignoring it. Participants in the public debate on the Ansari scandal have made this point in numerous ways:

[Men] think everything is fine because they want it to be. For them, it's best when it's OK because that "OK" gets them what they want.³⁵

If your view of women is that they are not really human beings, you're going to have a problem hearing what they say when you're determined to have sex with them.³⁶

These kinds of results can be avoided by taking a more "objective" view of subjective probability—rather than being determined by the agent's beliefs, we take them to be determined by the evidence available to her regarding the moral value of her potential actions. But again, there are problems with this view that are brought out particularly well by thinking about issues of consent. As Zimmerman points out, if permissibility is determined by the evidence one has on hand about the moral value of one's potential actions, then one is not obliged to seek out more information about those facts. But in cases of risk—for instance, the risk of violating another's bodily integrity—it seems clear that agents are obliged to seek out more information rather than acting on the evidence available to them at the time.

35 King, "Aziz Ansari Allegations Show that People Have a Lot to Learn about Consent."

36 Reilly, "Aziz Ansari and the Issue of Consent."

Moreover, there is another more general problem with this approach, which is that the conception of consent it delivers fails to satisfy the normative role of consent. Recall that consent's normative role is being a manifestation of the consentor's agency that can introduce permissions (for the consentee) absent independent overriding considerations. If we take facts about what is permissible for the consentor to be determined by facts about her *own* mental states or evidence, this implies that it is not—strictly speaking—the *consentor* who alters those permissions. Granted, the consentor can perform some behavior that has the contingent causal effect of altering the mental states and evidence of the consentee. Nonetheless, her action is ultimately not playing a metaphysically robust role in altering the normative landscape—while her behavior may have some contingent, *causal* influence over the consentee's mental states and total body of evidence, it is the latter and not the former that *metaphysically determines* what is permissible for him. The consentor, in this case, has no power to *settle the fact of the matter* about what is permissible for the consentee; she may have some causal influence over this, but the matter is ultimately settled by facts about the consentee's mental states. Thus, the change in the normative landscape is not a manifestation of the consentor's agency. This implies that there is nothing that plays the normative role of consent, given that it metaphysically divorces facts about what is permissible for the consentee from the agency of the consentor. Thus the behavioral consent* theorist will have a difficult time denying the impermissibility of violating mental consent* without committing herself to an implausible view of consent—or abandoning the notion altogether.

4.3. *Appealing to Independent Reasons*

I have claimed that in order to resist the argument that the normative role of consent is not essential or unique to mental consent*, the behavioral consent theorist must deny that the violation of mental consent* can constitute a wrong (and vice versa). I have argued, however that the behavioral consent* theorist cannot deny that violating mental consent* is impermissible without committing herself to an unpalatable theory of consent, and that the mental consent* theorist cannot deny that violating of behavioral consent* is impermissible without committing herself to an unpalatable moral theory. There is a second option for both positions, however: each can accept that the violation of the other may constitute a wrong, but deny that this wrong is *consent related*. Instead, they can characterize that wrong as one of the “independent” moral considerations bracketed off by the definition of the normative role of consent.

The characteristic normative power of consent is not limitless; we expect there to be limitations on this power, and so the fact that any particular theo-

ry of consent predicts such limitations will be a virtue rather than a drawback of that theory. The idea is that there are certain actions such that lack of consent provides a certain kind of *pro tanto* moral reason *against* performing them. Absent overriding reasons in *favor* of performing such actions, lack of consent renders them impermissible; when consent is obtained, this particular source of prohibition is eliminated. But, of course, there may be independent moral reasons that count against performing the action—and because consent’s power to transform prohibitions into permissions is a function of eliminating only *certain sources* of prohibition (those generated by lack of consent), we should expect it to be limited to cases in which there are not independent, overriding sources of prohibition. For instance, an act of adultery may be consensual but nonetheless morally prohibited because of the harm it would cause a third party. Our characterization of the normative role of consent brackets off such cases: we have said that consent is a content-bearing act of individual agency that introduces a permission for its receiver *absent independent overriding considerations*. This opens the door for a flatfooted way for both the mental consent* theorist and the behavioral consent* theorist to respond to the arguments above; each may claim that the source of the prohibition identified by the other is *independent* rather than *consent related*. For instance, some mental consent* theorists have noted that violation of behavioral consent* may constitute an “independent” act of wrongdoing; Ferzan, e.g., does not characterize it as risk running, as I have, but rather an act of disrespect.³⁷ A behavioral consent* theorist could, in turn, acknowledge that violating mental consent* is in some cases wrong, but characterize this wrongdoing as independent and non-consent related.

There are two problems with this response. The first is that while it gets around the non-essentiality counterexamples, it does not get around the non-uniqueness counterexamples. I argued above neither mental consent* nor behavioral consent* plays the normative role of consent *essentially*, given that (1) there are contexts in which mental consent* fails to introduce permissions in the absence of behavioral consent*, and (2) there are contexts in which behavioral consent* fails to introduce permissions in the absence of mental consent*. The mental consent* theorist can resist the claim that mental consent* fails to display the normative role *essentially* by suggesting that absence of behavioral consent* merely introduces an *independent* source of prohibition, and vice versa. While this response allows each theorist to insist that her conception of consent plays the normative role of consent *essentially*, it does not provide her with a way to claim that it does so *uniquely*; there will still be cases in which both mental con-

37 Ferzan, “Consent, Culpability, and the Law of Rape.”

sent* and behavioral consent* function to eliminate prohibitions in the particular way that is characteristic of consent.

The second problem with this response is that, in the context of the debate over the ontology of consent, to simply stipulate that the opposing conception of consent is an independent and non-consent-related moral consideration is to flatly beg the question at hand. Neither conception seems to respond better to our ordinary, pre-theoretical conception of consent (as evidenced by the debates in the literature and in the public sphere)—a fact that is readily acknowledged by theorists on both sides, who often cede that use of the word “consent” is polysemous between various readings. The same has been shown to be true in legal contexts.³⁸ Moreover, both conceptions are equally relevant to evaluating moral behavior, often in the very same cases—tellingly, those in which we take consent to be morally relevant. The point is not merely that both conceptions have moral weight in such situations, but that in both cases their normative role functions in the peculiar way that we have pre-theoretically characterized as belonging to *consent*: they both involve manifestations of individual agency directed toward the act in question, and they both introduce permissions for the consentee absent overriding considerations. Given that fit with this role is used to adjudicate between competing ontologies of consent, and given that both mental consent* and behavioral consent* fit it equally well, simply stipulating that the opposing conception is not really consent is a question-begging response to counterexamples to one’s preferred theory.

5. PROBLEMS WITH A HYBRID APPROACH

A natural response would be to adopt a hybrid view, in which an act of consent is taken to require both mental consent* and behavioral consent*.³⁹ This seems to address the worry that neither mental consent* nor behavioral consent* is sufficient on its own to introduce relevant permissions in every context. On this approach, consent is usually characterized as a type of action that necessarily has a mental component—for instance, it is common to claim that consent is an action performed *intentionally*.⁴⁰ The problem with the hybrid view is that it

38 See Westen, *The Logic of Consent*; and Dougherty, “Affirmative Consent and Due Diligence.”

39 See Dougherty, “Yes Means Yes.”

40 See, e.g., Owens, “The Possibility of Consent”; Archard, *Sexual Consent*; Dougherty, “Yes Means Yes”; and Wertheimer, *Consent to Sexual Relations*. Other views that are sometimes categorized under the heading of “hybrid” treat consent as a type of action but add the qualification that a certain kind of mental state may be needed in order for it to be morally transformative. See Miller and Wertheimer, “Preface.”

collapses the structure behind the mechanism that gives consent its normative power; the reason that consent is able to change prohibitions into permissions is not because mental states and behavior combine to produce a mechanism that generates this power. Rather, there are two independent mechanisms, each exhibiting the normative power of consent when taken individually. Mental consent* has this power because it is prohibited to wrong others, and mental consent* can change the facts about whether an action will wrong a person. In contrast, behavioral consent* has this power because it is prohibited to put others at risk of being wronged, and behavioral consent* can change facts about risk. Their combination does not create a new moral reason, as it were, but merely aggregates the weight of two independent moral reasons. In treating consent as a unified phenomenon, the hybrid view threatens to mask the complexity behind consent's normative power.

A second and related problem is that the hybrid view is too strong: the conjunction of mental consent* and behavioral consent* is not necessary for changing prohibitions into permissions.⁴¹ As mental consent* theorists have repeatedly pointed out, there are cases in which behavioral consent* is not necessary to change prohibitions into permissions, because mental consent* alone is sufficient. These are cases in which the risk of violating mental consent* is sufficiently low; in such cases, acting without behavioral consent* does not constitute running a prohibitively high risk of wronging another. Consider, for instance, the following scenario: Mona and Lisa are long-term partners with an active and trusting sexual relationship. One morning Lisa sleeps through her alarm clock, and so Mona wakes her up with a kiss, an act that Lisa mentally consents* to.⁴² Though Mona did not obtain behavioral consent* from Lisa before performing this action, their history together provided her with sufficient evidence that Lisa would mentally consent* to it; the fact that she did not obtain behavioral consent* did not amount to unjustifiably running a risk of wronging her.⁴³ Since Lisa did in fact mentally consent*, Mona did not wrong her in performing this action—neither by violating her preferences nor by running an unjustifiably high

41 Bolinger's version of a hybrid account is disjunctive rather than conjunctive ("Moral Risk and Communicating Consent"). This makes the view too weak (in the sense of being susceptible to the non-essentiality arguments of the last section) rather than too strong.

42 I assume that people can have preferences and other mental states while sleeping. Alternatively, we can imagine that Lisa is merely pretending to be asleep.

43 One might respond that Mona obtained behavioral consent from Lisa, given Lisa's past behavioral cues. But, as campus campaigns for *active, ongoing* consent bring out, such a notion of consent is problematic; contexts and preferences are constantly in flux, and for an agent *A* to consent to sexual activity with another agent *B* at one time and context *does not* amount to her consenting to sexual activity with *B* at all future times and contexts.

risk of doing so. Mental consent*, in this case, was sufficient to transform what would have otherwise been a prohibited act into a permissible one. The hybrid view gets such cases wrong by demanding the presence of behavioral and mental consent* in every context.⁴⁴

6. CONCLUSION

The guiding methodology in developing an ontology of consent has been to identify the kind of thing that plays its normative role—that is, the kind of thing that is a content-bearing manifestation of individual agency that introduces a permission for another absent independent, overriding circumstances, thus altering the normative relationship between consenter and consentee. I have argued that we ought to reject a tacit assumption underling the debate over the ontology of consent—which is that there is a single kind of thing that does this essentially or uniquely. Both mental consent* and behavioral consent* play this role in certain contexts, through different mechanisms; moreover, neither is sufficient on its own to play this role in every context. Recognizing this, we can see how the tacit assumption that consent is a unary kind has led to an impasse in the literature on the ontology of consent: not only is the *ordinary* conception of consent ambiguous between mental and behavioral consent*, but they both play its normative role—often in the very same contexts. This has forced both parties to the debate to dig in their heels and adopt question-begging responses to the counterexamples presented by the other. Though theoretical unity may be preferable, all things being equal, I have argued that all things are not equal; the considerations presented above suggest that we ought to reject the assumption that consent is a unary phenomenon and adopt a pluralistic approach.⁴⁵ Because mental states and behaviors introduce different *kinds* of *pro-tanto* permissions—each of which can be overridden by lack of the other—it is necessary to weigh each of their independent moral contributions in order to understand how consent functions to alter permissions across disparate contexts.⁴⁶

University of Leeds
keiserjessicalynn@gmail.com

44 It is less clear to me that behavioral consent can transform prohibitions into permissions absent mental consent. But all that is needed to show that the hybrid view is too strong is that there are cases in which at least one conjunct is not necessary.

45 Thanks to David Enoch for pushing me to acknowledge the value of theoretical unity.

46 Thanks to Jess Isserow, Renee Bollinger, David Enoch, and Larry Alexander for extensive feedback on written drafts, and to the audience at the Cornell Philosophy of Law confer-

REFERENCES

- Alexander, Larry. "The Moral Magic of Consent (II)." *Legal Theory* 2, no. 3 (September 1996): 165–74.
- . "The Ontology of Consent." *Analytic Philosophy* 55, no. 1 (March 2014): 102–13.
- Alexander, Larry, Heidi Hurd, and Peter Westen. "Consent Does Not Require Communication: A Reply to Dougherty." *Law and Philosophy* 35, no. 6 (December 2016): 655–60.
- Anderson, Michelle J. "All-American Rape." *St. John's Law Review* 79, no. 3 (Summer 2005): 625–44.
- . *Sexual Consent*. Boulder: Westview Press, 1998.
- Beran, Harry. *The Consent Theory of Political Obligation*. London: Croom Helm, 1987.
- Bolinger, Renée Jorgensen. "Moral Risk and Communicating Consent." *Philosophy and Public Affairs* 47, no. 2 (Spring 2019): 179–207.
- Brandt, Richard. *A Theory of the Good and the Right*. Oxford: Oxford University Press, 1979.
- Darwall, Stephen. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press, 2006.
- Dougherty, Tom. "Affirmative Consent and Due Diligence." *Philosophy and Public Affairs* 46, no. 1 (Winter 2018): 90–112.
- . "Fickle Consent." *Philosophical Studies* 167, no. 1 (January 2014): 25–40.
- . "Sex, Lies, and Consent." *Ethics* 123, no. 4 (2013): 717–44.
- . "Yes Means Yes: Consent as Communication." *Philosophy and Public Affairs* 43, no. 3 (2015): 224–53.
- Estlund, David M. *Democratic Authority: A Philosophical Framework*. Princeton: Princeton University Press, 2008.
- Faden, Ruth R., and Tom L. Beauchamp. *A History and Theory of Informed Consent*. New York: Oxford University Press, 1986.
- Ferzan, Kimberly. "Consent, Culpability, and the Law of Rape." *Ohio State Journal of Criminal Law* 13, no. 2 (2016): 397–439.
- Gibbard, Allan. "Wise Choices, Apt Feelings." Cambridge, MA: Harvard University Press, 1990.
- Guerrero, Alexander A. "The Epistemology of Consent." In *Applied Epistemology*, edited by Jennifer Lackey. Oxford: Oxford University Press, forthcoming.
- Hobbes, Thomas. *Leviathan*. 1651. In *Early Modern Philosophy: Essential Read-*

ence for helpful discussion. Thanks also to an anonymous referee for this journal, and to Tom Dougherty, whose work on consent has helped me to think through these issues.

- ings with Commentary*, edited by A. P. Martinich, Fritz Allhoff, and Anand Jayprakash Vaidya, 61–69. Hoboken, NJ: Blackwell, 2006.
- Hurd, Heidi M. “The Moral Magic of Consent.” *Legal Theory* 2, no. 2 (June 1996): 121–46.
- Jackson, Frank. “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection.” *Ethics* 101, no. 3 (April 1991): 461–82.
- King, Kristen. “Aziz Ansari Allegations Show that People Have a Lot to Learn about Consent.” *Teen Vogue*, January 16, 2018. <https://www.teenvogue.com/story/aziz-ansari-consent-op-ed>.
- Kleinig, John. “The Nature of Consent.” In Miller and Wertheimer, *The Ethics of Consent*, 3–24.
- Locke, John. *The Two Treatises of Civil Government*. 1689.
- MacKinnon, Catharine A. *Toward a Feminist Theory of the State*. Cambridge, MA: Harvard University Press, 1991.
- Manson, Neil C. “Permissive Consent: A Robust Reason-Changing Account.” *Philosophical Studies* 173, no. 12 (December 2016): 3317–34.
- Manson, Neil C., and Onora O’Neill. *Rethinking Informed Consent in Bioethics*. Cambridge: Cambridge University Press, 2007.
- Mill, John Stuart. *Utilitarianism*. Edited by Roger Crisp. Oxford: Oxford University Press, 1998.
- Miller, Franklin G., and Alan Wertheimer, eds. *The Ethics of Consent: Theory and Practice*. Oxford: Oxford University Press, 2010.
- . “Preface.” In Miller and Wertheimer, *The Ethics of Consent*, ix.
- Nagel, Thomas. “Moral Luck.” In *Moral Luck*, edited by Daniel Statman 141–66. Albany, NY: State University of New York Press, 1993.
- Owens, David. “The Possibility of Consent.” *Ratio* 24, no. 4 (December 2011): 402–21.
- Pateman, Carole. “Women and Consent.” *Political Theory* 8, no. 2 (May 1980): 149–68.
- Reilly, Natalie. “Aziz Ansari and the Issue of Consent.” *Sydney Morning Herald*, January 15, 2018. <https://www.smh.com.au/entertainment/celebrity/why-im-not-surprised-by-aziz-ansari-allegations-20180115-hoi92c.html>.
- Schulhofer, Stephen J. “The Feminist Challenge in Criminal Law.” *University of Pennsylvania Law Review* 143 (1995): 2151–2207.
- Sherwin, Emily. “Infelicitous Sex.” *Legal Theory* 2, no. 3 (September 1996): 209–31.
- Shiffrin, Seana Valentine. “Promising, Intimate Relationships, and Conventionalism.” *Philosophical Review* 117, no. 4 (October 2008): 481–524.

- Sosa, David. "Consequences of Consequentialism." *Mind* 102, no. 405 (January 1993): 101–22.
- Steward, Helen. *A Metaphysics for Freedom*. Oxford: Oxford University Press, 2012.
- Wallace, R. Jay. "Reasons, Relations, and Commands: Reflections on Darwall." *Ethics* 118, no. 1 (October 2007): 24–36.
- Wertheimer, Alan. *Consent to Sexual Relations*. Cambridge: Cambridge University Press, 2003.
- West, Robin. "Sex, Law, and Consent." In Miller and Wertheimer, *The Ethics of Consent*, 221–50.
- Westen, Peter. *The Logic of Consent: The Diversity and Deceptiveness of Consent as a Defense to Criminal Conduct*. Farnham, UK: Ashgate, 2004.
- Zimmerman, Michael J. "Is Moral Obligation Objective or Subjective?" *Utilitas* 18, no. 4 (December 2006): 329–61.

UTILITARIANISM, ALTRUISM, AND CONSENT

Christopher J. G. Meacham

ACCORDING to *objective act utilitarianism*, an act is morally permissible iff it maximizes total utility. Objective act utilitarianism (henceforth utilitarianism) yields appealing verdicts in a number of cases. But there are other cases in which its verdicts are less appealing. Consider the following case:

Parental Sacrifice: Your child's birthday is coming up. You have the option of working overtime in order to buy your child a gift. Doing so would significantly decrease your utility, a decrease greater than the utility your child would receive from the gift. No one else's utility would be affected.

Since working overtime would not maximize total utility, utilitarianism entails that it is impermissible to work overtime to buy your child the gift. But this strikes many as the wrong verdict. For while it is not obligatory to work overtime, it is surely permissible to do so.¹

The standard diagnosis of this problem is that utilitarianism provides implausible prescriptions in cases like Parental Sacrifice because it does not accommodate an important "self-other asymmetry." In particular, it does not accommodate the fact that it is permissible to disregard one's own utility when making decisions about how to act.

The standard remedy is to modify utilitarianism by adopting an altruistic version of what I will call a "dual-maximizing" theory. Dual-maximizing theories provide a particular way of allowing agents to (partially or wholly) disregard the utility of some group of subjects. And such views have been employed to address a number of objections to utilitarianism. In response to "nearest and dearest" objections to utilitarianism, some have suggested adopting dual-maximizing views that permit partially disregarding the utility of those one is not close to, and thus privileging the utility of one's friends and family. In response to "demandingness" objections to utilitarianism, some have suggested adopting dual-maximizing

¹ I have made your child the beneficiary of your sacrifice in order to make the case easy to imagine. But there is nothing important about the beneficiary being your child. If you wanted to make such a sacrifice to benefit a stranger, most would still feel it is permissible to do so.

views that permit partially disregarding the utility of others, and thus privileging one's own utility. And in response to "altruistic" objections to utilitarianism like Parental Sacrifice, some have suggested adopting dual-maximizing views that permit disregarding one's own utility, and thus privileging the utility of others.²

I will argue that both the standard diagnosis and the standard remedy are mistaken. I will argue that the standard diagnosis is mistaken because cases like Parental Sacrifice do not demonstrate that it is permissible to disregard one's own utility. Rather, what such cases show us is that it is permissible to disregard the utility of those who *consent*.

And I will argue that the standard remedy is mistaken because the way in which dual-maximizing theories permit disregarding the utility of groups of subjects is problematic. I will argue that if we want to modify utilitarianism to permit disregarding, we should do so in a different way, employing what I will call a "variable-disregarding" theory. And this is true not just in the context of "altruistic" objections like Parental Sacrifice, but in every case in which one might want to permit disregarding.

Note that these two theses are independent of one another. One can accept that it is permissible to disregard the utility of those who consent, and deny that the best way to permit disregarding is via a variable-disregarding theory. And one can accept that the best way to permit disregarding is via a variable-disregarding theory, and deny that it is permissible to disregard the utility of those who consent. However, these two theses are independently attractive, and together they entail that cases like Parental Sacrifice are best accommodated by adopting a variable-disregarding view that permits disregarding those who consent.

Two preliminary comments before we proceed. First, the issues I discuss in this paper are relevant to accounts of both *subjective obligation* (where an agent's beliefs mediate what they ought to do) and accounts of *objective obligation* (where they do not).³ But, for simplicity, I focus my attention here on theories of objective obligation.

Second, the issues discussed in this paper are not just relevant to utilitarian-

- 2 For a classic presentation of the "nearest and dearest" objection, see Smart and Williams, *Utilitarianism*. For a prominent discussion of "demandingness" objections, see Scheffler, *The Rejection of Consequentialism*. For an influential discussion of "altruistic" objections, see Slote, "Morality and Self-Other Asymmetry." For suggestions on how to use dual-maximizing views (or multi-maximizing views; cf. note 8) to address these objections, see Sider, "Asymmetry and Self-Sacrifice"; Portmore, "Dual-Ranking Act-Consequentialism"; and Vessel, "Supererogation for Utilitarianism."
- 3 For some recent discussions of these issues, see Zimmerman, "Is Moral Obligation Objective or Subjective?"; Sobel, "Subjectivism and Idealization"; and Graham, "In Defense of Objectivism about Moral Obligation."

ism, but to many theories that take utility into account.⁴ For example, consider a hybrid rights theory that takes rights to be side constraints, but otherwise tells you to maximize utility. In cases in which no one's rights are being threatened, this theory will yield the same prescriptions as utilitarianism. Thus one can raise "nearest and dearest," "demandingness," and "altruistic" worries for this kind of rights theory too. And it is natural to deal with these worries by allowing agents to disregard the utility of certain groups of subjects. In a similar vein, a typical person-affecting view will yield the same verdicts as utilitarianism in cases in which all of the same individuals exist regardless of what one does. Thus one can raise "nearest and dearest," "demandingness," and "altruistic" worries for these person-affecting views. And again it is natural to address these worries by allowing agents to disregard the utility of certain groups.

That said, it is easiest to see our way through these issues when our underlying theory is as simple and straightforward as possible. So in what follows I will focus my attention on how one might apply these kinds of modifications to utilitarianism.

The rest of this paper will go as follows. In section 1, I present dual-maximizing theories, show how one can use them to handle cases like Parental Sacrifice, then raise some worries for such theories. In section 2, I present variable-disregarding theories, show that one can use them to handle cases like Parental Sacrifice, then raise some worries for the standard diagnosis of these cases in terms of some kind of self-other asymmetry. In section 3, I make the case for thinking that the morally important issue that cases like Parental Sacrifice raises is not the distinction between self and other, but the distinction between giving and withholding consent. I then present a variable-disregarding theory that accommodates the role of consent, show how it handles all the cases discussed so far, and discuss the conditions required for morally relevant consent. In section 4, I consider several objections. In section 5, I briefly summarize these results.

1. DUAL-MAXIMIZING THEORIES

1.1. *The Structure of Dual-Maximizing Theories*

Utilitarianism says that we should treat all subjects equally. But various objections to utilitarianism, such as the "nearest and dearest" objections, "demandingness" objections, and "altruistic" objections, might make one reconsider this

4 For some recent examples of other views that one could modify in this way, see Roberts, *Child versus Childmaker*; Preda, "Rights Enforcement, Trade-Offs, and Pluralism"; and Meacham, "Person-Affecting Views and Saturating Counterpart Relations."

tenet. For while it is presumably permissible to treat everyone equally, one might also want to permit partially or wholly disregarding the utility of some subjects, whether it is other people, those one does not feel close to, or oneself.

If we want to modify utilitarianism in order to permit (but not require) disregarding some subjects, one popular option is to adopt a dual-maximizing theory.⁵ We can formulate dual-maximizing theories as follows, with respect to some condition ϕ that picks out the disregarded subjects, and some weight $w \in [0, 1]$ that corresponds to the degree to which disregarded subjects are taken into account (with $w = 0$ if we give the utility of disregarded subjects no weight, and $w = 1$ if we give the utility of disregarded subjects full weight).

Dual-Maximizing Theories: An act a is permissible iff either (i) a maximizes utility, or (ii) a maximizes the sum of (the utility of those who do not satisfy ϕ) plus (w times the utility of those who do satisfy ϕ).

We will spend most of our time focusing on cases in which $w = 0$, and the utility of disregarded subjects is completely discounted. In these cases, dual-maximizing theories consider two things: whether an act maximizes utility overall, and whether an act maximizes utility for the subjects we are not disregarding. If the answer to either question is yes, then the act is permissible. If the answer to both questions is no, then the act is impermissible.

As Portmore notes, dual-maximizing theories have several attractive features.⁶ As we saw above, we can use dual-maximizing theories to accommodate self-centered options, special consideration for one's friends and family, and self-sacrificing options. And we can employ dual-maximizing theories to make sense of a broad range of supererogatory acts, e.g., given a self-centered dual-maximizing theory, there will often be a number of permissible acts that are worse for the agent, and better for others, than other permissible options.⁷

- 5 If one wants to modify utilitarianism to *require* agents to disregard a group of subjects, then this is easy to do: one can either change one's characterization of utility to simply exclude the group in question, or (equivalently) adopt a theory that tells you to ignore the utility of such subjects when evaluating acts. Since the question of how to require disregarding is not contentious (unlike the question of how to *permit* disregarding), I will not address it in the text.
- 6 Strictly speaking, Portmore makes these claims regarding a broader class of theories that he calls "dual-ranking theories," which consist of any theory with a pair of conditions (i) and (ii) such that an act is permissible iff it satisfies one of these conditions ("Dual-Ranking Act-Consequentialism"). But all of the particular theories Portmore considers are either dual-maximizing theories or multi-maximizing theories (see note 8).
- 7 Of course, how to understand supererogation is controversial, and one might want to cash this notion in a number of different ways. And as Vessel notes, on this understanding of supererogation, even utilitarians will admit to some supererogatory acts (since there can be

Dual-maximizing theories are a special case of a more general kind of theory we might call “multi-maximizing theories,” which allow for any number of different disregarded groups with different weights. But to keep things simple, I will bracket these complications and focus on dual-maximizing theories in what follows.⁸

1.2. Self-Other Utilitarianism

Cases like Parental Sacrifice invoke the intuition that there is a distinction between how one takes one’s own utility into consideration and how one takes the utility of others into consideration. In particular, the intuition is that it is permissible to disregard one’s own utility when making decisions, but not permissible to disregard the utility of others. One way to modify utilitarianism in order to accommodate this intuition is to adopt a dual-maximizing theory, where the condition is being oneself, and the associated weight w is 0. Doing so yields the following theory, proposed by Sider:⁹

Self-Other Utilitarianism (SOU): An act a is permissible iff (i) a maximizes the utility of all subjects, or (ii) a maximizes the utility of others.

SOU allows agents to perform acts that decrease overall utility, as long as by doing so they do not decrease the utility of anyone but themselves. Thus this theory allows for permissible self-sacrifice.

SOU yields the desired prescriptions in Parental Sacrifice. Suppose we represent your options as follows (where the numbers indicate the utility of the relevant subject given that act):

pairs of acts that both maximize overall utility, one that is worse for the agent and better for others, and one that is better for the agent and worse for others) (“Supererogation for Utilitarianism”). Nevertheless, the point remains that there are natural ways of understanding supererogation according to which self-centered, dual-maximizing theories can recognize a much broader range of supererogatory acts than typical utilitarian theories can.

- 8 We can formulate multi-maximizing theories as follows, with respect to m conditions ϕ_i with associated weights w_i :

Multi-Maximizing Theories: An act a is permissible iff for some i ’s, a maximizes the sum of (the utility of those who do not satisfy ϕ_i) plus (w_i times the utility of those who do satisfy ϕ_i).

Multi-maximizing theories offer an appealing generalization of dual-maximizing theories, as they can simultaneously permit disregarding others, those one is not close to, and oneself. For formulations of multi-maximizing theories that incorporate multiple kinds of disregarding, see Portmore, “Dual-Ranking Act-Consequentialism”; and Vessel, “Supererogation for Utilitarianism.”

- 9 Sider, “Asymmetry and Self-Sacrifice.”

Options	Self	Child	Total
Do Not Work Overtime	10	10	20
Work Overtime	8	11	19

According to utilitarianism, it is impermissible for you to work overtime, since this option fails to maximize total utility. But according to SOU, both options are permissible. It is permissible for you not to work overtime because this option maximizes total utility. And it is permissible for you to work overtime because this option maximizes the utility of others.

1.3. Problems for Dual-Maximizing Theories

Although dual-maximizing theories like SOU are attractive in some respects, they have implausible consequences. To see this, let us consider a counterexample to SOU inspired by Splawn that is helpful in both identifying the problem with dual-maximizing theories and identifying how to remedy this defect.¹⁰

Consider a version of Parental Sacrifice in which you have the option of making both small and large sacrifices:

Variable Parental Sacrifice: Your child's birthday is coming up. You have the option of working overtime for a week in order to buy your child a gift, or for two weeks in order to buy your child an even better gift. Both options would significantly decrease your utility—in both cases, this decrease would be greater than the utility your child would receive from the gift. No one else's utility would be affected.

In this case we might represent your options as follows:

Options	Self	Child	Total
Do Not Work Overtime	10	10	20
Work One Week of Overtime	8	11	19
Work Two Weeks of Overtime	7	12	19

According to SOU, it is permissible not to work overtime, since this option maximizes total utility. Likewise, it is permissible to work two weeks of overtime, since this option maximizes the utility of others. But it is impermissible to work one week of overtime, since this option maximizes neither total utility nor the utility of others. This is an implausible result.¹¹ Given the range of options usu-

¹⁰ Splawn, "The Self-Other Asymmetry and Act Utilitarianism."

¹¹ Splawn offers a criticism of SOU along these lines ("The Self-Other Asymmetry and Act Utilitarianism"). Portmore offers a response, although he assesses the worry with respect

ally available to agents, it would follow that agents could only make total-utility-decreasing self-sacrifices of the most extreme kind. Modest gifts and sacrifices for loved ones would not be permissible. Only the most extreme and lavish gifts would be permitted.

This points to a general problem with dual-maximizing theories. The problem is that dual-maximizing theories are fundamentally disjunctive. They treat maximizing everyone's utility and maximizing the appropriately weighted utility of privileged and disregarded subjects as distinct goals, and take acts to be permissible iff they are the best at achieving either of these goals. But by treating these goals as distinct, they are unable to allow for natural trade-offs between them.

How might one modify dual-maximizing theories in order to allow for such trade-offs? Here is one natural thought. The difference between the first and second goals of *SOU* can be seen as a difference in the value of the weight w . The first goal—maximizing utility—is a case where $w = 1$, and the utility of disregarded subjects is not discounted, while the second goal—maximizing the utility of the privileged subjects—is a case where $w = 0$ and the utility of disregarded subjects is completely discounted. So if we want to allow for trade-offs between these two goals, one natural thought is to permit any act that maximizes weighted utility for any value of w between these two extremes.

Unfortunately, this natural thought does not pan out. On this proposal, the total weighted utility of each act will be:

Options	Self	Child	Weighted Total
Do Not Work Overtime	$10w$	10	$10 + 10w$
Work One Week of Overtime	$8w$	11	$11 + 8w$
Work Two Weeks of Overtime	$7w$	12	$12 + 7w$

to a kind of multi-maximizing theory he calls "Schefflerian utilitarianism" instead of *SOU*, where Schefflerian utilitarianism takes an act to be permissible iff it either (i) maximizes the utility of others, or (ii) maximizes the sum of (the utility of others) plus (ten times one's own utility) ("Dual-Ranking Act-Consequentialism"). Like *SOU*, Schefflerian utilitarianism has the implausible consequence that only the first and third options in Variable Parental Sacrifice are permissible. Portmore responds to this objection by arguing that working only one week of overtime *should* be impermissible, because you would be "unreasonably selfish" to sacrifice your utility in this inefficient way. If you work two weeks of overtime then others get two-thirds of the utility you sacrifice, whereas if you work one week of overtime then others get only one half of the utility you sacrifice. I think it is false that non-maximizing self-sacrifices are only permissible when they are maximally efficient. That stance would entail that if we could make a sacrifice to work a week of overtime to buy our child a present, but could also make a more efficient sacrifice by selling our leg in order to allow our child to go to college, then only selling our leg (or making no sacrifice at all) would be permissible. This is implausible.

But no value of w will make working one week of overtime maximize weighted utility. If $w = 1$, then not working overtime will yield a higher weighted utility (20) than working one week of overtime (19). And if $w < 1$, then working two weeks of overtime will yield a higher weighted utility than working one week of overtime. This is because working one week of overtime is just like working two weeks of overtime, except it shifts one unit of utility from your child to yourself. And since we are discounting the utility assigned to you (because $w < 1$), this shift will decrease weighted utility.

Why does working one week of overtime seem permissible? It seems permissible because it seems you should be free to determine how much of your utility you are willing to sacrifice. You can choose not to sacrifice any of your utility and not work overtime, you can choose to sacrifice a lot of utility and work two weeks of overtime, or you can choose to just sacrifice some of your utility—up to two utility, say—and work one week of overtime. The problem with the natural thought sketched above is that what we want to vary is not the *degree* to which you discount your utility, but the *amount* of utility that you are willing to discount. Let us turn to theories that allow us to do that.

2. VARIABLE-DISREGARDING THEORIES

2.1. *The Structure of Variable-Disregarding Theories*

In this section, I will introduce *variable-disregarding theories*. In the interest of accessibility, I will introduce them in two stages. In section 2.1.1, I will describe these theories informally. Then in section 2.1.2, I will describe these theories more formally, in a manner mirroring my presentation of dual-maximizing theories in section 1.1.

2.1.1. *Variable-Disregarding Theories, Take 1*

We want to modify utilitarianism to permit giving less weight to the utility of some group of subjects. And, as we saw in the last section, we also want our theory to consider various amounts of utility that we might discount. Let us look at one way to do this.

There are some acts that will bring about more utility for a group than any other act, i.e., will bring about the maximum utility for that group.¹² And there are other acts that will lead to drops in utility for this group, relative to this maximum. One way to give less weight to the utility of this group is to assign less

12 In certain infinity cases, there will be no act that maximizes utility; I am bracketing such complications here (cf. note 14).

importance to these drops—that is, to treat their *effective* utility as closer to their maximum than it really is. So suppose we are giving no weight to the utility of a subject. If their maximum utility is 10, and their utility given a certain act is 6, we can disregard this drop in their utility by treating their utility as 10 for the purposes of evaluating this act. Alternatively, suppose we are giving half weight to the utility of a subject. If their maximum utility is 10, and their utility given a certain act is 6, we can give half weight to this drop in their utility by treating their utility as 8 for the purposes of evaluating this act.

Now, we also want the theory to permit discounting various *amounts* of utility. Here is one way to do that. Consider different magnitudes of drops in utility. Some acts will maximize effective utility if we discount drops in a group's utility of up to some positive finite size. Some acts will maximize effective utility if we discount *all* drops in a group's utility (equivalently: will maximize effective utility if we discount all drops in a group's utility of up to size ∞). And some acts will maximize utility *simpliciter* (equivalently: will maximize effective utility if we discount all drops in a group's utility of up to size 0). We can ensure that our theory permits discounting various amounts of utility by taking all of these different acts to be permissible.

Putting these thoughts together, we can say that an act is permissible iff, for some amount of utility, discounting drops in a group's utility of up to that amount makes the act maximize effective utility.¹³ I will call a theory of permissible action of this form a *variable-disregarding theory*.

Let us work through an example. Suppose we want a variable-disregarding theory that allows you to give no weight to the utility of those you do not feel close to. Then we can take an act to be permissible iff there is some amount of utility of those you are not close to that you can disregard that makes that act maximize effective utility. So consider the following case:

Dividing Goods: You are considering whether to split some good evenly between a friend of yours and a stranger, or give all of the good to your friend. If you split the good evenly both will get a fair amount of utility, whereas if you give all of it to your friend, your friend's utility will be slightly higher, but the stranger's utility will be much lower.

We might represent your options in Dividing Goods as follows:

13 I am simplifying slightly here by talking in terms of a group's utility instead of the utility of individuals in a group; for a more precise presentation, see section 2.1.2.

Options	Friend	Stranger	Total
Split Evenly	9	9	18
Favor Friend	10	5	15

Which of these acts is permissible? Let us start with Split Evenly. Is there some amount of utility of the stranger that you can discount that makes Split Evenly maximize effective utility? Here the answer is trivially yes—if we do not discount *any* amount of utility, then Split Evenly will maximize utility. So Split Evenly is permissible.

Now consider Favor Friend. Is there some amount of utility of the stranger that you can discount that makes Favor Friend maximize effective utility? Well, suppose we discount an arbitrarily large amount of the stranger's utility. Favor Friend drops the utility of the stranger by 4, relative to Split Evenly. And if we give that drop in utility no weight—we give it none of the importance we normally would—then we can treat the stranger's effective utility as 9. So the total effective utility of Favor Friend will be 19. By contrast, disregarding drops in the stranger's utility will not change the effective utility of Split Evenly, since in Split Evenly there is no drop in the stranger's utility to disregard. Thus the effective utility of Split Evenly will still be 18. Since Favor Friend has a higher effective utility, it follows that Favor Friend is also permissible.

Options	Friend	Stranger	Effective Total
Split Evenly	9	9	18
Favor Friend	10	5 (9)	19

Of course, we did not have to discount an *arbitrarily large* amount of the stranger's utility to get this result. Discounting drops in the stranger's utility of up to size 4 would yield the same result.

Now consider a different theory. Suppose we wanted a theory that only allows you to *partially* disregard—say, give half weight to—the utility of those you are not close to. What acts would be permissible in Dividing Goods?

As before, Split Evenly will be permissible, since if we do not disregard *any* utility, Split Evenly will maximize effective utility. What about Favor Friend? Is there some amount of utility of the stranger you can partially disregard (give half weight to) according to which Favor Friend maximizes effective utility?

Well, suppose we discount an arbitrarily large amount of the stranger's utility. Favor Friend decreases the stranger's utility by 4. And if we give that drop in utility half weight—we give it half of the importance we normally would—then we can treat the stranger's effective utility as 7. So the total effective utility of Favor

Friend will be 17, while the total effective utility of Split Evenly will remain 18. So Favor Friend will not maximize effective utility in this case either.

Options	Friend	Stranger	Effective Total
Split Evenly	9	9	18
Favor Friend	10	5 (7)	17

More generally, a little thought reveals that there is no amount of utility we can discount that will make Favor Friend maximize effective utility. So if we are required to give the stranger's utility at least half weight, giving all of the good to your friend is impermissible.

2.1.2. Variable-Disregarding Theories, Take 2

Now let us formulate variable-disregarding theories more precisely. Let us begin by introducing some terminology. As in section 1.1, let ϕ be the condition that picks out the subjects we are disregarding, and let $w \in [0, 1]$ be the weight assigned to the utility of the disregarded subjects. Let $u_a(s)$ be the utility of subject s given act a . And let MAX_s be the maximum utility that subject s could end up with in a given decision problem.¹⁴ (Thus in Dividing Goods, $\text{MAX}_{\text{friend}}$ is 10, and $\text{MAX}_{\text{stranger}}$ is 9.)

Now let us introduce the *effective utility* $_{(\phi, w, \nu)}(a)$ of an act a , where ν is a sequence of values $\nu = \{\nu_{s1}, \nu_{s2}, \dots\}$ assigned to the ϕ subjects that corresponds to the amount of their utility we are discounting. The effective utility $_{(\phi, w, \nu)}(a)$ of an act a is the sum for each subject s of (1) $\text{MAX}_s - w(\text{MAX}_s - u_a(s))$, if (given a) s satisfies condition ϕ and s 's utility is within ν_s of MAX_s , or (2) $u_a(s)$, otherwise.

So when $w = 0$, the effective utility $_{(\phi, 0, \nu)}$ of an act is just the sum of the utilities of each subject, with the following exception: if a subject satisfies ϕ , and has a utility within ν_s of their maximum, then we treat their utility as that maximum. So the effective utility $_{(\phi, 0, \nu)}$ of an act ignores drops in utility of up to size ν_s for each subject s who satisfies ϕ .

When $w > 0$, these drops in utility are not ignored, just given less weight. So when $w = 0.5$, the effective utility $_{(\phi, 0.5, \nu)}$ of an act is the sum of the utilities of each subject, except that subjects who satisfy ϕ and have a utility within ν_s of their maximum are treated as having a utility that is only half as far from their maximum as it actually is, e.g., if their utility is 6 and their maximum is 10, they

¹⁴ In certain infinity cases there might not be a maximum utility a subject could have (e.g., a case in which you can choose any natural number n , and receive that much utility). Since such cases pose problems for both standard-maximizing theories and variable-disregarding theories, I will ignore these complications here.

are treated as having a utility of 8. Thus the effective utility $_{(\phi,0.5,v)}$ of an act gives half weight to drops in utility of up to size v_s for each subject s who satisfies ϕ .

With this terminology in hand, we can formulate variable-disregarding theories as follows, with respect to some condition ϕ that picks out the disregarded subjects, and some weight $w \in [0, 1]$ that corresponds to the degree to which disregarded subjects are taken into account:

Variable-Disregarding Theories: An act a is permissible iff, for some assignment v of values to subjects who satisfy ϕ , a maximizes effective utility $_{(\phi,w,v)}$.

So when $w = 0$, variable-disregarding theories tell us that a is permissible iff there is some assignment of values v to the ϕ subjects such that disregarding drops in utility of up to those amounts makes a maximize effective utility. And when $w > 0$, variable-disregarding theories tell us that a is permissible iff there is some assignment of values v to ϕ subjects such that partially disregarding (to a degree determined by w) drops in utility of up to those amounts makes a maximize effective utility.

Let me pause for a moment to say a bit more about the roles of ϕ , w , and v . The first two are parameters that are fixed by a variable-disregarding theory. In spelling out what variable-disregarding theory we are using, we have to specify a particular condition ϕ and a particular value w . By contrast, the v 's—the amounts we are disregarding—are not fixed by a variable-disregarding theory. For any given variable-disregarding theory, we consider *all* of the different possible v 's—all of the different possible amounts of utility we might disregard. And we take an act to be permissible iff, for at least one of these v 's, that act maximizes effective utility.

To get a feel for this terminology, let us work through the Dividing Goods case again. Suppose we adopt a variable-disregarding theory that allows you to disregard the utility of those you do not feel close to. So ϕ is the condition of being someone you do not feel close to; call this condition nc . And let us suppose that the theory allows you to completely disregard the utility of those you are not close to, so that $w = 0$. Recall that your options in the Dividing Goods case are as follows:

Options	Friend	Stranger	Total
Split Evenly	9	9	18
Favor Friend	10	5	15

If we set $v_{\text{stranger}} = 0$, and so do not disregard any drops in the stranger's utility, then Split Evenly maximizes effective utility $_{(nc,0,0)}$:

Options	Friend	Stranger	Effective Total ($v = 0$)
Split Evenly	9	9	18
Favor Friend	10	5	15

If we set $v_{\text{stranger}} = \infty$, and so disregard drops in the stranger’s utility of any size, then Favor Friend maximizes effective utility $_{(nc,0,\infty)}$:

Options	Friend	Stranger	Effective Total ($v = \infty$)
Split Evenly	9	9	18
Favor Friend	10	5 (9)	19

Since each act maximizes effective utility for some value of v , this theory takes both acts to be permissible.

Suppose instead we adopt a more plausible theory on which $w = 0.5$, so that we can only *partially* disregard drops in utility for those we are not close to. If we set $v_{\text{stranger}} = \infty$, then we will give half weight to drops in the stranger’s utility (because $w = 0.5$), and we will do so for drops in utility of any size (because $v_{\text{stranger}} = \infty$). Thus Split Evenly will maximize effective utility $_{(nc,0.5,\infty)}$:

Options	Friend	Stranger	Effective Total ($v = \infty$)
Split Evenly	9	9	18
Favor Friend	10	5 (7)	17

More generally, one can show that Split Evenly will maximize effective utility for any value of v . Thus if $w = 0.5$, we will get the result that only Split Evenly is permissible.

In what follows, we will mostly focus on cases in which $w = 0$ and the utility of disregarded subjects is completely discounted. To simplify notation, I will leave the w index implicit when it is clear that $w = 0$.

Variable-disregarding theories share all of the advantages of dual-maximizing theories. We can use them to permit self-centered options, options that favor one’s friends and family, and options that favor others. And variable-disregarding theories also allow for a broad range of supererogatory acts.

Just as dual-maximizing theories are a special case of multi-maximizing theories, variable-disregarding theories are a special case of what we might call “multivariable-disregarding theories,” which allow for any number of different disregarded groups with different weights. But for simplicity, I will focus on variable-disregarding theories in what follows.¹⁵

15 Let ϕ_1, \dots, ϕ_m be a sequence of conditions, and let $v_i = \{v_i^{s1}, v_i^{s2}, \dots\}$ be a sequence of values assigned to the ϕ_i subjects. Let the *effective utility* $_{(\phi_1, \dots, \phi_m, w_1, \dots, w_m, v_1, \dots, v_m)}(a)$ of an act a

2.2. *Self-Discounting Utilitarianism*

Suppose we want to handle Parental Sacrifice by modifying utilitarianism to allow us to disregard ourselves. As we have seen, we can do this by adopting a dual-maximizing theory, SOU. But we can also do this by adopting a variable-disregarding theory. Taking the condition ϕ to be being oneself (call this condition s), and taking the associated weight to be $w = 0$, yields the following theory:

Self-Discounting Utilitarianism (SDU): An act a is permissible iff, for some value ν , a maximizes effective utility $_{(s,\nu)}$.

So SDU tells us that a is permissible iff there is some value ν such that disregarding drops in one’s own utility of up to that amount makes a maximize utility.

SDU yields the desired prescriptions in Parental Sacrifice. If we set $\nu = 0$, and so do not disregard any drops in our utility, then not working overtime maximizes effective utility $_{(s,0)}$:

Options	Self	Child	Effective Total ($\nu = 0$)
Do Not Work Overtime	10	10	20
Work Overtime	8	11	19

If we set $\nu = \infty$, and so disregard drops of any amount to our utility, then working overtime maximizes effective utility $_{(s,\infty)}$:

Options	Self	Child	Effective Total ($\nu = \infty$)
Do Not Work Overtime	10	10	20
Work Overtime	8 (10)	11	21

Thus both options are permissible.

Unlike SOU, SDU also yields the desired prescriptions in Variable Parental Sacrifice. If we set $\nu = 0$, and so do not disregard drops of any amount to our utility, then not working overtime maximizes effective utility $_{(s,0)}$:

in a given decision problem be the sum over each subject s of (1) $\text{MAX}_s - w_i(\text{MAX}_s - u_a(s))$, if (given a) for some i , s satisfies ϕ_i and s ’s utility is within ν^i of MAX_s , or (2) $u_s(a)$, otherwise. We can formulate multivariable-disregarding theories, with respect to m conditions ϕ_i with associated weights w_i , as follows:

Multivariable-Disregarding Theories: An act a is permissible iff, for some m assignments ν_1, \dots, ν_m of values to subjects, a maximizes effective utility $_{(\phi_1, \dots, \phi_m, w_1, \dots, w_m, \nu_1, \dots, \nu_m)}$.

Options	Self	Child	Effective Total ($\nu = 0$)
Do Not Work Overtime	10	10	20
Work One Week of Overtime	8	11	19
Work Two Weeks of Overtime	7	12	19

If we set $\nu = 2$, and so disregard drops of up to 2 units to our utility, then working one week of overtime maximizes effective utility $_{(s,2)}$:

Options	Self	Child	Effective Total ($\nu = 2$)
Do Not Work Overtime	10	10	20
Work One Week of Overtime	8 (10)	11	21
Work Two Weeks of Overtime	7	12	19

If we set $\nu = \infty$, and so disregard drops of any amount to our utility, then working two weeks of overtime maximizes effective utility $_{(s,\infty)}$:

Options	Self	Child	Effective Total ($\nu = \infty$)
Do Not Work Overtime	10	10	20
Work One Week of Overtime	8 (10)	11	21
Work Two Weeks of Overtime	7 (10)	12	22

So SDU yields the desired result that all three options are permissible.

2.3. Further Problems

SOU yields plausible verdicts in Parental Sacrifice, and SDU yields plausible verdicts in both Parental Sacrifice and Variable Parental Sacrifice. But both SOU and SDU yield strange results in cases where other subjects are willing to make sacrifices as well, such as the following:

Two Parental Sacrifices: Your child's birthday is coming up. You have the option of working overtime in order to buy your child a gift. Doing so would significantly decrease the utility of both you and your partner—the decrease for each of you would be greater than the utility your child would receive from the gift. But your partner is willing to make the sacrifice, and no one else's utility would be affected.

In this case we might represent your options as follows:

Options	Self	Partner, Child	Total
Do Not Work Overtime	10	10, 10	30
Work Overtime	8	8, 11	27

According to SOU it will be impermissible to work overtime, since doing so maximizes neither total utility nor the utility of everyone else. Likewise, according to SDU it will be impermissible to work overtime, since regardless of how much of your utility we disregard, not working overtime will maximize effective utility $_{(s,\nu)}$. If $\nu < 2$, then the effective utility $_{(s,\nu)}$ of not working overtime will be 30 instead of 27:

Options	Self	Partner, Child	Effective Total ($\nu < 2$)
Do Not Work Overtime	10	10, 10	30
Work Overtime	8	8, 11	27

And if $\nu \geq 2$, then the effective utility $_{(s,\nu)}$ of not working overtime will be 30 instead of 29:

Options	Self	Partner, Child	Effective Total ($\nu \geq 2$)
Do Not Work Overtime	10	10, 10	30
Work Overtime	8 (10)	8, 11	29

So no matter what we set ν to, not working overtime maximizes effective utility. Thus SDU takes not working overtime to be obligatory.

These are implausible verdicts, especially when combined with the verdicts that SOU and SDU make in Parental Sacrifice. If it is permissible for you to willingly sacrifice some of your well-being for your child, then it should be permissible for your partner to do so as well. The fact that in Two Parental Sacrifices your partner cannot themselves act in order to make this sacrifice does not seem morally relevant.

Moreover, if we accept SOU or SDU, we have to accept an odd conflict between our permissions and those of others. To see this, return to the Parental Sacrifice case. When it is the parent making the decision, SOU and SDU both yield the verdict that it is permissible for the parent to work overtime. But suppose that working overtime also requires the parent's manager to sign off. Then it will be the manager making the decision. And since the manager is not one of the parties involved, SOU and SDU will both yield the same verdicts as standard utilitarianism—namely, that it is impermissible for the parent to work overtime. So while SOU and SDU allow the parent to make this self-sacrifice, it will not allow the manager to let this self-sacrifice take place.

This yields an uncomfortable tension between our permissions and those of others. While SOU and SDU will permit you to make self-sacrifices, they will also require everyone else to prevent you from doing so. This has the consequence that it would be virtually impossible to carry out these self-sacrifices in a society where everyone did the right thing. This is because most of our potential sacrifices require the acquiescence of others to go through, and others would be forbidden from allowing these sacrifices to take place. Banking agents would be obligated not to allow the relevant checks to clear, store owners would be obligated not to sell the relevant products to the relevant people, bosses would be obligated to refuse to allow the relevant people to work overtime, and so on. This would be a strange state of affairs.

Let us take a step back. Two Parental Sacrifices seems similar in morally relevant respects to Parental Sacrifice and Variable Parental Sacrifice, and it seems the explanation of why it is permissible to work overtime in each case should be the same. But while a self-other asymmetry could potentially explain the permissibility of working overtime in Parental Sacrifice and Variable Parental Sacrifice, it cannot explain the permissibility of working overtime in Two Parental Sacrifices. For in Two Parental Sacrifices, working overtime does not just lower your utility (which a self-other asymmetry could allow you to ignore), it also lowers the net utility of others. Thus we have reason to be skeptical that a self-other asymmetry is what is at the heart of these cases.

More generally, this suggests that the various authors who have taken utilitarianism to have a special problem with “altruistic” objections have missed the forest for the trees. They have identified a particular batch of problem cases, and have taken those problem cases to encapsulate the whole problem. What cases like Two Paternal Sacrifices show is that these “altruistic” objections are really just an instance of a broader interpersonal phenomenon. And in order to address this broader phenomenon, we need to do more than just permit agents to disregard their own utility.

3. CONSENT

3.1. *Consent-Discounting Utilitarianism*

While discussing these “altruistic” objections, Slote raises the following thought:

It has been suggested to me that the reason we are allowed to harm ourselves or avoid some benefit, where we should not be permitted to harm another person or prevent her from receiving a similar benefit, lies in the consent implicit in actions we do to ourselves. If I harm myself to avoid a

benefit, I presumably do this willingly, whereas the agent whom I refuse to benefit does not consent to this neglect (and when she does there is nothing wrong with what I do). It might be then thought that the moral asymmetry we have noted is not a deep feature of morality, but rather derivative from and justifiable in terms of the moral importance of consent.¹⁶

Although Slote goes on to reject this suggestion, I think this is exactly right. The key moral distinction in these cases is not the distinction between self and others, it is the distinction between those who consent and those who do not.

If this thought is correct, then the modification of utilitarianism suggested in section 2.2 is too cautious. Instead of modifying standard utilitarianism to permit disregarding oneself, we should modify it to permit disregarding those who consent. That is, we should change the condition ϕ picking out who we can disregard from “oneself” to “those who consent to the act.” Then we can explain the appearance of a self-other asymmetry in such cases as deriving from the asymmetry in what subjects typically consent to.¹⁷

In some of the literature, the term “consent” is used to mean something like waiving a right, or releasing someone from a duty to you.¹⁸ That is, consent is used to mean something that is by definition morally significant. I am using the term “consent” in its more colloquial sense, to denote something like agreement or acquiescence. And while one might plausibly take agreement or acquiescence to have moral significance, we can use these notions without presupposing that they have moral significance.

It is plausible that consent (so understood) must satisfy certain conditions in order to be morally relevant. For example, we might want to require the consenting subject to be informed, competent, free from coercion, and so on, in order for their consent to count. For now, I will simply take for granted that there are some conditions of this kind, and I will call consent that satisfies these conditions “morally relevant consent,” or “consent_m.” We will return to examine the question of what consent_m is in section 3.2.

Now let us return to the task at hand: formulating a theory that permits disregarding those who consent. We can construct a variable-disregarding theory that

16 See Slote, “Morality and Self-Other Asymmetry,” 190–91.

17 The suggestion is that *this* apparent asymmetry can be explained in terms of consent. It is not that *all* self-other asymmetries can be explained in terms of consent. For example, if one chooses to deal with “demandingness” objections by allowing agents to partially disregard the utility of others, then one is also positing a kind of self-other asymmetry. But this self-other asymmetry cannot be explained by consent.

18 E.g., see Liberto, “Intention and Sexual Consent.”

does this by taking the condition ϕ to be consent_m to the act in question (call this condition c), and taking the associated weight to be $w = 0$:

Consent-Discounting Utilitarianism (CDU): An act a is permissible iff, for some assignment of values v to subjects who satisfy c , a maximizes effective utility $_{c,v}$.

So CDU tells us that a is permissible iff there is some assignment of values v to subjects such that disregarding drops in utility of up to those amounts for subjects who consent_m to a makes a maximize utility.¹⁹

Now let us see how CDU handles the three cases we have discussed. In all of these cases you consent_m to whichever action you perform, and for concreteness I will assume your child also consents_m , though we get the same results regardless of whether your child consents_m . Let us start with Parental Sacrifice. If we set $v = \{0, 0\}$ (i.e., set $v_{\text{self}} = 0$ and $v_{\text{child}} = 0$), then not working overtime maximizes effective utility $_{(c,\{0,0\})}$:

Options	Self	Child	Effective Total ($v = 0, 0$)
Do Not Work Overtime	10	10	20
Work Overtime	8	11	19

And if we set $v = \{\infty, 0\}$, then working overtime will maximize effective utility $_{(c,\{\infty,0\})}$:

Options	Self	Child	Effective Total ($v = \infty, 0$)
Do Not Work Overtime	10	10	20
Work Overtime	8 (10)	11	21

So CDU will take both options to be permissible.

Now consider Variable Parental Sacrifice. If we set $v = \{0, 0\}$ (i.e., set $v_{\text{self}} = 0$ and $v_{\text{child}} = 0$), then not working overtime maximizes effective utility $_{(c,\{0,0\})}$:

19 Note that although I have described CDU as disregarding *drops* in utility of those who consent, it also effectively disregards *increases* in utility of those who do not consent to the increase. For example, consider a subject who does not consent to the option a that maximizes their utility, and instead prefers an option b that brings them a lower utility. Since the subject consents to b , they will be treated as if they had the utility they would have given a for the purposes of evaluating b , and thus the increase in utility that a brings them will not end up telling in favor of a .

Options	Self	Child	Effective Total ($\nu = 0, 0$)
Do Not Work Overtime	10	10	20
Work One Week of Overtime	8	11	19
Work Two Weeks of Overtime	7	12	19

If we set $\nu = \{2, 0\}$, then working one week of overtime maximizes effective utility $_{(C,\{2,0\})}$:

Options	Self	Child	Effective Total ($\nu = 2, 0$)
Do Not Work Overtime	10	10	20
Work One Week of Overtime	8 (10)	11	21
Work Two Weeks of Overtime	7	12	19

And if we set $\nu = \{\infty, 0\}$, then working two weeks of overtime maximizes effective utility $_{(C,\{\infty,0\})}$:

Options	Self	Child	Effective Total ($\nu = \infty, 0$)
Do Not Work Overtime	10	10	20
Work One Week of Overtime	8 (10)	11	21
Work Two Weeks of Overtime	7 (10)	12	22

So CDU will take all three options to be permissible.

Now consider Two Parental Sacrifices. In this case there is a third subject, your partner, who will also consent $_m$ to whatever act you perform. If we set $\nu = \{0, 0, 0\}$ (i.e., set $\nu_{\text{self}} = 0$, $\nu_{\text{partner}} = 0$, and $\nu_{\text{child}} = 0$), then not working overtime maximizes effective utility $_{(C,\{0,0,0\})}$:

Options	Self	Partner, Child	Effective Total ($\nu = 0, 0, 0$)
Do Not Work Overtime	10	10, 10	30
Work Overtime	8	8, 11	27

And if we set $\nu = \{\infty, \infty, 0\}$, then working overtime maximizes effective utility $_{(C,\{\infty,\infty,0\})}$:

Options	Self	Partner, Child	Effective Total ($\nu = \infty, \infty, 0$)
Do Not Work Overtime	10	10, 10	30
Work Overtime	8 (10)	8 (10), 11	31

So CDU will take both options to be permissible.

3.2. Morally Relevant Consent

3.2.1. Informed, Competent, and Uncoerced Consent

Now let us consider what kinds of conditions we want an account of consent_m to take into consideration.²⁰

First, we want consent_m to require that the subject is informed. If a subject would consent to having their money taken, but only because they have been deceived into thinking that they would get it back, we should not take them to be consenting_m. Likewise, if a subject would consent to giving up their salary bonus, but only because they falsely believe that if they do it will go to charity, we should not take them to be consenting_m.²¹

Second, we want consent_m to require that the subject is competent. If a subject would consent to being lit on fire, but only because they are mentally ill, we should not take them to be consenting_m. Likewise, if a subject would consent to having all of their property taken, but only because they are very young or on powerful drugs, we should not take them to be consenting_m.²²

Adding this competence clause raises a potential worry. Suppose one identifies competence with rationality. On some conceptions of rationality, a rational agent would never consent to a decrease in utility, and never dissent to an increase in utility. If so, and we require competence for consent_m, then no one will ever consent_m to acts that decrease their utility. But then CDU becomes equivalent to standard utilitarianism, since the deviations from maximizing utility that CDU allows will never come into play.

This worry requires an understanding of competence that equates “competent” with “ideally prudentially rational,” or something of that kind. This strong understanding of the competence condition is not what I have in mind. The notion of competence I have in mind here is something weaker, something like the

20 In addition to the informed, competent, and uncoerced conditions described below, some have suggested to me that one might add a fourth condition that requires the act to line up with the subject’s desires or deep commitments in order for them to consent_m. (For a discussion of a view along these lines, see Killoren, “Relationship-Sensitive Utilitarianism About Animals.”) This clause would rule out consent that was merely made because the subject was trying to be nice, or felt obligated to consent, even though they did not really want to. This is an interesting suggestion, and I take a version of CDU incorporating such a clause to be a viable option.

21 For some discussions of how to spell out the notion of informed consent, see Manson and O’Neill, *Rethinking Informed Consent in Bioethics*; Beauchamp and Childress, *Principles of Biomedical Ethics*; and Eyal, “Informed Consent.”

22 For a discussion of some of the issues that arise with respect to assessing competence, see Grisso and Appelbaum, *Assessing Competence to Consent to Treatment*; Beauchamp and Childress, *Principles of Biomedical Ethics*; and Charland, “Decision-Making Capacity.”

kind of competence required in order for one's actions to be subject to moral evaluation. For example, normal human adults might be thought to be subject to such evaluation, while dogs are not. Thus one might take normal human adults, but not dogs, to be competent.²³ And if we assess CDU using this weaker notion of competence, this worry will not arise.

Third, we want consent_m to require that the subject is uncoerced. Suppose you live under the rule of a sadistic dictator who can scan your brain to detect whether you have consented to various things. And suppose the dictator will scan your brain and react as follows: if you consent to their cutting off your hand then they will do so, and if you do not consent to their cutting off your hand then they will kill you. If you know all of this then you might reasonably consent to them cutting off your hand. But we should not take you to be consenting_m to their cutting off your hand.²⁴

3.2.2. *Actual versus Ideal Consent*

We want consent_m to be informed, competent, and uncoerced. But there are a couple of different ways in which one might incorporate these kinds of conditions into an account of consent_m. First, one might take a subject to consent_m to *a* iff they consent to *a* and satisfy the relevant conditions. We might call this the *actual* way of incorporating these conditions, since it requires the subject to actually satisfy these conditions in order to count as consenting_m. Second, one might take a subject to consent_m to *a* iff the following counterfactual is true: if the agent were to perform this act, and if *s* were informed, competent, and uncoerced, then *s* would consent to the act.²⁵ We might call this the *ideal* way

- 23 Granted, this notion of competence seems to come in degrees. Likewise, the extent to which a subject's actions are subject to moral evaluation seems to come in degrees (see Charland, "Decision-Making Capacity"). If one wanted to take this into consideration, one might consider modifying CDU to take *degrees* of competent consent into account. Alternatively, one might simply take the strength of deontic obligations to be something that admits of variable degree—a highly competent agent can do something *WRONG*, while a barely competent agent can only do something (*sotto voce*) wrong. These are interesting issues, but not ones I will take up here. For simplicity, I will simply proceed on the assumption that a binary notion of competence will suffice.
- 24 Although I am not taking a stand here on how to understand coercion, those looking for a more concrete account could do worse than assuming something like Nozick's account ("Coercion").
- 25 There are further important details here regarding how we evaluate this counterfactual, of course. In particular, we want to ensure that when evaluating this counterfactual we prioritize holding fixed facts regarding the subject's utility (so that their utility given each act remains unchanged) and other facts regarding their personality that bear on whether they would consent to the act. Pete Graham has suggested that we might also consider a different notion of ideal consent_m linked to a counterfactual that does not have the performance of

of incorporating these conditions, since it only requires an idealized version of the subject who satisfies these conditions to consent in order for the subject to count as consenting_m.

I take both the actual and ideal approaches to have intuitive appeal.²⁶ And in some cases our intuitions regarding these approaches conflict. For example, suppose a member of an isolated tribe with no experience of modern medicine is bleeding to death. And suppose a medic has found the subject and must insert a needle into their arm in order to give them the blood transfusion needed to save their life. In light of the subject's lack of familiarity with modern medicine, and their lack of understanding of what the needle insertion is for, they do not consent to having the needle inserted into their arm. But suppose that if the subject were informed of what the needle was for, and understood that this was necessary to save their life, they would consent.

Now, in this case should we treat the subject as consenting_m to have the needle inserted into their arm? That is, with respect to the morally relevant notion of consent, should we treat them as consenting? I think most people would feel torn about this case. On the one hand, the tribe member does not actually consent to having the needle inserted into their arm, and that seems morally important. On the other hand, they clearly would consent if they were appropriately informed, and that seems morally important as well.

I think both ways of understanding consent_m are viable. So in what follows, I will leave it open which notion of consent_m we are working with. And in the few places where our choice between actual and ideal consent_m makes a difference, I will note how these two approaches diverge.

4. OBJECTIONS

4.1. *Objections to Consent-Discounting Utilitarianism*

Now let us turn to consider some potential objections to CDU. One objection to CDU is that in cases where $w = 0$, it will permit choosing dominated acts—i.e., permit choosing an act even though there is some other act available that is strictly better for disregarded subjects and just as good for privileged ones. And one might object that this is implausible.

the act in the antecedent. This alternative notion of ideal consent would largely yield the same results, though it would yield a different response to Slote's first argument discussed in section 4.3 (cf. note 37).

²⁶ For a discussion of some of the merits and demerits of appealing to something like hypothetical or ideal consent, see Thomson, *The Realm of Rights*; Stark, "Hypothetical Consent and Justification"; and Enoch, "Hypothetical Consent and the Value(s) of Autonomy."

CDU will have this consequence, but this is a consequence that we should accept. To see this, consider a version of Parental Sacrifice in which your child would end up being equally happy with or without the gift. (While your child would greatly enjoy the gift, if you did not buy it then they would end up finding something else to enjoy.) In this case, I take it that it is still permissible for you to work overtime, even though the option of working overtime is dominated by the option of not working overtime. And CDU yields this verdict, i.e., if we set $v = \infty$, then both options maximize effective utility $_{(c,\infty)}$:

Options	Self	Child	Effective Total ($v = \infty$)
Do Not Work Overtime	10	10	20
Work Overtime	8 (10)	10	20

And so CDU will take both options to be permissible, as desired.

Likewise, consider a version of Two Parental Sacrifices in which your child would end up being equally happy with or without the gift. Again, I take it to be permissible for you to work overtime if both you and your partner are willing to make the sacrifice, even though working overtime is dominated by not working overtime. And again, CDU yields this verdict, i.e., if we set $v = \infty$, then both options maximize effective utility $_{(c,\infty)}$:

Options	Self	Partner, Child	Effective Total ($v = \infty, \infty, 0$)
Do Not Work Overtime	10	10, 10	30
Work Overtime	8 (10)	8 (10), 10	30

As before, CDU will yield the desired verdict that both options are permissible.²⁷

27 Here is another dominance-violating case that one might take to be more damning (thanks to Pete Graham). Suppose someone is drowning, and you can either rescue them (at no cost to yourself) or let them drown. If the person consents to your letting them drown, CDU would hold that it is permissible to do so. But (the objection goes) surely that is not right. Here is my response to this objection. When we fill in the details regarding this case, we naturally imagine the subject as someone who is severely depressed or is suffering from a mental illness of some kind. But such a subject would not be competent, and so could not consent $_m$ to your letting them drown. Thus CDU would not say it is permissible to let them drown. Could we fill in the details of this case in a way that makes it plausible that the subject is informed and competent? I find this very hard to do. But I find that the more I do to make it plausible that a subject could provide informed and competent consent to your letting them drown, the more plausible it becomes that it is permissible to let them drown. In this respect this case is similar to the brainwashing case discussed below (see especially note 31).

CDU's treatment of these dominance cases entails that it will sometimes violate the kind of "bang for your buck" principle defended by Graham, which (roughly) requires you to "make the most" out of losses inflicted on others ("The Bang for Your Buck Principle"). Thus

Here is a second objection to CDU. Consider the following variant of Variable Parental Sacrifice:

Options	Self	Child	Total
Do Not Work Overtime	10	10	20
Work One Week of Overtime	8	11	19
Work Two Weeks of Overtime	7.9	11.9	19.8

CDU will still take working one week of overtime to be permissible, since working one week of overtime maximizes effective utility $_{(c,\{2,0\})}$. But one might worry that in this case it should be impermissible to work one week of overtime, since working two weeks of overtime will only cost you an additional 0.1 utility while granting your child an additional 0.9 utility. I take the intuition behind this worry to be this: working one week of overtime is impermissible because, given the other options available, it is an inefficient sacrifice. If you work one week of overtime you sacrifice 2 utility to give 1 utility for your child, while if you work two weeks of overtime you sacrifice 2.1 utility to give 1.9 utility for you child. And one might hold that if one is going to make a utility-decreasing sacrifice, then one is obligated to do so in as efficient a manner as possible.

This worry is similar to one I addressed in section 1.3 (note 11). But since it has been repeatedly raised in conversation as a worry for CDU, I will address it again here. I think this worry is not compelling because I think it is false that permissible utility-decreasing sacrifices must be maximally efficient. That would entail that if we could work a week of overtime to buy our child a present, or could make a more efficient sacrifice by selling our leg in order to allow our child to go to college, then only selling our leg (or making no sacrifice at all) would be permissible. That is implausible.

I think the right stance to take here is the one I described at the end of section 1.3. Namely, you should be free to determine how much of your utility you are willing to sacrifice when performing utility-decreasing altruistic acts. You can choose not to sacrifice any of your utility and not work overtime, you can choose to sacrifice a lot of utility and work two weeks of overtime, or you can choose to sacrifice no more than 2 utility and work one week of overtime.

Here is a third objection to CDU. Consider an evil dictator who has brainwashed everyone else, so that they now consent to anything they might do. Then

if we consider a case where I can choose between lowering your utility by 1 to increase my utility by 1, lowering your utility by 1 to increase my utility by 2, or doing nothing, the first option would be impermissible according to the bang for your buck principle. But if you consent to all three options, then CDU will entail that the first option is permissible.

it seems like CDU would make it permissible for the dictator to torture everyone, since they all consent to this decrease in utility. This seems like the wrong result—it should not be permissible for a dictator to do whatever they want just because they have brainwashed everyone. Of course, everyone would not consent to being brainwashed in the first place, so presumably the act of brainwashing itself would not be permissible. So the evil dictator would still be doing something wrong at some point in the process. But one might think that, past events aside, the dictator is also doing something wrong *now* by torturing everyone.

How the proponent of CDU will reply to this worry depends on the notion of “brainwashing” in play. On the one hand, everyone may be “brainwashed” in the sense that they only consent to (say) being tortured by the dictator because they are not informed, competent, or uncoerced. If they were informed, competent, and uncoerced, they would not consent to being tortured. On this understanding of the case, the populace consents to being tortured but does not consent_m to being tortured. Thus torturing everyone would be impermissible.²⁸

On the other hand, one might construct a scenario in which everyone is “brainwashed” to consent in a manner that leaves them informed, competent, and uncoerced. In that case, everyone would consent_m to being tortured. This case is hard to envision—it is difficult to think of a scenario in which a populace full of informed, competent, and uncoerced subjects would still consent to being tortured. But if they consent despite being informed, competent, and uncoerced, then it sounds like they are not so much brainwashed as convinced.²⁹ And if this is how we understand the case, then it is not clear it is wrong for the dictator to torture everyone after all.³⁰

28 I take it that this is the way brainwashing is usually understood, since brainwashing is generally taken to undermine competence.

29 On this understanding, we are employing the term “brainwashed” to mean something like “convinced of something that radically alters your beliefs or consenting attitudes” or “convinced to adopt beliefs or consenting attitudes that you (initially) strongly disagree with.” This does seem to be one way in which we use the term “brainwashed.” Those who are convinced to enter a cult, for example, are often called brainwashed, even though in at least some cases it seems like those who enter the cult are competent, relatively informed, and so on.

30 It is important here not to confuse the plausibility of there being a case in which everyone would consent to being tortured despite being informed, competent, and uncoerced, with the plausibility of its being permissible for the dictator to torture them in such a case. To appropriately evaluate the latter question, we need to first come up with a case where it seems plausible to us that everyone would consent to being tortured despite being informed, competent, and uncoerced. It is only when we have managed to flesh out such a case that we can intuitively evaluate whether it is plausible in such a case for the dictator to permissibly torture them. And in my experience, the more we do to make it plausible that everyone

4.2. Objections to Variable-Disregarding Theories

Now let us consider an objection to variable-disregarding theories more generally, as a means of incorporating permissible disregarding into a utilitarian framework.

The first objection to CDU described in section 4.1 concerned the fact that it is permissible to choose an option that is dominated by another option that is strictly better for disregarded subjects and just as good for privileged ones. I argued that this is a desirable feature of CDU—if a subject consents_m to an option that leaves them with a lower utility, then *ceteris paribus* that option is permissible.

But one might worry that while this reply is plausible in some cases of disregarding, such as those involving consent, it is not plausible in others. For example, if we want to modify utilitarianism to permit disregarding those whom you are not close to, we surely do not want to permit acts that do nothing to benefit your friends and family, but make other people worse off! That is, suppose we set up a choice between acts like the following:

Options	Others	Friends and Family	Total
a_1	10	10	20
a_2	9	10	19

Surely a_2 should be impermissible. If that is right, then one might worry that while variable-disregarding theories may correctly model how to permit some kinds of disregarding (e.g., oneself, those who consent), they do not correctly model how to permit other kinds of disregarding (e.g., those one is not close to, others).

I agree that in the above case a_2 is impermissible. But this is not an objection to variable-disregarding theories, it is an objection to assigning disregarded subjects no weight in such cases. What distinguishes the two kinds of disregarding just described (disregarding oneself or those who consent, versus disregarding those one is not close to or others) are what weight assignments w to disregarded subjects are plausible. In the first kind of disregarding—disregarding oneself or those who consent—it is plausible to permit completely discounting the utility of the disregarded subjects and taking $w = 0$. In the second kind of disregarding—disregarding those one is not close to or others—it is plausible that the utility of disregarded subjects must be given at least some weight and $w > 0$. For while morality might allow one to *partially* disregard those one is not close to or others, it surely does not allow us to *completely* disregard those one is not close to or oth-

would consent to being tortured despite being informed, competent, and uncoerced, the more plausible it becomes that it is permissible for the dictator to torture them.

ers. And as long as we give some weight to the utility of the disregarded subjects, then the kinds of dominated options described above will not be permissible.³¹

4.3. *Slote's Objections*

I have suggested that one can accommodate the intuitions behind these “altruistic” objections by modifying utilitarianism in order to accommodate the role of consent. But Slote has argued that considerations involving consent cannot accommodate these intuitions.³² And the proponents of theories like *SOU*, such as Sider and Portmore, have endorsed these arguments for why appealing to consent will not work. So let us turn to assess these arguments.³³

Slote offers two arguments for why consent will not suffice to explain these intuitions. He presents the first argument in the following passage:

If someone irrationally asks me to harm or kill him, it will presumably be irrational and wrong of me to kill him, more wrong at any rate than if I irrationally choose to kill myself; yet the consent seems equal in the two cases.³⁴

The argument seems to be this. If someone irrationally consented to my killing them, it would be wrong for me to kill them. But if I irrationally want to kill myself (and so consent to doing so), it would not be wrong (or not as wrong) for me to kill myself. So even though we both (irrationally) consent to an act that kills us, and so the status of the act with respect to consent is the same, the moral status of these acts is still different. So consent cannot explain this moral difference.

By “irrational” I take Slote to mean irrational to a degree that undermines competence. For if we understand Slote to be talking about a notion of irrationality that does not undermine competence—e.g., merely failing to maximize the subject’s expected utility—then the claims he makes are not plausible.³⁵

31 For example, consider a variable-disregarding view on which the condition ϕ is not being close to the agent (call this condition nc), and suppose the weight assigned to nc subjects is $w = 0.5$. And consider the case described above. If $v = \{0, 0\}$ then the disregarding utilities $(nc, 0.5, \{0, 0\})$ will be 20 for a_1 and 19 for a_2 . If $v = \{\infty, \infty\}$ then the disregarding utilities $(nc, 0.5, \{\infty, \infty\})$ will be 20 for a_1 and 19.5 for a_2 . But for all v , a_1 will have a higher effective utility than a_2 . Thus a_2 will be impermissible, as desired.

32 Slote, “Morality and Self-Other Asymmetry.”

33 Sider, “Asymmetry and Self-Sacrifice”; Portmore, “Dual-Ranking Act-Consequentialism.”

34 Slote, “Morality and Self-Other Asymmetry,” 191.

35 For example, consider a terminal cancer patient facing an extremely painful death. At times close enough to their death, it will be rational for them to choose death, since the expected utility of continuing to live will be negative. At times far enough from their death, it will be irrational to choose death, since the expected utility of continuing to live will be positive. Now

Now let us see whether Slote's argument is compelling. First, let us assume an ideal notion of consent_m. Suppose that in both cases killing the person would lower their utility, and that no one else's utility would be affected. What will CDU say about my killing someone who incompetently consents to my killing them? To work this out, we need to assess whether, if they were competent, they would have consented to my killing them. I take it that in Slote's example the subject only consents because they are not competent—if they were competent, then they would not consent. If so, then CDU would say that it is impermissible for me to kill them, just as Slote says.

What will CDU say about the act of incompetently killing myself? Again, there is the question of whether, if I were competent, I would consent to my decision to commit suicide. Here the answer is trivially yes—if I were competent and decided to kill myself, I would consent to killing myself. So CDU will take it to be permissible for me to kill myself, just as Slote says. Thus a consent-based account can explain the moral difference between these two acts.³⁶

Suppose instead that we adopt an actual, instead of ideal, notion of consent_m. Given the way I have suggested we understand the notion of competence in sec-

consider a time just before the expected utility of death flips from positive to negative. And suppose the terminal cancer patient competently requests euthanasia at this time. This request will count as irrational, but it does not seem wrong to carry out their request.

Taking a step back, it is worth highlighting the strangeness of Slote's claim if we understand his talk of irrationality to just mean something like failing to maximize expected utility. We are trying to work out an account of *objective* obligation, tracking the actual consequences of our acts. Given this, it seems bizarre to maintain that the *rationality* of one's request—a subjective notion mediated by one's *beliefs* about the consequences—determines the wrongness of one's acts. To drive this home, consider two terminal cancer patients who are in identical situations, facing identical futures, who both request euthanasia. Suppose both patients are, according to the usual standards, informed and competent—perhaps they are both doctors and have a reasonably good grasp of what their future experiences will be like. And suppose that one patient is slightly more optimistic than the other, so that while living and dying have the same *expected* utility for the first patient, living has a slightly higher expected utility than dying for the second patient. According to Slote (so understood), while it could be objectively permissible to carry out the first patient's request for euthanasia, it would be objectively impermissible to carry out the second patient's request. And this is so even though they are in the exact same situation and facing the exact same futures.

36 That said, I will confess that my intuitions here are equivocal. If you only kill yourself because you are not competent, I find (*contra* Slote) that the verdict that your act was wrong to be as plausible as the verdict that your act was permissible. Given this, it is interesting to note that if we adopt the alternative notion of ideal consent_m suggested by Pete Graham (cf. note 25), CDU will hold that killing yourself in these circumstances would be impermissible after all. That is because, given this alternative notion, you will not ideally consent_m to your suicidal act, and so your utility will not be disregarded. Thus the act will be judged according to the usual utilitarian calculus, and the act of killing yourself will be judged impermissible.

tion 3.2.1, it follows that if I am not acting competently, then my act is not subject to moral evaluation. So if I am not competent and I kill myself, I cannot be acting wrongly, any more than a dog who killed itself could be acting wrongly. On the other hand, if someone else is not competent, and I kill them, I certainly can be acting wrongly! Again, we find that a consent-based account can explain the moral difference between these two acts.

So Slote is incorrect to think that consent-based views cannot yield the desired verdicts in this case. The moral difference he raises can straightforwardly be cashed out in terms of consent.

Let us turn to Slote's second argument. He presents his second argument in the following passage:

If I can avoid either an enduring pain to myself or a short-lived pain to you, you and I might both agree that it would be foolish of me to prevent the shorter pain to you; judging the matter objectively, you might not consent to my taking the longer pain upon myself in order to save you from the shorter pain. Yet there would be nothing morally wrong ... in such a sacrifice. But when the positions are reversed and I can avoid a short-lived pain to myself or a longer-lived one to you and it is morally right that I should do the latter, you will presumably not consent to my doing the former and it will be wrong if I do so. Again, consent or lack of consent seems not to make the relevant common-sense moral difference.³⁷

The argument seems to be this. If you do not consent to my performing an act a_1 , which would impose a large decrease in my utility to prevent a small decrease in your utility, it would still be permissible for me to do a_1 . But if you do not consent to my performing an act b_1 , which would impose a large decrease in your utility to prevent a small decrease in my utility, it would not be permissible for me to do b_1 . Since you do not consent to either, consent cannot explain the difference in moral status of these two acts.

Let us see whether this argument is compelling. Assume that these acts will have no effect on the utility of anyone else. And assume that both I and the other person in question are informed, competent, and uncoerced. What will CDU say about my performing act a_1 , which brings about a large decrease in my utility to prevent a small decrease in your utility, versus act a_2 , in which I do nothing? From the description of the case, we can assume that I consent_m to whichever act I perform, but the other person only consents_m to my performing a_2 . If we set $v = \{\infty, \infty\}$ (i.e., set $v_{\text{self}} = \infty$ and $v_{\text{other}} = \infty$) this yields the result that both acts maximize effective utility_(C, {∞, ∞}):

37 Slote, "Morality and Self-Other Asymmetry," 191.

Options	Self	Other	Effective Total ($v = \infty, \infty$)
a_1	5 (10)	11	21
a_2	10	10 (11)	21

Thus CDU will take a_1 to be permissible.

What will CDU say about my performing act b_1 , which brings about a large decrease in your utility to prevent a small decrease in my utility, versus act b_2 , in which I do nothing? From the description of the case, we can assume that I consent_m to whichever act I perform, but the other person only consents_m to my performing b_2 . And inverting the utilities above gives us the following values:

Options	Self	Other	Total
b_1	11	5	16
b_2	10	10	20

With a little thought we can see that b_2 will maximize effective utility given any assignment v . For the disregarding utilities for b_2 can range from 20 to 21, while the disregarding utilities for b_1 will always be 16. So CDU will take b_1 to be impermissible.

So again, Slote is incorrect to think that consent-based views cannot yield the desired verdicts in this case. As before, the difference he raises can be straightforwardly cashed out in terms of consent.³⁸

5. CONCLUSION

One common complaint about utilitarianism is that it does not allow agents to disregard anyone. For example, as altruistic objections like Parental Sacrifice show, it does not allow agents to disregard their own well-being. In response to these complaints, a number of authors have endorsed replacing utilitarianism with dual-maximizing (or multi-maximizing) views, and have endorsed handling altruistic objections by adopting something like Sider's Self-Other Utilitarianism.³⁹

I have argued that these suggestions are off track in two respects. First, I

38 Portmore ("Dual-Ranking Act-Consequentialism") cites Gert (*Brute Rationality*) as another source of arguments for why consent cannot accommodate the intuitions motivating these "altruistic" objections. Gert's arguments have a slightly different flavor, since, unlike Slote, Gert's arguments have to do with subjective (not objective) obligation. But once we shift to (say) evaluating CDU with respect to consenting increases and dissenting decreases in *expected* utility, Gert's arguments take the same form as Slote's first argument, and can be handled in the same way.

39 Sider, "Asymmetry and Self-Sacrifice."

have argued that if we want to modify utilitarianism to permit disregarding, we should adopt variable-disregarding views, not dual-maximizing views. Second, I have argued that, upon closer examination, the cases that motivate positing this kind of self-other asymmetry reveal that the morally relevant distinction is not between self and other, it is between those who do and do not consent. And by combining these two insights—adopting a variable-disregarding view that disregards those who consent—we end up with an attractive theory that yields the desired verdicts in a broad range of cases.⁴⁰

University of Massachusetts Amherst
cmeacham@philos.umass.edu

REFERENCES

- Beauchamp, Tom L., and James F. Childress. *Principles of Biomedical Ethics*. Oxford: Oxford University Press, 2008.
- Charland, Louis. “Decision-Making Capacity.” *Stanford Encyclopedia of Philosophy* (Summer 2011). <https://plato.stanford.edu/archives/sum2011/entries/decision-capacity/>.
- Enoch, David. “Hypothetical Consent and the Value(s) of Autonomy.” *Ethics* 128, no. 1 (October 2017): 6–36.
- Eyal, Nir. “Informed Consent.” *Stanford Encyclopedia of Philosophy* (Fall 2012). <https://plato.stanford.edu/archives/fall2012/entries/informed-consent/>.
- Gert, Joshua. *Brute Rationality: Normativity and Human Action*. Cambridge: Cambridge University Press, 2004.
- Graham, Peter. “The Bang for Your Buck Principle.” Unpublished manuscript.
- . “In Defense of Objectivism about Moral Obligation.” *Ethics* 121, no. 1 (October 2010): 88–115.
- Grisso, Thomas, and Paul S. Appelbaum. *Assessing Competence to Consent to Treatment: A Guide for Physicians and Other Health Professionals*. Oxford: Oxford University Press, 1998.
- Killoren, David. “Relationship-Sensitive Utilitarianism About Animals.” Unpublished manuscript.

⁴⁰ I would like to thank Eric Chwang, Maya Eddon, Linda Eggert, Fred Feldman, Brad Hooker, David Killoren, Hallie Liberto, Brad Skow, the audiences of the 2019 Formal Ethics conference and 2019 ROME conference, and an anonymous referee for helpful comments and discussion. I owe a special thanks to Pete Graham for detailed comments and nearly a decade of discussions on these issues.

- Liberto, Hallie. "Intention and Sexual Consent." *Philosophical Explorations* 20, supp. 2 (2017): 127–41.
- Manson, Neil C., and Onora O'Neill. *Rethinking Informed Consent in Bioethics*. Cambridge: Cambridge University Press, 2007.
- Meacham, Christopher J. G. "Person-Affecting Views and Saturating Counterpart Relations." *Philosophical Studies* 158, no. 2 (2012): 257–87.
- Nozick, Robert. "Coercion." In *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*, edited by Sidney Morgenbesser, Patrick Suppes, and Morton White, 440–72. New York: St Martin's Press, 1969.
- Portmore, Douglas W. "Dual-Ranking Act-Consequentialism." *Philosophical Studies* 138, no. 3 (2008): 409–27.
- Preda, Adina. "Rights Enforcement, Trade-Offs, and Pluralism." *Res Publica* 17, no. 3 (2011): 227–43.
- Roberts, Melinda. *Child versus Childmaker: Future Persons and Present Duties in Ethics and the Law*. Lanham, MD: Rowman & Littlefield, 1998.
- Scheffler, Samuel. *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations underlying Rival Moral Conceptions*. Oxford: Oxford University Press, 1994.
- Sider, Theodore. "Asymmetry and Self-Sacrifice." *Philosophical Studies* 70, no. 2 (May 1993): 117–32.
- Slote, Michael. "Morality and Self-Other Asymmetry." *Journal of Philosophy* 81, no. 4 (April 1984): 179–92.
- Smart, J. J. C., and Bernard Williams. *Utilitarianism: For and Against*. Cambridge: Cambridge University Press, 1973.
- Sobel, David. "Subjectivism and Idealization." *Ethics* 119, no. 2 (January 2009): 336–52.
- Splawn, Clay. "The Self-Other Asymmetry and Act Utilitarianism." *Utilitas* 13, no. 3 (November 2001): 323–33.
- Stark, Cynthia A. "Hypothetical Consent and Justification." *Journal of Philosophy* 97, no. 6 (June 2000): 313–34.
- Thomson, Judith Jarvis. *The Realm of Rights*. Cambridge, MA: Harvard University Press, 1990.
- Vessel, Jean-Paul. "Supererogation for Utilitarianism." *American Philosophical Quarterly* 47, no. 4 (October 2010) 299–319.
- Zimmerman, Michael J. "Is Moral Obligation Objective or Subjective?" *Utilitas* 18, no. 4 (December 2006): 329–61.

FORGIVENESS AND THE SIGNIFICANCE OF WRONGS

Stefan Riedener

SUPPOSE THAT, since a recent thunderbolt, the oak by Lila's house is frail. Hearing of an upcoming storm, she asks her neighbor Jack if he could help her cut it. She tells him that the tree is a hazard, that one cannot fell it alone, and that she will hire a saw in a distant city for the logging. Jack promises to be there to help. However, though he knows he is forgetful, he does not take measures to remind himself. And when the day comes, he fails to think of his promise, is nowhere to be found, and Lila has to return that saw without having cut a branch. What would it mean for Lila to *forgive* Jack for his promise breaking? A number of different accounts have been proposed.¹ But a common idea is that Lila forgives Jack if she first feels a negative reactive attitude toward him for having broken his promise, and then overcomes this attitude in a certain way.² I will call this the *standard account* of forgiveness.

I think this account is promising. But it raises a number of questions. A first range of questions concerns the *nature* of forgiveness—or how exactly Lila must overcome her negative attitudes in forgiving. If she tricked herself into believing that Jack was not blameworthy in the first place, and therefore no longer has any negative attitudes toward him, she has not really forgiven. So there is a question about when an overcoming of such attitudes amounts to forgiveness.³ A second range of questions concerns the *norms* of forgiveness—or when exactly Lila is

- 1 For instance, some people hold that forgiveness is the exercise of a normative power (e.g., Warmke, "The Normative Significance of Forgiveness"; and Bennett, "The Alteration Thesis"), or the deliberate refusal to punish (e.g., Zaibert, "The Paradox of Forgiveness"), or a plurality of things with no clearly delineable essence (e.g., Neblett, "Forgiveness and Ideals").
- 2 The (contested) *locus classicus* of this is Butler, *Fifteen Sermons*. See also, e.g., Murphy and Hampton, *Forgiveness and Mercy*; Richards, "Forgiveness"; Holmgren, "Forgiveness and the Intrinsic Value of Persons"; Hughes, "What Is Involved in Forgiving?"; and Darwall, *The Second-Person Standpoint*, 72 (to name just a few authors in this tradition).
- 3 For a particularly clear statement of this problem, see, e.g., Hieronymi, "Articulating an Uncompromising Forgiveness."

warranted to forgive, or perhaps would not be warranted not to. If the oak destroyed her house in that storm and yet Jack never showed any remorse, it would seem problematically condoning for Lila to forgive him straight away.⁴ On the other hand, if Jack has honestly apologized, organized the logging himself, and become a model of trustworthiness in the twenty-five years since then, it would seem problematically unforgiving for Lila to still bear her grudge.⁵ So there is a question about when the relevant overcoming of such attitudes is warranted. And these two questions are related. We can only really say when forgiveness is warranted once we know what it is. And we can determine what forgiveness is by reflecting on how overcoming such attitudes can be warranted.

My aim in this paper is to begin to answer these questions about the standard account. There are many proposals about how to do so. But my sense is that they often cannot explain how forgiveness may be warranted, or do so in a manner that distorts what it is. To illustrate what I mean, I will outline the challenge of accounting for the norms of forgiveness in section 1. In section 2, I will introduce one of the most prominent versions of the standard account, due to Lucy Allais.⁶ If my arguments are sound, her account of why forgiveness can be warranted is dubious, and her picture of its nature both too narrow and too broad. But Allais will provide us with an important insight. So I will then introduce a novel version of the standard account by elaborating on what I take from her. My core idea is that the reactive attitudes are fundamentally agent relative. They are a fitting response not just to someone's blameworthiness, but to their blameworthiness being *significant* for you, or worthy of your *caring*, in virtue of your relationship to it. Someone's blameworthiness is significant for you to the extent that you are bound up with what grounds it—e.g., with the wrongdoer's being a participant in human relationships, or the victim's being a source of demands. Thus you may fittingly not care about someone's blameworthiness if it is sufficiently insignificant for you—e.g., if they lived in a distant time and place. And forgiveness revolves around this. Lila (fittingly) forgives Jack for his fault if and only if, partly out of goodwill toward him, she (fittingly) does not care about his blameworthiness anymore. To explain and support these ideas, I will elaborate in sections 3, 4, and 5 on the reactive attitudes in general. In section 6, I will relate this to forgiveness, and outline the proposal I have just sketched.

Two clarifications before I begin. First, I will focus on *resentment* as the paradigm attitude at stake. Actually, I think other attitudes—e.g., forms of disap-

4 For a forceful defense of this, see, e.g., Anderson, "When Justice and Forgiveness Come Apart."

5 See, e.g., Watson, "Standing in Judgment."

6 Allais, "Wiping the Slate Clean."

pointment or contempt—may also play a role in forgiveness.⁷ And I think the story I will tell about resentment could be applied to these attitudes too. But for simplicity, I will focus just on resentment. Second, I will focus on *overcoming* as the paradigm process at issue. Some people hold that, in order to forgive, rather than overcoming your resentment in the right manner, you need to *forswear* it, or *endorse* your lack of resentment no matter why you do not feel it.⁸ I think there is an important sense in which you have not forgiven me if you have forsworn but not yet overcome your resentment, and forgiven me if you have overcome your resentment (in the right manner) but have not forsworn it, and similarly, an important sense in which you have not forgiven me if you actually do not resent me out of sheer contempt but endorse that lack of resentment for the right reasons, and forgiven me if you have overcome your resentment (in the right manner) but do not yet endorse your lack of it. But not too much hinges on this. The alternative versions of the account face similar questions, about whether any forswearing or endorsing amount to forgiveness and when such stances are warranted. So much of what I will say could be employed for these alternative accounts as well. But I will focus on overcoming.

1. THE PROBLEM

Let us start with the question about the norms of forgiveness. Suppose that, in breaking his promise, Jack responsibly evinced a wrongful attitude toward Lila. She has the standing to blame him for it—e.g., is not notoriously unreliable herself. And she knows this. If this was not so, it would in different ways be problematic for Lila to feel resentment in the first place. But, given these assumptions, there is a question about how she may be warranted in overcoming it. Resentment is an emotion, so the precise nature of the question will depend on the nature and norms of emotions. There is a large debate about this, which I cannot enter here. But here are three standard assumptions that I will accept for this paper. First, emotions have standards of “fittingness.” A certain fear of a dangerous snake is fitting—whereas fear of a totally harmless mouse, or excessive panic about a marginally dangerous dog, or a naively mild respect vis-à-

7 On disappointment, see, e.g., Richards, “Forgiveness,” 78; Smith, “Moral Blame and Moral Protest,” 38; Blustein, *Forgiveness and Remembrance*; and Fricker, “What’s the Point of Blame?” 172. On contempt, see, e.g., Pettigrove, *Forgiveness and Love*, 29.

8 On forswearing, see, e.g., Strawson, “Freedom and Resentment,” 6; and Fricker, “Forgiveness.” On endorsing, see Schönherr, “When Forgiveness Comes Easy.” I will come back to one issue that Schönherr raises in note 48. Still other formulations include “forbearing or withdrawing” (Darwall, *The Second-Person Standpoint*, 72) or “letting go of” (Griswold, *Forgiveness*, 40) resentment.

vis an extremely dangerous moose, are not. Second, emotions have fittingness conditions in virtue of being representational: (intense/mild) fear represents its object as (very/moderately) dangerous, and is fitting if and only if this representation is correct. So if some proposition p is among the fittingness conditions of an emotion, that is because p is part of this emotion's representational content.⁹ Third, fittingness is normative. That is, if the representational content of some emotion is p , then p is a reason for you to have that emotion, and not- p is a reason for you not to have it.¹⁰ We might say it is a "fittingness reason" (not) to have the relevant emotion.

These assumptions are rough. There is a question, say, about the precise nature of emotional representation. Perhaps to have an emotion is to make a sort of judgment, or to (quasi-)perceive the world as being a certain way, or to have an intentional feeling toward it.¹¹ There is also a question about why fittingness matters. Perhaps it is simply *constitutive* of our adopting specific emotions that we take their fittingness to provide reasons.¹² Or perhaps there are (in addition) bridge principles linking fittingness to other sources of normativity: perhaps you have a *moral* reason to have fitting emotions about instantiations of *moral* value properties (like moral wrongness, unfairness, or blameworthiness), or it is part of the good life to generally have fitting emotions, or objectively good.¹³ And there is a question about whether there are other reasons for emotions—"non-fittingness reasons," or reasons of the "wrong kind." Suppose someone offers Lila \$1,000 if, or tortures a cat unless, she resents Jack for his promise breaking. Perhaps this gives her a non-fittingness reason to resent him.¹⁴ Or perhaps it does not give her a reason to resent him, but only a reason to want to resent him, or to take actions to bring it about that she will.¹⁵ These questions

9 For such an account of fittingness, see, e.g., D'Arms and Jacobson, "The Moralistic Fallacy"; Tappolet, "Values and Emotions"; and Rosen, "The Alethic Conception of Moral Responsibility." For a dissenting voice on the idea that emotions are representational, see, e.g., Hutto, "Truly Enactive Emotion"; also Deonna and Teroni, "Emotions as Attitudes." For a survey of different accounts of fittingness, see Howard, "Fittingness."

10 See Howard, "Fittingness," 3; for a dissenting voice, see, e.g., Maguire, "There Are No Reasons for Affective Attitudes."

11 On emotions as judgments, see, e.g., Nussbaum, *Upheavals of Thought*; and Solomon, "Emotions and Choice." On quasi-perceptions of the world, see, e.g., Prinz, *Gut Reactions*; and Tappolet, *Emotions, Values, and Agency*. On intentional feeling, see, e.g., Goldie, *The Emotions*.

12 See, e.g., Sharadin, "Reasons Wrong and Right."

13 For a theory along the lines of the final idea, see, e.g., Hurka, *Virtue, Vice, and Value*, ch. 1.

14 See, e.g., Rabinowicz and Rønnow-Rasmussen, "The Strike of the Demon."

15 See, e.g., Kelly, "The Rationality of Belief and Some Other Propositional Attitudes"; Parfit,

need not concern us. In what follows, I will only be concerned with fittingness reasons, without specifying the way in which emotions are representational, or why correct representation matters.¹⁶ For even these rough standard assumptions raise a puzzle about forgiveness.

What is the representational content of resentment? A common answer seems to be that if you feel resentment toward me for an action, you represent me as thereby having manifested an attitude toward you that was morally wrong and for which I was responsible; and the stronger your resentment, the more seriously wrong you represent my attitude as having been, or the more fully responsible you represent me as having been for it.¹⁷ So on this understanding, you can misrepresent the facts by feeling resentment where there was no fault (or responsible wrong) at all, or by feeling an excessively strong form in the face of what was (in terms of wrongness and responsibility) a relatively harmless fault, or a trivializingly mild annoyance in the face of what was (in these terms) an outrageous one. But this raises a problem. It suggests that in feeling less and less resentment toward Jack or eventually overcoming it altogether, Lila would represent his attitudes as less seriously wrong or him as less fully responsible—or eventually as not being at fault at all. After all, on the present understanding, these are the only dimensions along which the content of resentment can vary. Yet Jack *was* at fault. So, on these assumptions, Lila's overcoming resentment seems to render her emotions unfitting. But we want an account on which Lila can forgive Jack fittingly, or without suggesting he was not at fault.

But perhaps that is too quick. We might try to resolve this by considering what it means that an emotion is fitting. If emotions have fittingness conditions in virtue of being representational, then that an emotion is fitting does not mean you *must* feel it in order not to misrepresent the facts. Suppose Angeline does not know of Jack's promise breaking in the first place, and therefore does not feel any indignation about it. Then she does not *mis*represent Jack's fault. She just does not represent it at all. Similarly, Lila need not come to misrepresent Jack's fault

On What Matters, ch. 2 and app. A; and Way, "Transmission and the Wrong Kind of Reason." For a helpful survey of the discussion about "right kinds of reasons," see Gertken and Kiese-wetter, "The Right and the Wrong Kind of Reasons."

- 16 For what it is worth, I find it plausible that there are no wrong kinds of reasons for feeling an emotion. And I find all of the suggestions about why fittingness matters plausible. The idea that it is morally good to have or problematic to lack fitting emotions about moral value facts, in particular, can explain why forgiving too readily or not readily enough might itself be morally blameworthy—a form of being condoning or unforgiving that is morally objectionable.
- 17 This is suggested, e.g., by Strawson, "Freedom and Resentment," 23. In Strawson's framework I am "responsible" for an attitude roughly just if it is *mine* (and I am a normal adult). I do not mean to suggest anything else with talk of "responsibility" throughout this paper.

in overcoming her resentment, but may just come to not represent it at all. And that is what she would do in forgiving, or so one might hold. But as an account of forgiveness, this is not convincing either. To begin with, it seems false that if Lila has forgiven she will lack any emotions about Jack's fault. In coming to forgive, she will typically feel various (continuously weaker) forms of resentment. And once she has forgiven, she might still feel sad that he wronged her, or be positively at peace with his fault, or even be grateful if it rendered her stronger. Indeed, it is unclear whether Lila *can* avoid any emotional representation of it. There is a difference between having *no* attitude toward something—e.g., simply not knowing about it—and feeling indifferent about it or reconciled with it. And at least as long as she knows of Jack's fault, it seems she must have *some* attitude or other toward it, be it insouciance, propitiation, acquiescence, or whatever. But in any case, most importantly, even if we could somehow describe her as not representing Jack's fault, this would not give us the picture we want. On the present assumptions, resentment would still be the one fitting response to Jack's negligence. So Lila could come to have an emotional state that is *not unfitting*, at best. But she would do so by *failing* to have the fitting response. She would manage to forgive only through a sort of emotional forgetting. Yet we arguably want an account on which Lila can have the (or at least *a*) positive fitting response to the fact that Jack wronged her.

We could also try to resolve our problem with a more complex account of resentment. For instance, we might suggest that if you feel resentment toward me for an action, you represent me as thereby having manifested an attitude toward you that was wrong and for which I was responsible *and* for which I still have not felt any remorse. This would explain how Lila may fittingly overcome her resentment, at least once Jack has felt remorse. It would raise the substantive question of whether remorse really *is* necessary for forgiveness. Personally I think it is not, and that Lila may sometimes forgive Jack even if he has not properly repented. But set this aside for the moment.¹⁸ The present idea has a more fundamental problem. It implies that once Jack has felt remorse, any further *resentment* would be unfitting. More generally, it now seems that once it is fitting for Lila to no longer resent Jack, it would *eo ipso* be unfitting for her to still resent him. But this too is unfortunate. It seems that Lila could, sometimes, fittingly not feel resentment

18 Kolnai, "Forgiveness." For the view that remorse or a "change of heart" is crucial, or that "unconditional forgiveness" is inappropriate, see, e.g., Novitz, "Forgiveness and Self-Respect," esp. 314; Griswold, *Forgiveness*, esp. 121–22; Couto, "Reactive Attitudes, Forgiveness, and the Second-Person Standpoint"; and Milam, "Reasons to Forgive." For the view that it need not be, see, e.g., Calhoun, "Changing One's Heart"; Holmgren, "Forgiveness and the Intrinsic Value of Persons"; Garrard and McNaughton, "In Defence of Unconditional Forgiveness"; Ware, "Forgiveness and Respect for Persons"; and Fricker, "Forgiveness."

even when she could also fittingly feel it. That is why forgiveness can be a “gift”: it is not something she must always either offer or refuse in order to respond fittingly to Jack’s fault. So we want an account on which Lila can have *discretion* in fitting forgiveness—on which she can, at least sometimes, fittingly forgive Jack but also fittingly refrain from forgiving.¹⁹

In sum, there is a question about why to overcome resentment is not to misrepresent the facts, or to simply omit any representation of them, or the *only* way to represent them correctly. How can we sometimes fittingly overcome our resentment without taking refuge to emotional forgetting and without it being unfitting to continue to feel it? Forgiveness seems either unwarranted or required or at best an emotional void. Echoing Aurel Kolnai, we might call this the (apparent) “paradox of forgiveness.”²⁰

2. ALLAIS’S PROPOSAL

Before I attempt to solve this problem, let us consider one of the most prominent developments of the standard account of forgiveness, due to Lucy Allais.²¹ If I am right, Allais’s proposal is not successful. But it can teach us important lessons. So it will help us to start here. The core of Allais’s account is a conception of what it is to blame someone. We generally have beliefs about other people’s characters. But Allais suggests that we also have a way of “feeling about” or “*affectively* seeing” them as people.²² This is not so much a matter of our beliefs, but of the emotions we are disposed to feel toward them, the expectations with which we encounter them, the patterns of attention with which we see them, and so on. Allais calls this our “affective attitude” toward a person. And she holds that Lila blames Jack for his promise breaking if she lets it influence her affective attitude toward him. In doing so she “affectively sees” this fault as “centrally attaching to

19 A related idea is sometimes expressed by saying that forgiveness is “elective”; see, e.g., Allais, “Elective Forgiveness.” However, Allais means by this that the wrongdoer cannot, or “paradigmatically” does not (642), have a moral claim to being forgiven. Again, my claim above is (at least in the first instance) about fittingness rather than moral obligations; and I am only claiming that we can *sometimes* fittingly forgive but also fittingly refrain from it—not that that is *always* so. For objections to Allais’s stronger claim, see, e.g., Milam, “Against Elective Forgiveness.”

20 Kolnai highlighted the rough problem that “forgiveness is either unjustified or pointless” (“Forgiveness,” 99). Yet my problem is somewhat different from Kolnai’s. I am concerned with the fittingness of emotions. Kolnai is not. As far as I can see, no one has stated the problem in precisely the manner I have.

21 Allais, “Wiping the Slate Clean.” See also Allais, “Elective Forgiveness.”

22 Allais, “Wiping the Slate Clean,” e.g., 51.

[his] character,” and thus “holds it against him.”²³ Correspondingly, Lila forgives Jack for his promise breaking if, without coming to believe that he was not culpable for it, she *ceases* to let it influence her affective attitude toward him. In doing so, she ceases to affectively regard it as attaching to his character, and thus “wipes the slate clean” with respect to this fault.²⁴

Allais provides an elaborate story about why this may be fitting. It relies on two claims. The first is skepticism about character judgments. Allais argues that we are “never perfectly positioned to judge people’s characters,” that such judgments are “always underdetermined by the evidence”—and specifically, that they are never “epistemically mandate[d]” by a person’s responsible wrongdoing.²⁵ In other words, even if Lila knows that Jack is culpable for his promise breaking, she will not have conclusive evidence about whether it expresses his character. The second claim is a view about fittingness. According to Allais, that an attitude is fitting does not mean that it is “epistemically mandated” by your evidence—but rather, that it is “epistemically permissible” or “not contradicted” by it.²⁶ And within the range of what is not contradicted by your evidence, she suggests, your attitude may permissibly be a matter of what you “pay attention to.”²⁷ In other words, an affective attitude toward Jack is fitting so long as it is not contradicted by conclusive evidence about his character. His promise breaking neither provides conclusive evidence that it centrally attaches to his character nor that it does not. So Lila may fittingly come to adopt an affective attitude toward Jack that is not influenced by his promise breaking, and thus forgive him. But she may also fittingly (continue to) have an affective attitude toward him that *is* influenced by it, or refrain from forgiving.

What is appealing about this is that Allais separates the belief in blameworthiness from the actual attitude of blame. According to Allais, there are different criteria for when we must believe someone is blameworthy and when we must actually blame them in order to have correct beliefs and fitting emotions: we can fittingly refrain from blaming the blameworthy. In this sense, blameworthiness does not *eo ipso* “mandate” blame, and blame is more than an emotional registering just of someone’s blameworthiness. I think this is right, and relevant for forgiveness.

But Allais’s way of explaining it is not fully convincing. Most fundamentally, her picture of attitudes is dubious. It is not true that an attitude is (subjectively) fitting as long as there is no conclusive evidence against it. Suppose Theodore

23 Allais, “Wiping the Slate Clean,” e.g., 51.

24 Allais, “Wiping the Slate Clean,” e.g., 51.

25 Allais, “Wiping the Slate Clean,” 60.

26 Allais, “Wiping the Slate Clean,” 62, 60.

27 Allais, “Wiping the Slate Clean,” 61.

and Luke tell Lila how they judge Glory. Theodore reports that Glory once kindly helped him, and that he generally knows her as generous and kind hearted. Luke says she once strangely refused him a favor, and that he generally finds her ill-natured and mean. Suppose Lila regards them as equally reliable, has no other evidence, and tries to form a fitting attitude toward Glory. Allais suggests that she could now choose what to “pay attention to”—and may fittingly disregard Theodore’s positive opinion and be outright critical about Glory, or vice versa. But this seems wrong.²⁸ Lila’s attitude is only (subjectively) fitting if it reflects her epistemic situation: if to some extent she withholds attitude, or perhaps tentatively regards Glory as being multifaceted. Inconclusive evidence makes certain specific stances (subjectively) fitting. *It* itself does not give us any discretion. So similarly, if Lila has insufficient evidence about whether Jack’s promise breaking reflects his character, her attitude toward him is only fitting if it reflects this: if she withholds attitude, or perhaps regards him as capricious. Her lack of evidence *per se* does not make it fitting for her to ignore our incident outright. If Lila has discretion in fitting forgiveness, it must be grounded in the facts, or in the positive evidence she possesses.

Also, the relevant facts do not seem to be just whether Jack’s fault is characteristic. It is an unduly narrow picture on which blame is all about character assessment. Of course, we sometimes have insufficient evidence about whether a wrong expresses someone’s character. But at least on an ordinary understanding of “character,” and ordinary standards of evidence, our evidence is often sufficient. And in some such cases we still have discretion in forgiveness. Lila might know that Jack’s promise breaking *is* characteristic—e.g., expressive of his general insensitivity to the commitments he has entered. But perhaps nothing much happened as a result of his fault. And perhaps Lila managed not to depend on Jack any further, to just enjoy his endearing cheerfulness in brief encounters, his childlike absorption in the moment, or the oblivious kindheartedness with which he often does her a favor—and thus to see his whole unreliability with enough distance to simply not bother about it anymore. In that case, she might arguably fittingly cease to resent this entire character trait of Jack’s. Or again, she might know that our promise breaking *is not* characteristic—e.g., if our incident lies in the past, and Jack has genuinely felt remorse and truly managed to become reliable. But perhaps that windstorm unrooted the oak so that it destroyed Lila’s nineteenth-century family house. And perhaps Lila cut off all further interactions with him afterward—and never got over the fact that due to the muddler he was she has lost that treasured place. And in that case, she might fittingly continue to feel a certain grudge about that promise breaking. These

28 A related point is made by Ware, “Forgiveness and Respect for Persons,” 253.

responses would not simply be unfitting, or mistaken, as resentment would be if Jack had not been at fault in the first place. Blame and forgiveness are not all about character assessment. They are much more complex responses, it seems.

In another respect, however, Allais's account of forgiveness seems too broad. She does not say *how* Lila must come to a changed attitude toward Jack. But some ways of wiping the slate clean do not amount to forgiveness. Suppose Lila saw it as detrimental to her well-being that she is holding his fault against Jack. And suppose she therefore banged her head on the table until the blows severed the associative ties she drew between him and the incident, and it no longer affected her attitude toward him. Or suppose that Lila's mother died on the evening after the promise breaking, and her associative ties between Jack's promise breaking and the death of her mother are so strong that she somehow never holds it against Jack—just because whenever the incident comes up, the question of how to evaluate her neighbor dwindles into oblivion. In these cases, it seems, she has not properly forgiven. Not all ways of ceasing to affectively regard Jack's fault as attaching to his character are forms of forgiveness.²⁹

Suppose all of this is right. Then our discretion in fitting forgiveness is grounded in the positive facts, not in our epistemic limitations. There is a complex variety of grounds for blame and forgiveness beyond character assessment. And yet not *any* way of wiping the slate clean amounts to forgiving. We must look for an alternative solution to our problem. So let us now turn to the account I propose.³⁰

3. THE CORE IDEA: THE SIGNIFICANCE OF A WRONG

My proposal will build on Allais's idea that we can fittingly refrain from blaming people even if they are blameworthy. But I will account for it differently. My core

29 Interestingly, Allais discusses the fact that some forms of overcoming negative reactive emotions do not amount to forgiveness (e.g., "Wiping the Slate Clean," 43–50, 57–58). She cites "forgetting the wrong," "putting the wrongdoer out of your mind," or "regarding her as beneath your concern" as examples (58). According to Allais, these things do not amount to forgiveness because they do not involve "a change in your [affective] view of the wrongdoer" (58) with respect to their wrong: they just mean you lack any such affective attitude toward them. My point is that even Allaisian changes in your affective view of the wrongdoer need not amount to forgiveness. This is because Allais does not add any constraints about *how* you must come to cease to affectively regard a fault as attaching to the character of your wrongdoer in order to forgive.

30 For some alternative proposals (differing from Allais's and mine), see, e.g., Kolnai, "Forgiveness"; Murphy and Hampton, *Forgiveness and Mercy*; Hieronymi, "Articulating an Uncompromising Forgiveness"; Calhoun, "Changing One's Heart"; Callard, "The Reason to Be Angry Forever"; or Na'aman, "The Fitting Resolution of Anger."

suggestion will be that the representational content of resentment is more complex than the common assumption has it. The reactive attitudes are a response to someone's (putatively) responsibly evincing wrongful attitudes. But they are not *just* a response to that. This can be argued without reference to forgiveness. So in this and the following two sections I will set forgiveness aside and sketch a picture of the reactive attitudes that is independently plausible.³¹ I will come back to forgiveness in section 6.

It is worth bringing to mind some commonplace facts. A myriad of people have at some point somewhere responsibly evinced wrongful attitudes, and we often know of that and respond to it emotionally. Yet we generally do not adopt *proportionally strong* reactive attitudes toward all of these faults. Suppose Lila knew that Emperor Hadrian was wittingly betrayed by his friend and predecessor Trajan in second-century Rome. And suppose she also knew that her husband, Robert, was wittingly betrayed by his friend Marcelle in precisely the same manner. Suppose she responds with some mild indignation to the first case, but with quite strong indignation to the second. This would be entirely natural. We would say that she *cares* more about Marcelle's fault than about Trajan's. Indeed, intuitively, such unequal caring seems warranted, or supported by reasons. Generally, there are reasons for you to care more or less about facts. The fact that Marcelle harmed *her husband* is a reason for Lila to care about her fault. But the fact that Trajan lived in ancient Rome, or was five feet, eight inches tall, or perhaps had the same number of hairs as Lila generally is not a reason for her to care about *his* fault. So given her relation to Robert, and her lack of any relationship to our Romans, it seems warranted for Lila to care more about Marcelle's fault than about Trajan's. Marcelle's fault concerns Lila more, or has more normative import or *significance* for her. Indeed, something even stronger seems true. If caring is a response to reasons, and if there is no good reason for her to care about Trajan's fault but plenty of reason to care about Marcelle's, it would in an important sense be unwarranted for Lila to care *equally* much about either. If she got worked up equally much, or little, about either fault, her caring would be at odds with her reasons to care, or with the grossly unequal significance that these faults have for her. She would ignore the import of her relations.

Now what kinds of reasons are these reasons to care? We might try to suggest that they are non-fittingness reasons. We might stick with the simple view about the representational content of the reactive attitudes (from section 1), on which the fittingness of a certain degree of indignation is determined fully by the de-

31 Here and in what follows, I say "putative(ly)" because whether you respond with a certain emotion ultimately depends on whether you *take* something to have a certain property, rather than on whether it actually has it.

gree of the wrongdoer's fault. On this view, the uniquely *fitting* response for Lila would be equal indignation at Trajan and Marcelle. But we might suggest that Lila has non-fittingness reasons to (get herself to) care more about Marcelle's fault: prudential or moral reasons to be emotionally invested in her husband, say, but insouciant about our Romans. So it is somehow *all-things-considered* reasonable for her to (get herself to) feel unequal indignation at them. But this seems dubious. First, it seems to intuitively mischaracterize caring. There are nonrepresentational feelings for which there are only non-fittingness reasons—like hunger, tiredness, or pain. But, intuitively, the attitude of caring seems different. It seems propositional or representational. Something's being significant for you seems a reason for you to care in much the same way as something's being dangerous is a reason for you to be afraid. Second and more specifically, the suggestion seems to misdescribe Lila. Suppose someone offers her \$1,000 if, or tortures a cat unless, she feels resentment at her chair. Perhaps this gives her prudential or moral reasons to (get herself to) resent it: to misrepresent the facts, or emotionally pretend her furniture was blameworthy while knowing it is not. But intuitively our case is different. In responding more strongly to Marcelle's fault than to Trajan's, she does not seem to misrepresent anything, or involve herself in a contradiction between her emotions and her beliefs. In particular, Lila's emotions need not represent Marcelle as having evinced worse attitudes than Trajan, or as being more responsible for them. Lila may know that Marcelle and Trajan are equal in these respects, and intuitively her attitudes need not contradict this. Indeed, third, recall that Lila's emotions are entirely standard. So the proposal implies that *all* of us are *constantly* involved in blatant misrepresentations of people's faults, or contradictions among our representational mental states. Such widespread attribution of error seems hard to bear. If anything, it seems, Lila's emotions would misrepresent the facts if she got worked up as little about Marcelle's fault as about Trajan's, or as much about Trajan's as about Marcelle's. In doing so, she would falsely pretend that these faults concern her equally little, or much. She would misrepresent their unequal significance for her.

I do not see any convincing way of salvaging a non-fittingness interpretation of Lila's reasons to care. More plausibly, these reasons of significance are fittingness reasons. Something's being significant for you makes it *fitting* for you to care about it. That is to say, its significance is part of the representational content of your response toward it. So in the remainder of this paper, I will try to turn these first intuitions into a more full-fledged account of the fittingness of reactive emotions. Let us first clarify the idea of significance. It is not that something either is or is not significant for you. Take Trajan's betrayal of Hadrian. This betrayal may make a range of emotions fitting in virtue of instantiating a range of value prop-

erties. Perhaps it proved the theory of some diligent twentieth-century historian, allowed her a publication, and got her a deserved career. If so, as something that was good for that historian, our betrayal may make it fitting to be glad. Perhaps it constituted an aesthetically compelling narrative, and as such may make it fitting to feel aesthetic appreciation. Perhaps it amazingly figured in a dream of a Peruvian farmer who knew nothing of ancient Rome, and as such may make it fitting to be amazed. And I assume it also makes it fitting to feel indignation or resentment. I will say it does so, specifically, *as* something for which Trajan is blameworthy. So our betrayal may be significant for you in different respects, or as an instantiation of different properties. For our diligent historian, the betrayal seems significant *as* something that was good for her. This property of the betrayal concerns her. But it does not seem significant for her as something for which Trajan is blameworthy. She does not have any personal relationship to our Romans. For Trajan's and Hadrian's friends, in contrast, our fact will be significant as something for which Trajan is blameworthy, but insignificant as something good for our historian. And for still other people, it might be significant in still other ways.

Here is how I will understand this more formally. By saying that some x is (more/less) *significant* for you *as* an instantiation of a certain objective value property F , I mean that, due to your relationship to the fact that x is F , it is fitting for you to be (more/less) affected by the emotions that x 's being F ultimately makes agent-neutrally fitting. If you are (more/less) so affected, due to your (putative) relationship to x 's being F , I will say you *care* (more/less) about x as an instantiation of F . Indeed, for you to care (more/less) about x as an instantiation of F just is for you to be (more/less) affected by the emotions that x 's being F ultimately makes agent-neutrally fitting due to your (putative) relationship to the fact that x is F . Put simply, significance makes it fitting for us to care, and caring is our response to (putative) significance.³² But caring is not an *extra* emotion, alongside indignation, gladness, or admiration, say. It is a component of your feeling another emotion to a certain degree. And in this sense, the significance of x as F for you will be part of the representational content of your response to x 's being F .

Let us spell this out for the reactive emotions. I will assume that what my

32 Note that, in this explication, x might be anything—a fact, an event, an object, or whatever. By “objective value properties” I mean value properties that are not essentially relative to you (e.g., to your epistemic situation), perhaps like something's being disappointing or surprising. What I say could be extended to such essentially relative properties too. But for simplicity, I focus on objective ones. By the emotions that some fact “ultimately makes agent-neutrally fitting,” I mean those it would make fitting for an impartial spectator, or regardless of its agent-relative significance.

being blameworthy for a fault ultimately makes fitting are specifically (and only) the negative reactive attitudes. So that my fault is (more/less) significant for you *as* something for which I am blameworthy means that due to your relationship to the fact that I am blameworthy for it, it is fitting for you to be affected (more/less) by the resentment or indignation my fault ultimately makes agent-neutrally fitting. And if you therefore feel (more/less) indignation at me, you care (more/less) about my fault *as* something for which I am blameworthy. Indeed, for you to care (more/less) about my fault in this respect just is for you to feel (more/less) indignation at me, in response to my fault's (putatively) being (more/less) significant for you as something for which I am blameworthy. Thus your overall degree of indignation among other things comprises your caring. In feeling a certain indignation at me, you represent me as thereby having manifested an attitude toward you that was morally wrong and for which I was responsible, *and* represent my fault as being significant for you. The stronger your resentment, the more seriously wrong you represent my attitude as having been, or the more fully responsible you represent me as having been for it, *or* the more significant you represent it as being for you. In this sense, the reactive attitudes, like other emotions, have an agent-relative element. They are not just a response to faults. They are a response to faults that are (putatively) significant for you.

This is the account I will develop in what follows. And this, I suggest, is why we can fittingly refrain from blaming people even if they are blameworthy. Lila's mild form of indignation at Trajan need not falsely represent him as not being equally blameworthy as Marcelle. It can correctly represent his fault as equally serious, but less significant for Lila than Marcelle's. To anticipate, I will suggest that Lila would forgive Jack for his promise breaking if she did not care about it anymore, as something for which he is blameworthy, and thus represented it as being sufficiently insignificant. But before I turn to forgiveness, let me elaborate more on this picture.

4. THE REASONS OF SIGNIFICANCE

I have indicated what *kind* of reasons may be relevant for overcoming resentment: fittingness reasons of significance, which you have in virtue of your relationship to someone's fault, as something for which they are blameworthy. The weaker your relationship to it, the weaker your reasons to care. But what relationship to something's instantiating a value property makes it significant for you as an instantiation of it? *When* do you have reasons to care about something?

This important question seems underexplored in the theory of emotions, and

would be worth devoting an entire book to.³³ But I suggest the answer has to do with what *grounds* the relevant property instantiation: x is significant for you as an instantiation of F to the extent that you are bound up with what grounds the fact that x is F . Trajan's betrayal is significant for you as something that was bad for Hadrian to the extent that you are bound up with the grounds of this badness—e.g., involved with Hadrian's suffering. It is significant for you as something that was good for our historian to the extent that you are bound up with the grounds of this goodness—e.g., confronted with her joy or the benefits she reaps from her career. In a thorough general answer to our question, the phrase of your being "bound up" with something would have to be spelled out more. But I will assume we have a certain intuitive understanding of it, and for present purposes this will suffice. So rather than elaborating on this general criterion, let me say how I think it applies to blame.

Suppose a wrongdoer is blameworthy for wronging a victim. If I am right, his blameworthiness will be significant for you to the extent that you are bound up with what grounds it. So what will that be? There are different stories, but let us assume the picture of Strawson.³⁴ On this picture, our blameworthiness is essentially grounded in our "involvement . . . in inter-personal human relationships."³⁵ For Strawson, it is a constitutive feature of a normal interpersonal relationship that we demand of others a certain form of goodwill or regard toward us.³⁶ The presence of this demand distinguishes such relationships from those we have toward children or psychopaths. And to make such a demand just is to respond with negative reactive attitudes to its violations.³⁷ If this is roughly right, then what grounds our wrongdoer's blameworthiness is the fact that the victim is a source of demands on our attitudes, that the wrongdoer is a responsible person with whom we stand in relationships shaped by such demands, and that his attitude toward the victim violated a demand she is a source of.

So, for one thing, his blameworthiness will be more significant for you the more you are bound up with the importance of the pertinent demand. Most prominently, it will be more significant for you the more you are bound up with the *victim*, or with the fact that she is not just something that is fascinating for biochemistry, say, but a source of moral demands. When precisely you are so bound up with her depends on what about her grounds these demands. But plausibly,

33 Though for very interesting explorations, see, e.g., Helm, *Emotional Reason*; Kolodny, "Which Relationships Justify Partiality?"; and Lord, "Justifying Partiality."

34 Strawson, "Freedom and Resentment."

35 Strawson, "Freedom and Resentment," 9.

36 Strawson, "Freedom and Resentment," 15.

37 Strawson, "Freedom and Resentment," 23.

you will be bound up with her in this manner if you are involved with the value of her leading the life she chooses, faring well, and being embedded as a member in equal standing in a moral community—if you have been confronted with this value and it matters or should matter to you. Other things equal, it is fitting for Lila to care more if someone got aggressive and broke her husband's nose than if that person got aggressive and broke the nose of a stranger. Indeed, other things equal, it would be unfitting for her to care more about the fault that affected a stranger than the fault that affected her husband. Her caring would be at odds with her reasons to care, or her involvement with the victims. She is more bound up with her husband as a source of demands than with a person she has never met.

For another thing, a wrongdoer's blameworthiness will be more significant for you the more you are bound up with the *wrongdoer*, or with the fact that he is not just an object of historical studies or social planning, but an object of moral demands, or a person in human relationships shaped by them. And this will be the case to the extent that you actually stand in such a relationship with him. When precisely you do, or to what extent, is a complex issue. Plausibly, you may have such a relationship even to people you do not literally interact with, as when you regard your favorite author or a stranger under the guise of a fellow person. But you are more fully involved with someone to the extent that you actually share your life or navigate a common space with them. Other things equal, it is fitting for Lila to care more about xenophobia in her mother than about the same xenophobia in a stranger, even if the relevant victims are the same. Indeed, it would be unfitting, or at odds with the import of these faults for her, if she cared more about the fault of a stranger than that of her mother. She is more bound up with her mother as an object of interpersonal demands.

But these are only roughly the main determinants of significance. There are a host of other factors beyond this. Most importantly, if your involvement with the wrongdoer matters, *his* response to his wrong matters too. At least insofar as you actually interact with him, you stand in a relationship with the *present* person, not with whom he once was. So the more *he* distances himself from his fault through remorse and a change for the better, the less it manifests a deficiency in a relationship in which you actually stand, or in a person whose being the object of demands you are bound up with. The more he separates himself from his wrongful attitudes, the less you are involved with his past demand violation, or with what grounded the fact that he was to blame. His past self will be a bit like a stranger. Other things equal, it is fitting for Lila to care more about her brother's wrongful outburst of rage if he has never felt remorse for it than if he has felt genuine guilt and has since tried hard and succeeded to be placid.

Moreover, it is not that you will simply have a more or less close relationship

to a wrongdoer. You have different relationships with different people, or relationships to specific aspects of them. So you are more fully bound up with the fact that a wrongdoer violated a demand the more that demand is important in the specific relationship you have to him. It might affect Lila that her neighbor Jack is endearingly cheerful in brief encounters, that he fully respects her privacy, and that he is generally very happy to do her a favor. But Jack's stubbornness in political matters, or his cowardice, or his characteristics as an employer, might be largely irrelevant to their relation. So while it is fitting for Lila to care more about Jack's political stubbornness than about that of a total stranger, other things equal, it is fitting for her to care even more about that stubbornness in Louisa, with whom she interacts in the local government. And that is so even if she does not have a closer relationship to Louisa than to Jack on the whole.

Similarly, it is not just your relationship to the victim that determines how much you are bound up with the relevant demand. There are other ways in which you might be bound up with it, or its grounds, or the grounds of why it is important. For instance, you are more involved with the importance of a demand if you have experienced how a violation of it can be harmful. You are more fully bound up with the demand to drive cautiously if as a result of someone's reckless driving you have once lost a child. The same may be true if you have dedicated your life to victims of road accidents. And something similar holds, all the more poignantly, if the *present* reckless driving had terrible consequences that you are involved with. Other things equal, it is fitting for Lila to care more if her friend Della drove recklessly and killed the child of Lila's neighbors than if she luckily did not harm anyone.³⁸

All of this is rough, and plausibly not exhaustive. But it illustrates the general idea. The emerging overall picture is that our patterns of significance form a complex web. On the one hand, significance is infectious: if some things are significant for you in certain respects, then so are others that are appropriately connected to them. If the value of that house is significant for Lila, then so is Jack's promise breaking, which puts it at peril. On the other hand—and as will become important later on—significance must always be *relative*: if certain things are significant for you in certain respects, then others that are not connected to them will have to become less significant for you *in comparison*. If on the same evening there is a train crash that kills Lila's mother, Jack's promise breaking will suddenly be relatively insignificant next to it. The more specific upshot is that it is fitting for you to care about a wrongdoer's blameworthiness to the extent that

38 I think this grounds an important kind of moral luck, which I explore in Riedener, "Don't Make a Fetish of Faults." Scanlon makes a similar claim, within his different conception of "blame" (*Moral Dimensions*, 147–51.).

you are bound up with the victim as a source of demands and the wrongdoer as a participant in human relationships—or specifically, to the extent that the wrong is still associable with the wrongdoer's attitudes, matters in your relationship to him, actually harmed you, and so on. These reasons are part of what reactive attitudes respond to, and of what they represent. Or so I claim.

5. THE DISCRETION IN CARING

I have suggested what kind of reasons our reasons of significance are. And I have made a proposal about when they arise. The significance of a value fact for you, we might say, is a quantity determined by how much you are bound up with its grounds, and determining how much caring is fitting in response to it. It remains to examine what the effect of these reasons will be. If significance is such a quantity, does that mean you must always care to one specific degree in order to have fitting attitudes? Or may you generally care to different degrees and still have fitting emotions?

At least in theory, there are two extreme views about how much discretion in fitting caring you have. On the one hand, one might think you have total such discretion. You can always care as much or as little as you please, without your caring ever being unfitting. On the other hand, one might think you have no such discretion at all. Given your relationship to a fault and to everything else, there is always just a single degree to which you can fittingly care. I have already suggested that the first view is intuitively implausible. Some forms of caring seem unfitting. Suppose Lila's close friend Della drove with outrageous recklessness, killed Lila's nephew, and hardly showed any remorse. And suppose someone named Edward drove with that very same recklessness a hundred years ago on the other side of the planet, but did not cause any harm, soon felt deep remorse, and drove with uttermost caution ever after.³⁹ Other things equal, Lila cannot hold that the second incident concerns her more than the first. It just does not. And if she cared very much about Edward's but very little about Della's fault, her attitudes would misrepresent the significance of these faults for her. In caring about a fault, there is generally a minimal degree to which you must and a maximal degree to which you may care about it, in order for your caring to be fitting.

Why is this so? Well, it just follows from the fact that caring is a represen-

39 Recall that I regard the degree of your fault as being fully determined by the wrongness of, and your responsibility for, your attitude. Whether or not, unluckily or luckily, you end up harming someone with a given attitude does not determine the degree of your fault. It determines its significance. So as I understand the cases involving reckless driving in this section, the faults of the agents are the same. They only differ in significance for Lila. (On significance and moral luck, see also note 38 and the corresponding paragraph.)

tational response to significance, in the manner I have suggested in section 3. Suppose you could feel extreme “shfear” in response to a totally harmless mouse, and no shfear at all in response to a very dangerous snake, without this being an unfitting response to the dangerousness of these animals. Then shfear would not be a representational response to danger. It would be a nonrepresentational feeling like hunger, or perhaps a representational response to something altogether different. Similarly, suppose you could care very much about one fact that intuitively did not concern you at all, and very little about another fact that concerned you very much, and your caring would not be an unfitting response to the significance of these facts for you. Then caring would not be a representational response to significance. It would be a nonrepresentational feeling, or a representational response to something else. But it is a representational response to significance, or so I have argued.

However, to *some* extent at least you are free in how much you care. It is not always just one degree of caring that is fitting. Suppose Lila’s acquaintance Grace once drove with a recklessness equal to Della’s and Edward’s. And say Lila has long had a good but rather impersonal relationship to Grace, that Grace’s driving took place in distant Anchorage and damaged a fence post, and that she felt considerable but not *quite* the remorse she should have. These facts may determine some limits to Lila’s caring. Other things equal, she can neither fittingly pretend that Grace’s fault does not concern her *at all* (or less than Edward’s), nor that it concerns her tremendously much (or more than Della’s). But in between these extremes, she arguably has a large *range* of degrees of caring that are fitting. Within this range, she can fittingly care more or less. Not just one degree of caring seems fitting.

Why is this so? It helps to consider other cases, independent of significance, where we are justified to have attitudes of varying strength. Take aesthetic value. Consider a good artwork, such as Klee’s little painting *Dieser Stern lehrt beugen*. And compare it to an absolute masterpiece like Tolstoy’s *War and Peace*, and to an amateur work like a clay sculpture Lila once produced in adolescence. Other things equal, intuitively, you cannot fittingly admire Klee’s small painting more than Tolstoy’s novel, or less than Lila’s sculpture. In doing so you would pretend it was more valuable than the former or less valuable than the latter, and that would be a misrepresentation. But compare Klee’s painting to another artwork that is far from dilettantish but not quite epoch making either—such as one of Bartók’s *Romanian Folk Dances*. Here, intuitively, you can fittingly feel the same degree of admiration toward them. But you can also fittingly feel somewhat

more admiration at the painting than at the composition, and vice versa. None of these combinations of attitudes must be misrepresenting their worth.⁴⁰

There might be different ways to explain this. But a standard and I think plausible way of doing so is in terms of *incommensurability* in values or reasons. That a certain form of value gives rise to incommensurability means that instances of it are not representable by a single value function, but only by a set of value functions. Your attitudes are fitting if and only if the values they represent things as having accord with *one* of the functions in the set. More intuitively, we might say, instances of the value are not representable as a single determinate quantity (precisely k on some scale), but only as a range of values (somewhere between n and m). And your attitudes are fitting if and only if the values they represent things as having fall within that range.⁴¹ Thus in order to have fitting attitudes, you must have some attitudes from this range, but may have any attitudes in it: you do not have conclusive (fittingness) reasons for one degree of emotion, but sufficient (fittingness) reasons for many. This can account for your discretion in response to art. Aesthetic value is not a precise determinate quantity. The value range of Klee's painting lies fully between the ranges of Lila's sculpture and Tolstoy's novel. So in order to have fitting attitudes, you must admire it more than the former but less than the latter. But since the value ranges of the works by Klee and Bartók overlap, you can fittingly feel equal admiration toward them, or somewhat more admiration for either.⁴²

- 40 Why not think the works of Klee and Bartók are precisely equally valuable? Here is the standard "small improvement argument." Had Bartók chosen a slightly more mesmerizing melody, his piece would have been determinately (slightly) better than it is. But it would still not necessarily be determinately better than Klee's painting. So the actual composition cannot be precisely equally as valuable as the painting. For classic defenses of the pervasiveness of such incommensurability, see, e.g., Raz, *The Morality of Freedom*, ch. 13; and Broome, "Is Incommensurability Vagueness?" For a classic challenge, see Chang, "Introduction."
- 41 Strictly speaking, the set- and range-theoretic statements are not equivalent. Let V be a set of value functions. It could be that on all individual functions in V , the value of x is greater than that of y —i.e., that for all v in V , $v(x) > v(y)$ —and yet that the ranges of values assigned to x and y by different functions in V overlap—i.e., that for some u and w in V and real number k , $u(x) = w(y) = k$. The set-theoretic formulation is more precise, or allows us to capture more fine-grained differences—and is thus standard in decision theory (see e.g., Dubra, Maccheroni and Ok, "Expected Utility Theory without the Completeness Axiom"; also Broome, "Is Incommensurability Vagueness?"). But since the range-theoretic statement is simpler, and the details do not matter for our purposes, I rely on that in what follows.
- 42 For an analysis of incommensurability in terms of permissible ranges of attitudes, see, e.g., Rabinowicz, "Value Relations Revisited," esp. 139n3; and Raz, "Incommensurability and Agency." There is a related discussion about reasons for action: Why are you generally justified to do various actions, rather than required to do any one? For an explanation of such discretion in action in terms of incommensurability, see, e.g., Raz, *Engaging Reason*. My picture

Incommensurability in this sense is arguably pervasive. There is ample incommensurability in badness, tragedy, or awesomeness, say—or rarely just a single degree of sadness, dismay, or awe you can fittingly feel about instantiations of these properties. Very plausibly, there is similar incommensurability in significance. Significance is not a precise determinate quantity. It does not give you conclusive (fittingness) reasons for one degree of caring, but sufficient (fittingness) reasons for different degrees. Lila cannot fittingly pretend that the recklessness of her acquaintance Grace in Anchorage concerns her less than Edward's or more than Della's. But compare Grace's fault with an equal fault by a stranger in Lila's neighborhood. In many such cases the significance resulting from Lila's involvement with the wrongdoer will not be fully commensurable to that resulting from her involvement with the wronged. So Lila can fittingly care equally about these faults, or somewhat more about either of them. The incommensurability of significance gives her discretion in fitting caring.

Indeed, it arguably gives her quite a lot of it. For one thing, the relevant incommensurabilities will be ample. Incommensurability arises especially when a property has a number of diverse and hard-to-measure factors. Aesthetic value is determined by emotional depth, beauty, and immediacy, by the interplay of harmony and tension, and so on. These factors are diverse and hard to measure, and this grounds the wide incommensurabilities in art.⁴³ Significance too is based on numerous factors, which are equally diverse and hard to measure. So here too incommensurabilities will be substantial. Moreover, they will multiply through their relations. Lila has a certain discretion with respect to how much she cares about Jack *tout court*, or about the demand for reliability as it applies to him, or the hours she has lost in fetching her saw. If she cares much, or little, about these background things, she may fittingly care correspondingly more or less about his fault. The incommensurabilities will grow with a ripple effect through the knots of the web of significance. And, indeed, if significance is always also *relative* as I have suggested, then the significance of Jack's promise breaking will also depend on how much she cares about the train crash of her mother, or about a recent earthquake in the Philippines and so on, even if those things are not intrinsically connected to it at all. So Lila will have a broad range of ways of caring about Jack's fault that are fitting.

is also reminiscent of Gert's account (see, e.g., "Requiring and Justifying"). Gert contrasts his account with the idea of incommensurability. But he also thinks the strength of reasons must be represented by value ranges, whose lower and upper bounds determine their "requiring" and "justifying" dimension respectively. The upshot of this is parallel to the picture I have sketched. For different accounts of our discretion in action, see, e.g., Greenspan, "Making Room for Options"; and Portmore, "Imperfect Reasons and Rational Options."

43 See, e.g., the standard arguments in Chang, "Introduction," 14–17.

Hence the following picture suggests itself. In caring about Jack's fault to a certain degree, she does not represent it as giving her conclusive (fittingness) reasons for this and only this degree of caring. She represents it as giving her sufficient (fittingness) reason to care to that degree. And due to the incommensurability in significance, she might have sufficient reasons for different degrees. Given the history between Lila and Jack, his own response to his negligence, the extent to which she was actually harmed, and so on, as well as a context of other value facts, there *will* be forms of caring about his promise breaking that are unfittingly detached, and others that are unfittingly involved. But in between these extremes, there will be a range of different fitting responses. In light of Lila's complex relationship to Jack's promise breaking, she may fittingly care considerably about it, and fittingly care rather less—just as you may fittingly feel different degrees of admiration for Klee's little painting.⁴⁴

6. AN ACCOUNT OF FORGIVENESS

This indicates a new account of forgiveness. Again, the stronger your resentment, the more serious you represent my fault as having been (in terms of wrongness and responsibility), *or* the more significant you represent it as being for you. And that you represent it as being very (or not so) significant for you means you represent it as being such that in virtue of your relationship to it, you have sufficient reasons to care greatly (or little) about it, as something for which I am blameworthy. That is how your blame registers more than my blameworthiness, and why you can fittingly refrain from blaming me even if I am blameworthy. And the overcoming of resentment in forgiveness revolves around this factor of significance.

44 In view of Allais's claim that "wrongdoing *entitles* us to resent, but this does not mean that it *obliges* us to" ("Elective Forgiveness," 647), it might be worth summarizing the senses in which, on the picture I have sketched, the fact that a certain emotion is fitting means you are "required" to feel it. There is a simple sense in which it does mean this. If the fact that x is F makes emotion E , and only E , fitting, then in order to have the (or a) fitting response to x 's being F , you must feel E . If you do not, you lack a fitting response to that fact. However, there are three different senses in which for all I have said "fittingness" is not "requiring," but may be merely "permissive." First, there is the more fundamental question about whether there is, say, a moral or prudential requirement to have fitting attitudes in the first place. Perhaps there is not. Thus even if you have unfitting attitudes, perhaps you are not violating any such requirement. Second, there is the form of emotional oblivion mentioned above. If you lack any representation of x 's being *or not* being F , you do not have unfitting emotions toward x 's being F . Indeed it might be that all of your emotions are fitting. You just lack the fitting response, specifically, to x 's being F . Third, if there is incommensurability, it might be that x 's being F makes both E_1 and (impossible) E_2 fitting. Then you might lack E_1 and still have a fitting response to x 's being F —*viz.*, by having E_2 .

If Lila forgives Jack for his promise breaking, she does not *care* about it anymore, as something for which he is blameworthy. She comes to represent it as not being *significant* for her anymore, in this specific respect. She comes to represent Jack's fault a bit like that of a stranger, at least as his blameworthiness goes. Now this does not mean that Lila would eschew *any* response to Jack's fault, or no longer care about it *tout court*. There are other responses to Jack's promise breaking besides the negative reactive emotions. Lila may respond with sadness to the incident and henceforth feel a sense of sorrow in Jack's presence; she may respond with a desire to stay somewhat out of his way; or she may respond with curiosity and attempt to find out why Jack broke his promise. But these responses would not ultimately be made fitting by Jack's being *blameworthy* for his fault. They would ultimately be made fitting, say, by the fact that Jack's promise breaking was bad for Lila (e.g., if the oak fell on her house), or by the fact that it is evidence for the undesirability of future involvements, or by the fact that it promises interesting insights (e.g., about how to prevent promises being broken in the future). Again, I take it that what our being blameworthy for a fault ultimately makes fitting are specifically and only the negative reactive attitudes. So if Lila had these other responses, she would care about Jack's fault in these other respects—as something that was bad, or evidence for the undesirability of involvement, or intellectually interesting. And all of these responses would be compatible with her having forgiven. In forgiving, she just comes to no longer care about the promise breaking specifically *as* something for which Jack is blameworthy. In this sense, she overcomes her resentment, due to its (putative) lack of significance for her in this specific respect.

If all I have said is correct, this solves the problem from section 1. In no longer feeling resentment, Lila would come to represent Jack's fault as being such that in virtue of her relationship to it she has sufficient reasons not to care about it anymore. And at least in some circumstances—perhaps if nothing much happened and Jack felt some remorse—this may certainly be true. So she may fittingly not feel resentment. In still feeling resentment, Lila would represent Jack's fault as being such that she has sufficient reasons to still care about it. And this may also be true, even in the very same circumstances. Since significance is not a precise determinate quantity, but involves incommensurabilities, these circumstances will give her sufficient reasons for several degrees of caring. So even when she may fittingly not feel resentment, she may also fittingly still feel it. And note that if she does care less, she does not just *lack* any representational response to Jack's fault. To represent it as insignificant is not to not represent it at all. So in no longer caring she would not just end up with emotions that are not unfitting. She may end up in a positively fitting emotional state. There is

no real “paradox” in forgiveness. Our discretion in fitting forgiveness is part of our more general discretion in how much we care about things.⁴⁵ As emphasized, that latter discretion has limits. But then our discretion in forgiveness has limits too. Sometimes—with remorseful wrongdoers whose wrong is long past and has not harmed us at all—resentment is no longer fitting. And sometimes—with unapologetic close wrongdoers who have just harmed us severely—resentment just *is* the fitting response. So I take it the proposal gets the measure of our discretion quite right.

However, no longer caring about a wrong as something for which the wrongdoer is blameworthy plausibly is not sufficient for forgiveness. If Lila sought peace of mind, took a tranquillizer pill, and eventually no longer blamed Jack due to the chemistry of this pill, she has not forgiven. Neither has she done so if she does not care about his wrong *only* because her mother died and this rendered her indifferent about that promise. Nor is it forgiveness if out of contempt she no longer cared about Jack *tout court*. So we need to say more. Nothing of what I have said so far depends on how we explain why these ways of ceasing to care are not forgiveness. But let me add a proposal. Lila’s lack of resentment must be reason responsive (not just the effect of some pill), and the relevant reasons must have their source at least partly in Jack (not *just* in some unrelated event) and in something that is lovable about him (not in what perhaps makes fitting contempt).⁴⁶ So I suggest that Lila forgives Jack for his promise breaking if and only if she no longer cares about it partly *out of goodwill* toward him. Or more precisely: Lila forgives Jack for his promise breaking if and only if, while believing that he was at fault for it, and partly out of goodwill toward him, she no longer cares about his promise breaking *as* something for which he is blameworthy. To forgive just is to no longer care in this manner.

What do I mean by this? As I will understand it, you ϕ out of an attitude if your having that attitude plays an unmediated causal role in why you ϕ . And it is in this sense that Lila might not care about Jack’s fault “out of goodwill” toward him. Her not caring might partly be the unmediated consequence of her having a positive attitude toward him. The phrase “unmediated” here is intended to rule out cases where Lila’s goodwill indirectly causes her not to care. If out of goodwill

45 One might think that Lila’s relationship to the victim will not be relevant for forgiveness. After all, *she* is the victim, and (one might think) will have a fixed relationship to herself. But note that what matters is her relationship to the victim *as* a source of demands. And she can be more or less bound up with herself in this respect. For instance, if she is struggling for a sense of self-worth, she will experience the importance of others’ respect toward her firsthand. In this case, she may be more bound up with herself as a source of the demand for respect than if she has a fully firm sense of her worth.

46 This is emphasized especially in Milam, “Reasons to Forgive.”

Lila intends to benefit Jack, and takes a pill that causes her not to care about his fault, she has not properly forgiven. In the sense I intend, she did not cease caring “out of goodwill,” but due to the effect of that pill. More positively, in a case of fitting forgiveness, there is some property of Jack that makes it fitting for Lila to feel goodwill toward him. Lila feels goodwill in response to that. But that property of Jack also makes his fault less significant for Lila as something for which he is blameworthy. So in coming to appreciate that property, Lila also comes to see his fault as less significant for her. And it is partly due to this that she ultimately does not care about it anymore as something for which he is blameworthy.

Let me illustrate this with examples. The property of Jack that makes Lila’s goodwill fitting might be something that renders his promise breaking *intrinsically* less significant for her. Suppose Jack repented his fault. This means Lila is less bound up with the grounds of his being blameworthy for it. So on the account from section 4, it makes it fitting for her to care less about it. But it arguably also makes it fitting for her to feel goodwill toward him. If Lila comes to appreciate his remorse, responding with goodwill toward it, and therefore (among else) no longer cares about his promise breaking, she forgives him. But I think the relevant property of Jack might also be something that renders his promise breaking only *relatively* less significant for Lila—i.e., makes other things more significant, and thus Jack’s fault less significant *in comparison*. Suppose Jack beams with a disarmingly endearing cheerfulness, an adorable childlike gaiety, and despite all his weaknesses it is a precious delight to have him around. Or suppose he is diagnosed with a terrible disease, and undeservedly soon faces death. These facts have nothing to do with the grounds of his blameworthiness for that promise breaking. So they do not render it intrinsically less significant for Lila. But they make it less significant in the overall scheme of things, and specifically in her dealings with him. And they arguably also make it fitting for her to wish him well. If Lila comes to appreciate these aspects of Jack, responding with goodwill to them—with a love for her scatterbrained yet congenially joyful or pitifully unlucky neighbor—and therefore (among else) no longer cares about his promise breaking in the pertinent manner, she arguably also forgives him.⁴⁷ But I say it must be *partly* out of goodwill that Lila no longer cares. There may be other facts that render his fault less significant, and do not warrant any

47 Intuitions about whether such cases amount to “forgiveness” will be more disputed: Milam (“Reasons to Forgive”), say, would presumably deny it. I think they do. But my account could be modified to imply that they do not. We might require that the relevant property of Jack must render his fault *intrinsically* (not just relatively) less significant for Lila. Note also that as I understand the phrase “out of goodwill,” it in effect implies a right-kind-of-reason requirement for forgiveness: in forgiving, Lila does not care due to something that (putatively) renders Jack’s fault less significant, or is a *fittingness* reason not to care.

such goodwill. Perhaps Lila's mother died on that evening, and this made our incident unimportant in comparison. Or perhaps Lila moved away, and is simply no longer involved with her former neighbor. These facts do not warrant goodwill toward Jack. So if it is *only* due to them that Lila no longer cares about his fault, she has not properly forgiven. But these facts do make his promise breaking less significant for her. So they can contribute to making it fitting for her not to care anymore, and help her actually not care. And as long as it is *partly* due to such facts that she does not care, and partly also out of goodwill—partly also in response to his remorse, say—they can contribute to making forgiveness fitting, and help her actually forgive. The threshold about when enough goodwill is involved will be vague. But this just mirrors that the concept of forgiveness is vague—and that there is a *certain* continuity between forgiveness and phenomena like oblivion or distraction.⁴⁸

So if all of this is correct, then *pace* Allais, not all ways of ceasing to hold the promise breaking against Jack amount to forgiveness. Forgiveness is a narrower phenomenon than that. It must in part be a genuine *response* to a (putative) fact about Jack. Also, Lila's discretion in fitting forgiveness does not hinge on her epistemic limitations. It is fully grounded in the facts, or in the actual incommensurabilities of significance. And finally, blame and forgiveness are not all about assessing our characters. Of course it matters whether Jack's promise breaking is (still) characteristic of him. If it is not, this renders his fault less significant for Lila, and may thus allow her to fittingly not feel resentment in response to it. Indeed, it also makes it fitting for her to feel goodwill toward him, and thus may allow her to fittingly *forgive*—or not blame him partly *out* of such goodwill. But the significance of Jack's fault for Lila hinges on many other factors too, which do not depend on whether his unreliability is (still) characteristic. It depends on how close a relationship she (now) has to him, on precisely what *kind* of relationship they have, and on what role the demand for reliability plays in it. It depends on what ramifications it had, for Lila's afternoon or her house or her sense of self-worth. And it also depends on other facts in the context, about Jack, but also

48 Note that I think it is sufficient for forgiveness if it is partly out of goodwill that Lila *now does not* care about Jack's fault. It need not have been partly out of goodwill that she *originally ceased* caring. Suppose she originally ceased caring about it out of contemptuous disregard. But she later overcomes that disregard, and *would* come to care about Jack's fault and resent him again, were it not for the goodwill she now feels toward him in response to his sincere apology. In an important sense, I think, Lila then forgives him. Thus the process of "overcoming" resentment that constitutes forgiveness need not end once Lila lacks any resentment. It ends once she lacks it for the right reasons—i.e., once she does not care partly out of goodwill. (To that extent I am convinced by the arguments in Schönherr, "When Forgiveness Comes Easy.")

about Lila's mother or even the Philippines. If these factors play out right, they too can make Jack's fault less significant for Lila, or allow her to fittingly not feel resentment in response to it. And some of them—such as his general cheerfulness or undeserved bad health—may also make it fitting to feel goodwill toward him, and thus allow her to fittingly forgive. So sometimes, Lila may fittingly forgive even if the promise breaking is expressive of Jack's character. And sometimes she may still fittingly resent him even though it is not. Being sensitive to all facets of significance, blame and forgiveness are complex responses indeed.⁴⁹

7. CONCLUSION

Let me wrap up. I have introduced a version of the standard account of forgiveness. I have explicated a notion of the significance of value facts, provided an account of when such facts are significant for us in terms of their grounds, and outlined what this normatively implies. And I have argued that an account of forgiveness based on this phenomenon dissolves the apparent paradox and satisfies a number of desiderata. The account is also intuitively plausible. If Lila is a forgiving person, she will know that people go wrong. But she will be disposed not to dwell on their faults more than necessary, to *distance* herself from them—a bit as if they had happened far off. She will care about the good in people. If Lila is an unforgiving person, she may know that people do admirable things, and may not see their faults as worse. But she will be disposed to dwell on these faults if she may.

Naturally, there are many open questions. Most importantly perhaps, I have raised a number of general issues—about the representational content of emotions, the grounds and shape of their agent-relativity, or the extent to and reasons for which we are required to feel them. Some of these topics seem underexplored. It would be important to elaborate more on them. Being a version of the

49 It might also help to compare the present account to Hieronymi, "Articulating an Uncompromising Forgiveness." Hieronymi thinks your resentment is warranted in response to an action that wronged you, for which the agent was responsible, and that still makes a threatening claim to the effect that you may be treated thus. The wrongdoer's apology will undo that claim, and thus make your resentment lose its "point" (548), or become inappropriate. To forgive, she suggests, is to retract your resentment in response to that. Hieronymi does not say so. But perhaps one may interpret that threatening claim as making the fault *significant* for you. If so, then besides explicitly understanding the relevant issue as the general category of significance, there are again three main ways in which the present proposal goes beyond hers. Most importantly, it accounts for the complexity of forgiveness (and especially for forgiving the unapologetic) by allowing that a fault's significance depends on much more than its posing a threat. Partly in virtue of this, it can (and does explicitly) account for our discretion in forgiving. And at the same time, it elucidates why not all ways of overcoming resentment (even in response to apology) must amount to forgiveness.

standard account, the proposal also raises familiar questions about forgiveness. Forgiveness is something that usually has *normative* effects—e.g., concerning the wrongdoer's duty toward the victim to repent.⁵⁰ It is something that some people (notably the victim) seem in a privileged position to offer.⁵¹ And it is something we sometimes do with a goal.⁵² An account of forgiveness must explain these facts. I am confident that the present proposal has the resources to do so. But these are matters for another paper.

It would also be interesting to explore further dimensions of significance. For instance, presumably the significance of a value fact matters not just for attitudes, but also for *action*. We have stronger reason to act in response to a value fact the more significant it is.⁵³ Also, presumably our freedom not to care extends to other emotions and beyond people's faults, onto facts of life more generally—pains and losses and defeats. Thus there will be ways of “forgiving life.” Some people get habitually worked up about the foul and ugly and deplorable things, and stand aloof from what is worthy of gratitude or joy. And some people are just the opposite. And the difference seems important. In some instances, and as with forgiveness proper, a heightened concern with the good might amount to a dubious blindness to reality. But then again, if done in the right manner, it might also constitute an innocuous form of positivity and a reliable path to contentment.⁵⁴

University of Zurich
stefan.riedener@philos.uzh.ch

- 50 See, e.g., Warmke, “The Normative Significance of Forgiveness”; or Bennett, “The Alteration Thesis.”
- 51 See, e.g., Murphy and Hampton, *Forgiveness and Mercy*, 32; Walker, “Third Parties and the Social Scaffolding of Forgiveness”; also Pettigrove, “The Standing to Forgive”; and Radzik, “Moral Bystanders and the Virtue of Forgiveness.”
- 52 See, e.g., Hieronymi, “Articulating an Uncompromising Forgiveness”; and Fricker, “Forgiveness.”
- 53 There is a straightforward link between attitudes and action if we regard your *motivation* to act in accordance with the reasons (for action) provided by a value fact as a fitting attitude toward it. Then the more significant a value fact is for you the more motivation to act on these reasons it is fitting for you to have. That the suffering of a friend is more significant for you than that of a stranger, as something bad, say, does not just mean it is fitting for you to respond with more sorrow. It also means it is fitting for you to be more motivated to help them. This might explain many of our intuitions of partiality.
- 54 I thank Barbara Bleisch, Susanne Boshammer, Samuel Hughes, Felix Koch, Benjamin Lange, Adam Lovett, Lukas Naegeli, Claire Plassard, Peter Schaber, Philipp Reichling, Felix Timmermann, R. Jay Wallace, two anonymous reviewers and an editor from the *Journal of Ethics and Social Philosophy*, and audiences in Zurich, Berne, Cologne, and Osnabrück for very helpful comments on earlier versions of this paper.

REFERENCES

- Allais, Lucy. "Elective Forgiveness." *International Journal of Philosophical Studies* 21, no. 5 (2013): 637–53.
- . "Wiping the Slate Clean: The Heart of Forgiveness." *Philosophy and Public Affairs* 36, no. 1 (Winter 2008): 33–68.
- Anderson, Pamela Sue. "When Justice and Forgiveness Come Apart: A Feminist Perspective on Restorative Justice and Intimate Violence." *Oxford Journal of Law and Religion* 5, no. 1 (February 2016): 113–34.
- Bennett, Christopher. "The Alteration Thesis: Forgiveness as a Normative Power." *Philosophy and Public Affairs* 46, no. 2 (Spring 2018): 207–33.
- Blustein, Jeffrey M. *Forgiveness and Remembrance: Remembering Wrongdoing in Personal and Public Life*. Oxford: Oxford University Press, 2014.
- Broome, John. "Is Incommensurability Vagueness?" In Chang, *Incommensurability, Incomparability and Practical Reason*, 67–89.
- Butler, Joseph. *Fifteen Sermons Preached at the Rolls Chapel*. London: J. and J. Knapton, 1726.
- Calhoun, Cheshire. "Changing One's Heart." *Ethics* 103, no. 1 (October 1992): 76–96.
- Callard, Agnes. "The Reason to Be Angry Forever." In *The Moral Psychology of Anger*, edited by Myisha Cherry and Owen Flanagan, 123–38. London: Rowman and Littlefield, 2017.
- Chang, Ruth, ed. *Incommensurability, Incomparability and Practical Reason*. Cambridge, MA: Harvard University Press, 1997.
- . "Introduction." In Chang, *Incommensurability, Incomparability and Practical Reason*, 1–34.
- Couto, Alexandra. "Reactive Attitudes, Forgiveness, and the Second-Person Standpoint." *Ethical Theory and Moral Practice* 19, no. 5 (November 2016): 1309–23.
- D'Arms, Justin, and Daniel Jacobson. "The Moralistic Fallacy: On the 'Appropriateness' of Emotions." *Philosophy and Phenomenological Research* 61, no. 1 (July 2000): 65–90.
- Darwall, Stephen. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press, 2006.
- Deonna, Julien A., and Fabrice Teroni. "Emotions as Attitudes." *Dialectica* 69, no. 3 (September 2015): 293–311.
- Dubra, Juan, Fabio Maccheroni, and Efe A. Ok. "Expected Utility Theory without the Completeness Axiom." *Journal of Economic Theory* 115, no. 1 (March 2004): 118–33.

- Fricker, Miranda. "Forgiveness: An Ordered Pluralism." *Australasian Philosophical Review* 3, no. 3 (2019): 241–60.
- . "What's the Point of Blame? A Paradigm Based Explanation." *Noûs* 50, no. 1 (March 2016): 165–83.
- Garrard, Eve, and David McNaughton. "In Defence of Unconditional Forgiveness." *Proceedings of the Aristotelian Society* 103, no. 1 (2003): 39–60.
- Gert, Joshua. "Requiring and Justifying: Two Dimensions of Normative Strength." *Erkenntnis* 59, no. 1 (July 2003): 5–36.
- Gertken, Jan, and Benjamin Kiesewetter. "The Right and the Wrong Kind of Reasons." *Philosophy Compass* 12, no. 5 (May 2017): e12412.
- Goldie, Peter. *The Emotions: A Philosophical Exploration*. London: Oxford University Press, 2000.
- Greenspan, Patricia. "Making Room for Options: Moral Reasons, Imperfect Duties, and Choice." *Social Philosophy and Policy* 27, no. 2 (July 2010): 181–205.
- Griswold, Charles L. *Forgiveness*. Cambridge: Cambridge University Press, 2007.
- Helm, Bennett W. *Emotional Reason: Deliberation, Motivation, and the Nature of Value*. Cambridge: Cambridge University Press, 2009.
- Hieronymi, Pamela. "Articulating an Uncompromising Forgiveness." *Philosophy and Phenomenological Research* 62, no. 3 (May 2001): 529–55.
- Holmgren, Margaret R. "Forgiveness and the Intrinsic Value of Persons." *American Philosophical Quarterly* 30, no. 4 (October 1993): 341–52.
- Howard, Christopher. "Fittingness." *Philosophy Compass* 13, no. 11 (November 2018): e12542.
- Hughes, Paul M. "What Is Involved in Forgiving?" *Journal of Value Inquiry* 27, no. 3 (December 1993): 331–40.
- Hurka, Thomas. *Virtue, Vice, and Value*. New York: Oxford University Press, 2001.
- Hutto, Daniel D. "Truly Enactive Emotion." *Emotion Review* 4, no. 2 (April 2012): 176–81.
- Kelly, Thomas. "The Rationality of Belief and Some Other Propositional Attitudes." *Philosophical Studies* 110, no. 2 (August 2002): 163–96.
- Kolnai, Aurel. "Forgiveness." *Proceedings of the Aristotelian Society* 74, no. 1 (June 1974): 91–106.
- Kolodny, Niko. "Which Relationships Justify Partiality? The Case of Parents and Children." *Philosophy and Public Affairs* 38, no. 1 (Winter 2010): 37–75.
- Lord, Errol. "Justifying Partiality." *Ethical Theory and Moral Practice* 19, no. 3 (June 2016): 569–90.
- Maguire, Barry. "There Are No Reasons for Affective Attitudes." *Mind* 127, no. 507 (July 2018): 779–805.

- Milam, Per-Erik. "Against Elective Forgiveness." *Ethical Theory and Moral Practice* 21, no. 3 (June 2018): 569–84.
- . "Reasons to Forgive." *Analysis* 79, no. 2 (April 2019): 242–51.
- Murphy, Jeffrie, and Jean Hampton. *Forgiveness and Mercy*. Cambridge: Cambridge University Press, 1988.
- Na'aman, Oded. "The Fitting Resolution of Anger." *Philosophical Studies* 177, no. 8 (August 2020): 2417–30.
- Neblett, William R. "Forgiveness and Ideals." *Mind* 83, no. 330 (April 1974): 269–75.
- Novitz, David. "Forgiveness and Self-Respect." *Philosophy and Phenomenological Research* 58, no. 2 (June 1998): 299–315.
- Nussbaum, Martha C. *Upheavals of Thought: The Intelligence of Emotions*. Cambridge: Cambridge University Press, 2001.
- Parfit, Derek. *On What Matters*. Vol. 1. Oxford: Oxford University Press, 2011.
- Pettigrove, Glen. *Forgiveness and Love*. Oxford: Oxford University Press, 2012.
- . "The Standing to Forgive." *The Monist* 92, no. 4 (October 2009): 583–603.
- Portmore, Douglas W. "Imperfect Reasons and Rational Options." *Noûs* 46, no. 1 (March 2012): 24–60.
- Prinz, Jesse. *Gut Reactions: A Perceptual Theory of Emotion*. Oxford: Oxford University Press, 2006.
- Rabinowicz, Wlodek. "Value Relations Revisited." *Economics and Philosophy* 28, no. 2 (July 2012): 133–64.
- Rabinowicz, Wlodek, and Toni Rønnow-Rasmussen. "The Strike of the Demon: On Fitting Pro-Attitudes and Value." *Ethics* 114, no. 3 (April 2004): 391–423.
- Radzik, Linda. "Moral Bystanders and the Virtue of Forgiveness." In *Forgiveness in Perspective*, edited by Christopher R. Allers and Marieke Smit, 69–88. New York: Rodopi, 2010.
- Raz, Joseph. *Engaging Reason*. New York: Oxford University Press, 1999.
- . "Incommensurability and Agency." In Chang, *Incommensurability, Incomparability and Practical Reason*, 110–28.
- . *The Morality of Freedom*. Oxford: Clarendon, 1986.
- Richards, Norvin. "Forgiveness." *Ethics* 99, no. 1 (October 1988): 77–97.
- Riedener, Stefan. "Don't Make a Fetish of Faults: A Vindication of Moral Luck." *Philosophical Studies* 178, no. 3 (March 2021): 693–711.
- Rosen, Gideon. "The Alethic Conception of Moral Responsibility." In *The Nature of Moral Responsibility: New Essays*, edited by Randolph Clarke, Michael McKenna, and Angela M. Smith, 65–88. Oxford: Oxford University Press, 2015.

- Scanlon, T.M. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA: Harvard University Press, 2008.
- Schönherr, Julius. "When Forgiveness Comes Easy." *Journal of Value Inquiry* 53, no. 4 (December 2019): 513–28.
- Sharadin, Nathaniel. "Reasons Wrong and Right." *Pacific Philosophical Quarterly* 97, no. 3 (September 2016): 371–99.
- Smith, Angela M. "Moral Blame and Moral Protest." In *Blame: Its Nature and Norms*, edited by D. Justin Coates and Neal A. Tognazzini, 27–48. New York: Oxford University Press, 2013.
- Solomon, Robert C. "Emotions and Choice." *Review of Metaphysics* 27, no. 1 (September 1973): 20–41.
- Strawson, P.F. "Freedom and Resentment." 1962. In *Freedom and Resentment and Other Essays*, 1–28. Abingdon, UK: Routledge, 2008.
- Tappolet, Christine. *Emotions, Values, and Agency*. Oxford: Oxford University Press, 2016.
- . "Values and Emotions: Neo-Sentimentalism's Prospects." In *Morality and the Emotions*, edited by Carla Bagnoli, 117–34. Oxford: Oxford University Press, 2011.
- Walker, Margaret Urban. "Third Parties and the Social Scaffolding of Forgiveness." *Journal of Religious Ethics* 41, no. 3 (September 2013): 495–512.
- Ware, Owen. "Forgiveness and Respect for Persons." *American Philosophical Quarterly* 51, no. 3 (July 2014): 247–59.
- Warmke, Brandon. "The Normative Significance of Forgiveness." *Australasian Journal of Philosophy* 94, no. 4 (2016): 687–703.
- Watson, Gary. "Standing in Judgment." In *Blame: Its Nature and Norms*, edited by D. Justin Coates and Neal A. Tognazzini, 282–301. New York: Oxford University Press, 2013.
- Way, Jonathan. "Transmission and the Wrong Kind of Reason." *Ethics* 122, no. 3 (April 2012): 489–515.
- Zaibert, Leo. "The Paradox of Forgiveness." *Journal of Moral Philosophy* 6, no. 3 (January 2009): 365–93.

POSITIVE RIGHTS TWO-PERSON CASES

David Alm

THIS PAPER is about positive rights, meaning rights to be saved from harm. The topic is large, and I will therefore restrict myself entirely to the simplest type of case in which we appeal to such rights, leaving more complex ones for later treatment. I will be concerned with what I call “two-person cases.” In such a case, there is just one agent and one patient; the patient is in dire need and can be saved by the agent, and the agent alone, and there is no one else the agent could save. For most of the paper, I will also restrict myself still further, to what we can call “easy-rescue cases,” in which the cost to the agent of aiding is small.

My question, broadly speaking, is that of why patients in two-person cases have a right to the agent’s aid, given a view of rights, positive and negative, on which rights are not exhaustively characterized as corresponding to (first-order) reasons for action, but consist rather in one person’s having the authority or the standing to make demands of another (and more on that later). More narrowly, though, I will concentrate on one central aspect of the broader question, as will become clearer below.

I will proceed as follows. First, and as necessary background, I will say something about *negative* rights (to noninterference) and how positive rights differ from these—and also, in light of those differences, why we should be talking about positive rights in the first place. In section II, I turn to two-person cases and distinguish two questions that we would eventually have to answer about them: (1) Why does the agent in such a case lack a negative right he would normally have? and (2) Why does the patient have a positive right he would not normally have? I also offer a quick (and admittedly inconclusive) answer to question 2. However, question 1 is particularly challenging on the view of rights suggested above, and therefore I devote the balance of the paper to answering it. (Perhaps ironically, then, the lion’s share of this paper nominally concerned with positive rights is really about negative rights.) In section III, I sketch an answer to question 1, building on the idea that the kind of authority involved in having

a negative right requires being able to demand noninterference “decently,” and I offer some clarifications. In section IV, I continue my exposé of this proposal by considering the relevant notion of a “demand” a bit more closely and asking what we might mean in talking about an “indecent demand.” Then, in section V, I attempt to explain why the ability to demand decently is necessary for authority. Before concluding, I turn in section VI, if only briefly, away from easy-rescue cases to consider some other two-person cases.

I

As will become clear shortly, understanding positive rights requires understanding negative rights, and I will therefore begin by outlining my view of the latter in a rough-and-ready way. I have elaborated and defended it at length elsewhere, and so refer readers to that discussion if they wish further details.¹ On my view, to have a negative right is to have a certain type of *authority* over some other person, or persons generally, with respect to some type of interfering action. This authority consists in its being *up to* the right holder whether a given person is under a duty to refrain from performing that action, meaning both that a right holder is in some sense entitled or licensed to *demand* that others refrain from interfering in certain ways and that he is able to waive this duty. By the same token, a demand will be invalid or illegitimate if the demander lacks the authority to make it, meaning that it does not answer to any right. It is precisely this authority that yields the characteristic structure of a duty corresponding to a negative right—namely, that the duty bearer has *exclusionary* reasons not to act for any (or at least some) reasons for not doing his duty.² Again, note that it is up to the right holder whether these exclusionary reasons exist: that is what his authority consists in. To violate a negative right is to set aside the right holder’s authority, to act as if it did not exist and it is up to oneself to weigh the reasons for and against the relevant action. It follows, I hold, that duties corresponding to negative rights are not simply (strong) reasons for (in)action, and also that whether a negative right is violated does not simply depend on the duty bearer’s balance of reasons.³ It is not as if the fact that an action amounts to setting aside another’s authority is itself an overriding reason against that action, or indeed a reason at all, on a par (say) with the fact that the action will cause harm. It also follows that negative rights cannot conflict with other rights, strictly speaking,

1 Alm, *Moral Rights and Their Grounds*.

2 The idea of exclusionary reasons comes from Raz, *The Morality of Freedom and Practical Reasons and Norms*. See especially *Practical Reasons and Norms*, 39.

3 The view of right violation rejected here is essentially that of Thomson, *The Realm of Rights*.

as it is impossible for both parties to such a supposed conflict to have authority over the action at issue.

I admit that a view of negative rights like the one just described will likely make one ask why we should speak of positive “rights” at all, for they are not rights in the way I have said that negative rights are rights. For one thing, talk of “authority” and “demands”—or, relatedly, of the right holder’s being “sovereign” over certain matters—fits negative rights against interference much better than it does positive ones to aid.⁴ Indeed, the problem here goes deeper than mere language, for arguably an agent able to aid some patient does not, in virtue of that fact alone, have exclusionary reasons not to act for reasons other than that particular patient’s need—as opposed to simply having reasons for aiding him.

Even so, it seems to me that it still makes sense to speak of positive rights, if only in a possibly narrow range of cases, on account of several features that such rights share with negative ones. Here it will suffice to mention three. In the first place, the requirement one is under when in a position to aid another will, in some cases, be *personal* (as opposed to impersonal). It is a kind of relation between agent and patient.⁵ I will not attempt to specify this relation exhaustively, but at least it would involve the patient’s having reason to resent the agent for failing to aid and the agent’s having reason to apologize to the patient (or to those close to him) for such failure, and perhaps to make amends in other ways. A second point, possibly a consequence of the first, is that a purported positive right can be *waived*. Third, positive rights can be *enforced*, at least to some extent, meaning that the right holder is permitted to do certain things that would normally be disallowed if they are necessary and sufficient to induce the duty bearer to act. As we will see below, this feature in particular is essential to explaining why it makes some sense to speak of the patient as “demanding” the agent’s help, as opposed to merely “requesting” it (though such talk may admittedly still seem a stretch to some).

II

Having said a little about positive rights generally, I turn to what I have called “two-person cases.” It will prove useful here to work with a concrete instance, and

4 For the phrase “sovereign,” see Hart, *Essays on Bentham*, 183.

5 Some philosophers reject positive rights to aid on the grounds that (1) all rights require a nonarbitrary specification of the corresponding duty bearer, and (2) such specification is impossible for positive rights (unless they are institutional). See, e.g., O’Neill, *Towards Justice and Virtue*, 131–34. I dodge that issue by sticking to two-person cases, in which a unique duty bearer can be identified unproblematically.

I will be referring frequently to it and variants of it. Consider, then, the following situation: Bob is drowning, through no fault of his own, and Betty, a complete stranger, can (with certainty) save him by throwing him a public life preserver located right next to her; there is no one else she could save from harm in the situation; and there is no one else around who could save Bob. Call this case *Drowning 1*.

It will also be helpful to make a pair of stipulations about this case. The first stipulation is that Betty's decision (i.e., that of whether to throw the life preserver) is such that she typically, or by default, has a right against interference with it.⁶ In other words, unless special circumstances obtain, forcing her to perform that action would wrong her (and would not merely be wrong)—and, we might add, she would wrong no one by not performing it. This right includes also a right to use a moderate amount of force to prevent others from interfering with her decision.

The second stipulation is that special circumstances *do* obtain, because if Betty does not throw Bob the life preserver, he will die, and so he has a positive right against Betty that she do just that. As we have seen, it is awkward to speak here of Bob's having "authority" over Betty—and indeed potentially misleading, if we tend to link that notion with exclusionary reasons. Therefore, I will say only that Bob has the "standing" to demand that Betty make a decision she would normally be authorized to make as she sees fit. As a consequence, Betty lacks the authority over Bob, in this matter, which she would normally have. As we saw in the preceding section, this second stipulation also entails that Bob's right has the three features listed there as characteristic of positive rights. That is, (1) the features of the case suffice to yield a personal moral relation of the kind described between Bob and Betty, (2) Bob can waive his right, and (3) Bob is allowed to use some measure of (normally impermissible) force to make Betty save him. (Specifying means of enforcement that might plausibly be at Bob's disposal is cordially left to the reader's ingenuity.)

Given these stipulations and the view of negative rights outlined in the preceding section, we need to answer two questions about *Drowning 1*. The first is why Betty (the agent) to some extent lacks authority that she would otherwise have over a certain action (i.e., throwing the life preserver). That is, why does she lack her "default" negative right? The second is why Bob (the patient) acquires his standing to make demands on Betty. In other words, why does Bob have a positive right to her aid? Betty's lack of a right, of course, entails only that Bob has a mere liberty to make her help him, compatible with her own liberty to refuse to help, and to resist his efforts to make her do so.

6 On the notion of "default" rights, see Walen, "The Restricting Claims Principle Revisited."

My main concern in this paper is with the first of these two questions. That is in part because I have addressed the second question elsewhere—if admittedly incompletely—but also because the first seems to be especially pressing on a view of negative rights like the one outlined in the preceding section.⁷ Indeed, on the alternative view that rights are simply the correlatives of strong reasons to act, the first question hardly arises at all, as that view does not make use of the notion of authority in the first place. By contrast, the alternative view would still have to explain why Bob has a right to Betty's aid, whether or not that right is understood in terms of what I have called "standing." For present purposes, I will make do with the following very sketchy, and inevitably insufficient, remarks in answer to the second question, again referring readers to *Moral Rights and Their Grounds* for a more complete treatment. It is in virtue of their *agency* that persons have rights at all, and this fact is reflected in the content of these rights and also in their being subject to their holders' control (by waiver). In particular, persons have rights to what they need to function adequately as agents, which may include various resources. The fact that such a right presupposes a need, even if it cannot simply be reduced to one, explains the limited nature of the standing to make demands, in contrast to the authority constitutive of negative rights, which does not presuppose needs. There are two important points of difference. First, if I possess a negative right, I could legitimately demand that you *not* do something over which I have authority, even if your action would in fact meet important needs of mine and would be of net benefit to me. In the most familiar case, I could refuse needed medical treatment. But, at least barring a special obligation, I could not demand that you (positively) act unless your action would actually meet one of my needs. To the extent that I have standing to demand that you act, then, that standing requires a genuine need. Second, a positive right's presupposing a need of a certain strength also explains why it does not correspond to exclusionary reasons—or so I would argue. To put the point simply, if a little misleadingly, the right inherits this strength as an essential feature and so the right could not exclude stronger competing reasons.

The above will have to suffice here in answer to the second question, and so I turn to the first.

III

We have noted that in Drowning 1 Betty lacks the authority to demand of Bob that he not interfere with her decision not to throw the life preserver. Now, if we ask *why* that is so, we are bound to alight on the huge difference in what is *at*

⁷ See Alm, *Moral Rights and Their Grounds*, ch. 6.

stake for Bob, relative to Betty. After all, were that difference significantly smaller, Betty would retain her authority (see section VI). But, we might ask, why is this difference in stakes so important? Why does Betty lack her authority to make demands of Bob when he has much more at stake than she does, but not otherwise?

It is not an adequate answer to the question to say simply that Betty lacks authority in Drowning 1 because her stake is so much smaller than Bob's. This is for two reasons. In the first place, that answer renders unclear the relation between rights and the balance of reasons. The large difference in stakes might certainly explain why Betty's reasons for resisting Bob's interference are weaker than her reasons for not doing so and her reasons for aiding him, but that is not the same as her lacking a right to noninterference. Second, and no doubt related, it is unclear how an appeal to relative stakes could explain why *Betty's* authority is lacking, as it is as much a fact about *Bob* as it is about Betty.

We can see a more promising answer, I believe, if we attend to a feature of situations like Drowning 1 that is occasionally noted in the literature.⁸ For there would (typically) be something *morally objectionable* about Betty's exercising her purported authority in the way we are currently considering—that is, by demanding that Bob not interfere with her decision not to throw the life preserver. We might say that such conduct is “contemptible” or “indecent” or “selfish” or “disgraceful,” precisely because of the huge difference in stakes between Betty and Bob.⁹ Here, it seems, we have a factor that can explain why Betty's authority is lacking in Drowning 1: she lacks it because she could not use it (or try to use it) in the circumstances without being subject to some sort of moral opprobrium. Note that we are then pointing to a moral fact about Betty, the purported right holder, and not merely to a fact about relative stakes—though the latter matters indirectly—and thus a fact that seems to be of the right kind to explain her lack of authority.

Turning the proposed explanation into an account of authority, we get something like the following, as a first approximation: one has the authority to demand noninterference just in case one can do so decently. To throw some light on this proposal, it is useful to begin by comparing it to a similar one defended by Skorupski. On his view (in simplified form), one has a (moral) right if and only if it is “prima facie morally permissible” for one to demand its fulfillment.¹⁰ Our proposals are similar above all in that both are defined precisely in terms of the moral assessment of *demands*. Consider Drowning 1. On both views, Betty

8 For examples, see Taurek, “Should the Numbers Count?” 308; Quinn, *Morality and Action*, 169n34; and Voorhoeve, “How Should We Aggregate Competing Claims?” 75.

9 However, I would not say that, as a rule, Betty's demand *wrongs* Bob. I thank an anonymous referee for raising this issue.

10 Skorupski, *The Domain of Reasons*, 311–12.

lacks the right against Bob that he not interfere with her decision not to throw the life preserver because her demand that he not interfere would necessarily be morally objectionable, but the relevant objection is precisely to her *demand* that Bob not interfere with her decision, rather than to the decision itself. To be sure, her decision is also objectionable, and it may be that her demand would not have been objectionable otherwise—after all, the demand (for noninterference with her decision) logically presupposes the decision. Even so, our concern is with the demand. It is worth noting that this feature of the view means that it is compatible with a “right to do wrong.”¹¹ I grant that some action that a person has a right to do could be such that he could not decently choose to perform it (meaning that it is, in one sense, an exercise of the right to do wrong). Yet it does not follow that it must also be objectionable for him to demand that others not interfere forcibly with this action—and it is the latter judgment that matters to whether he has a right against such interference.

However, there are also some significant differences between my proposal and Skorupski’s, a couple of which can be noted here. In the first place, and as indicated in section I, I do not endorse the view of justifiably overridden (negative) rights that goes with Skorupski’s formulation. For that reason, I would also dispense with the qualifier “*prima facie*.” Second, I would hesitate to speak of “permissibility”; that is why I have used terms such as “indecent” and “selfish” (or the somewhat nebulous “objectionable”). I will say more about this last point in sections IV and V. In section IV, I will also describe a further important difference between the two accounts.

I end with a caveat. The account I presented above concerns authority, and so negative rights, exclusively. The same is arguably true of Skorupski’s account as well, as he does not seem to recognize rights to aid based simply on need.¹² It is possible to extend the suggested account also to positive rights—saying, presumably, that a person has a right to demand aid just in case he could do so decently. I do not think, however, that such an account would be plausible: while the condition stated may be necessary, it is hardly sufficient. I am further inclined to believe that we would also eventually have to modify the suggested account of authority (and, arguably, Skorupski’s account of rights) in light of these worries (and perhaps others). However, as these matters are not crucial for the argument to follow, I will set them aside here. At least as long as we limit our attention to two-person cases and consider only the information available in them, I believe that the proposed account is defensible—once properly understood.

11 See Waldron, “A Right to Do Wrong.”

12 See Skorupski, *The Domain of Reasons*, 312–13.

IV

I have made heavy use of the notion of demanding, and it is time now to consider it more carefully. As a preliminary, let me signal a restriction of scope: I am concerned exclusively with demands of the type made in Drowning 1 and similar cases; that is, demands that do not presuppose any kind of institutional background. Indeed, I do not pretend to offer a complete analysis even of “pre-institutional” demanding. This is because my main interest is really in the separate question of what would make such a demand morally objectionable (or selfish, indecent, or contemptible).

Turning to that question, I note first that it is useful to contrast a demand with a (mere) *request*. After all, even if a demand, such as Betty’s in Drowning 1, is objectionable, the same is typically not true, or at least not nearly to the same degree, of a request with the same content. The key difference is that demands are linked somehow to *enforcement*, as defined in section I, but requests are not. Unfortunately, there is no agreement in the literature about what role, if any, enforcement plays in an account of demanding.¹³ However, considering the purpose for which we appeal to the notion of a demand—that is, to make sense of the notion of a right, which is clearly linked to enforcement—I choose to embrace the following necessary condition: a demander must be willing to enforce his demand somehow, at least if he could do so at little cost to himself and third parties. It should be clear that such willingness matters to our moral evaluation of the relevant demand.

Having identified the morally distinctive features of demands, we return to our question of what makes a demand morally objectionable. Unfortunately, there has not been much discussion of that question.¹⁴ If we are to answer it, I believe, we must first realize that we could mean two rather different things when we say that a demand is morally objectionable. As we will also see, one of these interpretations is decidedly more promising than the other as a part of the account of authority suggested in the preceding section. Let me address these two interpretations now.

First, though, I wish to add something that applies to both of them and makes them more precise. I have spoken about a demand’s being morally objectionable,

13 For contrasting answers, see, e.g., Guttchen, “Demands,” 121; and Schaber, “Demanding Something,” 68. On Skorupski’s view, “to demand is to imply that enforcement would, if necessary be permissible” (*The Domain of Reasons*, 310).

14 Skorupski is an exception. He holds that “a permissible demand is a request that it is morally permissible to enforce” (*The Domain of Reasons*, 310). There are problems with that claim that will become apparent below.

and about its being indecent, but it was implicit in section III that a demand is objectionable or indecent because it is *selfish*, and that is indeed what I will assume in the sequel. (I will consider a complication below.) And so we can state the first interpretation, which is also likely the one that will first come to mind. It is that in saying that a demand is selfish, we are attributing selfishness to an *action*—namely, the act of demanding. Further, what makes that act selfish is its being explained or motivated by some desire or attitude, also selfish. In saying that a selfish desire or attitude *explains* an action, I mean to say at least that the demand would not have been made if the demander's desire or attitude had not been selfish.¹⁵ Call the above the “act interpretation.” Is it adequate?

Now, we should remind ourselves why we are interested in the notion of a selfish demand to begin with: it is because of the suggestion that a person has the authority to demand something of another if and only if he can do so without being (objectionably) selfish. On a closer look, though, it becomes clear that this ability is neither necessary nor sufficient for having such authority if the notion of a selfish demand is understood along the lines of the act interpretation.¹⁶ To see that it is not necessary, consider the following variant of Drowning 1. Bob is in no serious danger yet still demands that Betty throw the life preserver. If she refuses, her demand that Bob not interfere with her decision, and that alone (including the willingness to enforce), will trigger a madman to kill several people (and she knows this). It seems impossible in these circumstances for Betty to make her demand without objectionable selfishness, as it would inevitably depend on her callous indifference to the madman's victims: had she cared more about them, she would neither have made nor enforced her demand. Yet surely her authority remains intact: it still seems that Bob is duty bound to respect her demand. After all, the madman cares not one whit about what *he* does.

To see that the ability to demand nonselfishly is not sufficient either, consider the following “inverse” of the madman case just presented: Bob is again in mortal danger, but Betty's refusal to aid and her demand for noninterference with that refusal, and only that (including the willingness to enforce), would spur an eccentric do-gooder to save several other lives. If this fact moves Betty to make her demand, it seems both pardonable and permissible, yet surely Bob is not required to heed it. Again, the do-gooder does not care about what *he* does.

To get around these difficulties, it seems to me, we need to shift our attention—

15 For simplicity, I will assume that the act of demanding and the willingness to enforce have the same explanation.

16 Counterexamples much like those I go on to discuss in the text would also seem to affect Skorupski's view. These examples admittedly stretch the definition of a two-person case. Still, an account that fits them does seem preferable to one that does not.

and, as a consequence, appeal to a different understanding of selfish demands.¹⁷ I have been asking whether a person is *able to demand something nonselfishly*. Again, the focus here is on a moral evaluation of the possible act of demanding itself. A different and, I believe, more appropriate focus is on *what is being demanded*—as indicated by our talking about “what could decently be demanded.” That is, we are not asking “Is this act of demanding selfish?” but rather “Is it selfish to demand this (rather than that)?” Once we ask the question in those terms, an answer suggests itself: what one could decently demand of another, and so what one has the authority to demand, is what one could decently *want* the other to do. (Note that it does not matter what one in fact wants the other to do, if anything.) Strictly speaking, then, the kind of selfishness relevant to authority belongs not to the act of demanding, but rather to the *desire for compliance* with that demand. Further, whether one could decently want another to comply depends on the stakes involved in the case. To clarify, the “desire” here should be understood as a preference relation: compliance is preferred over noncompliance. In Drowning 1, Betty’s desire for compliance means that she prefers Bob’s leaving her alone (leading to his drowning) to his interfering (and, let us assume, making her save him)—and hence, in practice, that she prefers her own convenience to Bob’s survival. Call what I have stated in this paragraph the “compliance interpretation” of the notion of a selfish demand. It dispenses with the counterexamples. After all, in the madman case, Betty could nonselfishly desire that Bob comply, even though her act of demanding would be selfish, and so she does have the authority to demand noninterference. By contrast, in the do-gooder case, Betty’s desire for Bob’s compliance would have to be selfish, even though her act of demanding is not, and so she does not have the authority to demand noninterference.¹⁸

Importantly, the compliance interpretation can make sense of the essential role of enforcement in accounting for the objectionableness of demands. This is because the demander may also prefer *enforced* compliance over noncompliance (and so prefer compliance no matter what the other person says), at least given that enforcement is necessary. Strictly speaking, it is this preference that matters to our evaluation. Again, it could be absent in a given case; what matters is whether it would be objectionably selfish if it did exist, and that this selfishness explains why the demand in question is one that could not decently be made

17 As advertised in section III, I here take a step further away from Skorupski’s view. In saying this, I effectively attribute an act interpretation to Skorupski. I will not pause to justify that attribution here, but one strong indicator is his treatment of the “coercion” case, discussed in note 19 below.

18 Note that it is futile to revise the counterexamples so that the madman and the do-gooder act contingently on Betty’s desire for compliance, rather than her act of demanding. This is because our evaluation of that desire does not depend on the consequences of having it.

in the circumstances. However, we can ignore the particular *means* of enforcement the demander is willing to use, even if resorting to such means would be objectionable, and even if no other means are available or would be effective in the situation. Again, the question we are asking is that of what the person could decently demand, and though the willingness to enforce *in some way* may be essential to a demand, as I have suggested, it is certainly not essential that one be willing to use whatever means of enforcement are necessary in the circumstances, no matter how harsh. Our question, then, is not about whether it would be objectionable to use whatever means of enforcement happen to be available or necessary in the context. It is rather about whether it would be objectionable to demand compliance *no matter which* method of enforcement one uses (assuming, of course, that the means chosen would normally be impermissible).

I should note an important complication. Offhand, Betty's demand that Bob not interfere could be both nonselfish and invalid, as long as she makes it on behalf of someone else—in the sense that her desire for compliance itself is motivated by a concern for that other person's welfare. In such a case, Betty need not prefer her own convenience to Bob's survival. For instance, she might fear that her saving Bob's life, or Bob's making her do so, will disturb Bernie's afternoon nap. In this case also, Bob would not wrong Betty by interfering. Now, cases like this one, involving three persons rather than two, need separate treatment, and here I can only make a couple of brief remarks pertinent to the worry just raised. In the first place, I would say, a person is typically authorized to make demands on behalf of himself alone. As a consequence, when we try to determine whether a given demand of his is valid, we are to ask whether it would be objectionably selfish of him to make the demand on his own behalf (meaning that he is motivated by a concern for his own welfare), whatever his actual motivation. Second, in the special case in which a person could indeed make a valid demand on behalf of another—perhaps due to the intimate relationship between them—the validity of such a demand depends on whether that other person could validly, and hence nonselfishly, make the same demand on his own behalf. In our variant case, Bernie could make no such demand of Bob, and so Betty's demand, on Bernie's behalf, is invalid (but not because it would be selfish or otherwise objectionable of Betty to make it).

I have addressed the question of what it means to say that a demand is selfish, distinguished two interpretations of that question, and tried to explain why one of these interpretations is more promising in the context of accounting for authority. I turn now to another question—namely, that of what it means to say that a demand, in the sense indicated, is (objectionably) selfish. Here I need make only one crucial point; namely, that the term “selfish” tends to imply *blame*—though there is of course no strict entailment—suggesting that the judgment

that a demand, and the corresponding desire for (enforced) compliance, is objectionably selfish is in the end a judgment of blameworthiness. The same goes for the other terms I have been using: “contemptible,” “indecent,” and “disgraceful.” (By contrast, Skorupski’s favored term “morally impermissible” does not carry the same implication, or at least does so less strongly.) I will return in the next section to the question of *why* I choose to put my account in these terms.

In the meantime, though, I need to consider some further objections against the compliance interpretation, reflecting precisely the idea that it is a mistake to make authority depend on blameworthiness—rather than on justification, and thus ultimately on the strength of competing reasons. These objections have to do with the fact that our judgments about blameworthiness are sensitive to the presence of various *excuses*, which do not seem relevant to authority. In one type of case, the desire for compliance is the result of manipulation or brainwashing, making us unwilling to blame the person for having it—although his authority presumably would be no more genuine for that.¹⁹ In a less far-fetched case, our demander is akratic. There may be cases in which, if he feels terrible about his selfish desire for compliance (and for acting on it), he would not be contemptible—though he would have been contemptible if he had not felt terrible about it. Yet surely he should be equally lacking in authority regardless. In dealing with cases like the ones just described, it is again important to remember that we are not concerned with assessing a particular act of demanding performed by some particular person in some particular set of circumstances. Rather, we are asking what one *could* decently demand of another in those circumstances, and hence what one *could* decently desire that person to do—meaning what *anyone* could decently desire in those circumstances, defined by the relative stakes of the persons involved and not by the peculiarities of any actual demander. The relevant case here, therefore, is one where the demander is acting with full responsibility—in his full capacity as agent, as it were—and hence we can simply ignore excusing conditions such as those just cited. The judgment we make remains one of blameworthiness, if admittedly a somewhat peculiar one.

The second type of problem concerns a different kind of excuse—namely, ignorance. Specifically, the compliance interpretation seemingly entails that a person

19 Skorupski considers a parallel problem case for his own account—and, incidentally, for the act interpretation—in which a person is *coerced* into making a demand (*The Domain of Reasons*, 312n27). (Skorupski modifies his account to deal with this case, but that modification does not seem able to accommodate the do-gooder case described in the text above.) Interestingly, Skorupski’s coercion case suggests that he, too, may ultimately be concerned with whether the agent is *blameworthy* for making his demand—despite stating his view in terms of *permissible* demands. After all, coercion is usually seen as an excuse rather than a justification. (See also section v below.)

may retain authority because he is (blamelessly) misinformed about the facts—but that cannot be right. Thus, in a version of Drowning 1, Betty's desire for compliance is not selfish or blameworthy because she falsely, but blamelessly, believes that Bob is not in grave danger. For all that, though, Bob does not wrong her in forcing her to aid. Now, the reason mentioned in the preceding paragraph for disregarding excuses may apply to the case of ignorance as well; but there is also another reason, peculiar to that case, and worthy of separate treatment. Unfortunately, I will have to be brief and dogmatic, as a full discussion would require a rather extensive foray into rights theory. On the view of rights I have presupposed, based on the related notions of authority and standing, rights have to be objective, in the sense that they do not depend on what their holders, or other people, believe. If not, we could find ourselves forced to acknowledge both that Betty has the right to defend herself against Bob's interference (because of what she believes) and that Bob has the right to interfere (because of what *he* believes). To avoid that result—which I take to be impossible—we must recognize that our moral judgments about a person's demands are only relevant to that person's authority (or standing) on the assumption that he is correctly informed about the situation, including the stakes involved. As a consequence, we cannot infer that Betty's demand is valid in the above variant of Drowning 1, and the problem I noted does not arise.

So much for counterexamples. In the next section, I will return to the general question of why we should understand authority in terms of blameworthiness.

v

And so we face the question: Why does selfishness undermine a person's authority to make demands? Actually, I will ask a related question: Why does selfishness render a demand invalid? I will do this because I believe that, on the view of selfish demands now being presupposed, and restricting our attention to two-person cases, a person who makes such a demand could not in the situation have made that same demand nonselfishly—for selfishness is a matter of relative stakes, which we assume are essential to the situation. Hence, the person lacks the authority to make the demand in question. Granting that much, I hold that we can derive the desired conclusion (that selfishness undermines a person's authority to make demands) from the following two premises.

1. An objectionably selfish demand is not worthy of respect *qua* demand (indeed, it is worthy of disrespect).
2. A valid demand is worthy of respect, or at least not worthy of disrespect, *qua* demand.

From these two premises, it follows that an objectionably selfish demand is not valid. Let us now consider this simple argument.

The meaning of the premises should be fairly clear from the preceding section, but let me make it quite explicit here. In the first place, I use the term “respect” in such a way that a demand is worthy of respect, or at least not worthy of disrespect, just in case one is not blameworthy for making such a demand in the situation. Second, a demand is worthy of respect *qua demand* not in virtue of the demander’s actual contingent motive for making it or enforcing it, which could of course be practically anything, but rather in virtue of the *constitutive* aim of a demand.²⁰ This aim, at least in part, is compliance. Compliance is not a constitutive aim of demands in the strong sense that a speech act does not count as a demand unless the demander in fact prefers compliance to noncompliance. Yet that preference is at least constitutive of the *practice* of demanding in the sense that cases in which the demander does not prefer compliance—and indeed enforced compliance—to noncompliance will necessarily be exceptional. That is why, in judging whether a demand is worthy of respect *qua demand*, we naturally look to the desire for compliance, even if no such desire is actually present.

Given this understanding of the two premises, both seem quite plausible. A critic is likely to object, however, that this impression is deceptive, because the argument equivocates by conflating two ways in which a demand can be worthy of respect. On the one hand, a demand could be worthy of respect in more or less the way I have described (in terms of blameworthiness). On the other, it could be worthy of respect precisely because it is successful, meaning valid. There is equivocation here, the critic argues, because the first premise of the above argument (that an objectionably selfish demand is not worthy of respect) will seem indubitable only given the first interpretation, while the second premise (that a valid demand is worthy of respect) clearly presupposes the second interpretation.

Before responding to that charge, though, I wish also to state a second one, as I believe it is related to the first, and is effectively subject to the same treatment. This second charge is, perhaps, better thought of as aimed at the argument’s conclusion, rather than the argument itself. The idea is that the phrase “worthy of respect” is itself question begging in the present context: a demand is worthy of respect just in case we ought to respect it, which in turn means that it corresponds to a genuine right—or perhaps simply that the balance of reasons favors respecting it. Indeed, this criticism could be taken further, for it could be held that the judgment that a given person’s use of his (purported) authority is

20 For this notion, see, e.g., Lance and Kukla, “Leave the Gun; Take the Cannoli!”—though the view of the constitutive aim taken in the text is different from theirs (if not obviously incompatible with it).

“objectionably selfish” (or whatever other term we care to employ) presupposes an antecedent account of rights, and so cannot prop up just such an account. For instance, the critic has it, what makes it objectionably selfish of Betty to demand that Bob leave her alone in Drowning 1 is precisely that, for independent reasons, she lacks such a right in the circumstances, rather than the other way around.

I respond to both objections by flatly denying that we can tell whether a demand is valid or successful independently of its moral status. After all, we are not dealing with a performance with a causal effect, observable independently of its moral properties. At any rate, I would insist that some argument be provided for such independent facts, and in cases of the type that now concerns us, involving demands for noninterference in two-person cases, I can see no such argument. I conclude, then, that there is no obvious reason why we must resort to an independent fact about rights to draw the line between sufficiently and insufficiently morally objectionable demands. It seems at least as plausible to suppose, on the contrary, that we draw conclusions about rights in such contexts from independent judgments about what would be too much to ask of another, and thus about when he could decently demand to be left alone. While in making such judgments we are indeed concerned with a difference in stakes between the parties involved, we have no way of telling which difference is too great in a given context to support a valid demand, independently of our sense of whether one would be to blame for making that demand.²¹

Now, and as I have stressed repeatedly, I believe we should treat blameworthiness, rather than justification, as explanatory in the present context. In part, that is because it is much less clear otherwise that we could avoid the charge of begging the question just discussed.²² Yet my choice is not simply an *ad hoc* maneuver designed to circumvent circularity. On the contrary, we talk about blameworthiness in these contexts rather than simply about reasons for action (or justification) because we are asking precisely about the degree to which one may favor oneself, and it is hard to answer that question in terms of reasons for action. After all, are these reasons to be understood as impersonal or “for-me” reasons? To be sure, we may end up saying, in confronting a case like Drowning 1, that Betty ought to help Bob “all things considered” or suchlike, *and* that this fact in-

21 Cf. Quinn, *Morality and Action*, who endorses (but does not defend) the following proposal: “A negative right may be justifiably infringed just in case it would be contemptible of its possessor to insist on it,” stressing that this criterion “requires intuition to apply” (169n34).

22 For that reason, the “begging-the-question” objection might cause more trouble for Skrupski. On his view, after all, one has a right only if one is *morally permitted* to demand compliance, and demanding involves enforcement. It seems less plausible, at least arguably, that the notion of morally permissible enforcement is explanatory in this context compared to that of being objectionably selfish.

dicates that her impersonal reasons for preventing Bob's death are "stronger than" her self-regarding reasons for avoiding her own inconvenience—but it does not follow that we arrive at the former conclusion by way of the latter, understood as an independently intelligible weighing of reasons, rather than by way of a judgment about what a person could decently ask that we leave for him to decide.

The conclusion to draw, then, is that the argument outlined at the start of the section does not equivocate, and therefore gives support to our conclusion, that selfishness undermines a person's authority to make demands.

VI

After this tortuous analysis of Drowning 1, I wish to consider, much more briefly, a few other two-person cases, variants of the original. Consider first a *hard-rescue* case in which the cost to the agent of saving the patient is substantial. Specifically, in *Drowning 2* Betty is only able to save Bob from drowning by heroically swimming across ice-cold, stormy, and polar bear-infested waters, risking her own life. In light of our earlier analysis of Drowning 1, we need to make two observations about this new case. First, and obviously, it would not be objectionably selfish of *Betty* to refuse to aid Bob in this case, nor to demand that Bob not interfere with her decision. That is enough to explain why she retains her right against Bob not to force her to aid him in this case. Second, we may or may not hold that it would be objectionably selfish of *Bob* to demand Betty's aid at such great risk to herself. However, for present purposes it does not matter in itself whether that is so (as noted in section III). What matters is that Betty's demand is not selfish, and hence that her authority is intact, and so Bob could not have standing to demand her aid. In making that judgment, we would of course have to rely on a comparative assessment of what is at stake for Bob and Betty, respectively, just as we would in judging whether Bob's demand would be selfish. But even so, it is the judgment about Betty (the supposed right holder), rather than the one about Bob (the supposed duty bearer), that matters.

In yet a third case—*Drowning 3*—Betty is still able to save Bob at little cost to herself, but Bob's enforcement would be quite harmful to Betty.²³ In this case, if the harm that Bob's enforcement would impose on Betty is grave enough, she would not count as selfish in demanding that he desist, and she therefore retains her right. To be sure, she could have avoided putting herself in this predicament to begin with if she had not selfishly refused to aid Bob. Yet what is decisive here, as usual, is whether her *demand* is (or would be) objectionably selfish, and in determining

23 To avoid irrelevant complications, let us grant that, for whatever reason, Bob has no other means at his disposal.

that, we must also consider the severity of the enforcement, or interference, it is intended to forestall (if need be with counterforce). As that enforcement would, in Drowning 3, cause Betty significant harm, it is not selfish of her to demand that Bob not enforce. That is compatible with her conduct being blameworthy overall.

Our final case, *Drowning 4*, is in a way intermediate between the two preceding ones. In it, the cost to Betty of saving Bob is significant, but not severe enough by itself to shield her from the charge of selfishness were she to refuse to aid. The same is true of Bob's enforcement. The question now arises of whether the *combined* cost of enforcement and action could be enough to avert the selfishness charge, even if neither cost suffices on its own. After all, Betty would suffer both costs if Bob makes her save him by enforcing his demand (and he may have to do this to survive). I am inclined to answer that question in the negative: for Betty to escape the charge of selfishness, either cost on its own would have to be sufficiently great to absolve her. The aggregate cost does not matter. Unfortunately, I am uncertain about how to argue for this answer and have chosen to leave the matter unresolved.

VII

Summing up, I have argued that it is possible to account for the patient's positive right to aid in two-person cases on a view of such rights as the standing to make demands of others, rather than simply in terms of reasons for action. I have concentrated on one part of that account—namely, explaining, in a way that does not also presuppose a view of rights as reducible to reasons for action, why the agent lacks a negative right to noninterference that he would normally have. I have proposed just such an explanation, arguing that there are certain demands of noninterference that an agent could not validly make. While the proposal faces some significant hurdles, I have argued that these can be overcome. The next challenge is to extend it to cover more complicated cases involving positive rights, such as those in which the agent must choose between saving several persons and cases involving purported aggregation of positive rights. I aim to take on this challenge in subsequent work.²⁴

Lund University
 david.alm@fil.lu.se

²⁴ For helpful comments, I am indebted to participants at the Higher Seminar in Practical Philosophy at the University of Lund and an anonymous referee for the *Journal of Ethics and Social Philosophy*. Work on this paper was funded by a grant from the Swedish Research Council.

REFERENCES

- Alm, David. *Moral Rights and Their Grounds*. New York: Routledge, 2019.
- Guttchen, Robert. "Demands." *Journal of Value Inquiry* 7, no. 2 (June 1973): 121–28.
- Hart, H. L. A. *Essays on Bentham*. Oxford: Oxford University Press, 1982.
- Kamm, F. M. *Morality, Mortality*. Vol. 1, *Death and Whom to Save From It*. Oxford: Oxford University Press, 1993.
- Lance, Mark, and Rebecca Kukla. "Leave the Gun; Take the Cannoli! The Pragmatic Topography of Second-Person Calls." *Ethics* 123, no. 3 (April 2013): 455–78.
- O'Neill, Onora. *Towards Justice and Virtue*. Cambridge: Cambridge University Press, 1996.
- Quinn, Warren. *Morality and Action*. Cambridge: Cambridge University Press, 1991.
- Raz, Joseph. *The Morality of Freedom*. Oxford: Oxford University Press, 1986.
- . *Practical Reason and Norms*. Revised ed. Oxford: Oxford University Press, 1990.
- Schaber, Peter. "Demanding Something." *Grazer Philosophische Studien* 90, no. 1 (January 2014): 63–77.
- Skorupski, John. *The Domain of Reasons*. Oxford: Oxford University Press, 2010.
- Taurek, John M. "Should the Numbers Count?" *Philosophy and Public Affairs* 6, no. 4 (Summer 1977): 293–316.
- Thomson, Judith Jarvis. *The Realm of Rights*. Cambridge, MA: Harvard University Press, 1990.
- Voorhoeve, Alex. "How Should We Aggregate Competing Claims?" *Ethics* 125, no. 1 (October 2014): 64–87.
- Waldron, Jeremy. "A Right to Do Wrong." *Ethics* 92, no. 1 (October 1981): 21–39.
- Walén, Alec. "The Restricting Claims Principle Revisited: Grounding the Means Principle on the Agent-Patient Divide." *Law and Philosophy* 35, no. 2 (April 2016): 211–47.

SHOULD LIBERTARIANS REJECT THE FREE MARKET? ON OLSARETTI'S POSITIVE ANSWER

Peter Bornschein

LIBERTARIANS should reject the free market.¹ Their principles require it. These bold claims are what Serena Olsaretti argues for in a book and a series of articles.² According to her, libertarians are wrong to believe that the free market—a market unconstrained by regulation and redistributive taxation—is compatible with the liberty of each individual to lead their life according to their own choices.³ The thrust of Olsaretti's argument is that libertarians rely on a problematic account of voluntary action. As part of this argument, she develops an alternative account of voluntary action. As she sees it, if the libertarian

- 1 Throughout this essay, by “libertarianism” I mean *right*-libertarianism. All libertarians are committed to the view that each person has full ownership over their body and mind, and that this places significant constraints on what governments may do. But libertarians differ over ownership of external resources (resources external to anyone's body and mind). Right-libertarians favor a regime of robust private property rights that is insensitive to considerations of inequality, while *left*-libertarians favor a regime of property rights that is highly sensitive to considerations of inequality. For authors and their works that are representative of right-libertarianism, see Machan, *Individuals and Their Rights*; Mack, “Moral Individualism and Libertarian Theory”; Narveson, *The Libertarian Idea*; Nozick, *Anarchy, State, and Utopia*; Rothbard, *For a New Liberty*. For authors and their works that are representative of left-libertarianism, see Otsuka, *Libertarianism without Inequality*; Steiner, *An Essay on Rights*; Vallentyne, Steiner, and Otsuka, “Why Left-Libertarianism Is Not Incoherent, Indeterminate, or Irrelevant”; Van Parijs, *Real Freedom for All*. Also see the anthologies by Vallentyne and Steiner, *The Origins of Left-Libertarianism* and *Left-Libertarianism and Its Critics*.
- 2 Olsaretti, *Liberty, Desert and the Market*, “Freedom, Force, and Choice,” “Debate,” “Coercion and Libertarianism,” and “Rescuing Justice and Equality from Libertarianism.”
- 3 I am using “free market” in the same sense as Olsaretti—namely, to “refer to the free or unregulated process whereby full private property rights are exchanged, so that goods and services over which people have rights are transferred and exchanged at whatever conditions the individuals whose rights these are choose” (*Liberty, Desert and the Market*, 2).

accepts her account of voluntary action, then the libertarian must abandon their opposition to redistributive taxation and government regulation of the market.

In what follows, I argue that by accepting her account of voluntary action, libertarians are not compelled to alter their position in the way that she claims. In fact, I argue that by accepting her account of voluntary action, libertarians would be compelled to alter their position in a way that Olsaretti would probably find objectionable.⁴

1. OLSARETTI'S ARGUMENT

Olsaretti takes the core commitment of libertarians to be the right of self-ownership.⁵ The right of self-ownership grants each individual the freedom to lead their life according to their voluntary choices. This implies that any interference with a person's freedom requires their voluntary consent (except when interference is necessary to ensure respect for libertarian rights), and, therefore, any enforceable obligation (beyond the obligation to respect libertarian rights) must be voluntarily chosen.⁶ But voluntariness is not just necessary for the assumption of an enforceable obligation, it is also sufficient. So if I have an enforceable obligation to provide you some good or service, that can only be because I voluntarily agreed to provide you that good or service. Additionally, what good or service I am to provide you, and under what conditions, is irrelevant (assuming it does not involve a violation of anyone's libertarian rights). It is sufficient that I voluntarily agreed to provide that good or service. This idea, "that voluntariness is a necessary and sufficient condition for the legitimacy of nearly all obligations and interference with individuals," Olsaretti dubs the *voluntariness requirement*.⁷

Self-ownership and the voluntariness requirement are not the only things libertarians believe in. Libertarians also believe that individuals should be free to acquire full ownership rights over external resources, and that these rights are nearly absolute. It is this belief that has drawn the most attention from libertarianism's critics.⁸ But Olsaretti's attention lies elsewhere. She does not chal-

4 For different criticisms of Olsaretti's argument, see Barnes, "Why Is Coercion Unjust?"; Garnett, "Coercion," 572–73.

5 Olsaretti, *Liberty, Desert and the Market*, 91.

6 By "enforceable obligation" I mean an obligation that it is permissible to compel someone, through the threat of physical force, to comply with. As Olsaretti puts it, "Having enforceable obligations makes one liable to legitimate interference" (*Liberty, Desert and the Market*, 1103).

7 Olsaretti, *Liberty, Desert and the Market*, 89.

8 For just a few examples, see Christman, *The Myth of Property*; Cohen, *Self-Ownership, Free-*

lenge the libertarian's theory of property acquisition, nor does she challenge the theory of self-ownership or the voluntariness requirement. What Olsaretti challenges is the libertarian's assumption that the free market uniquely satisfies the voluntariness requirement.⁹

Olsaretti criticizes libertarians for having a flawed conception of voluntary action. These criticisms, however, are not my focus.¹⁰ My focus is on Olsaretti's alternative account of voluntary action. She presents this account, strictly speaking, in terms of what makes an action *nonvoluntary*. According to her, "a choice is non-voluntary if and only if it is made because the alternatives which the chooser believes she faces are unacceptable."¹¹ This analysis includes a number of components that require elaboration.

With respect to what makes an alternative unacceptable, Olsaretti assumes that there is an objective fact of the matter.¹² An agent's believing that her alternatives are unacceptable does not make them so. Although she does not give a complete account of what makes an alternative objectively unacceptable, Olsaretti does think an example would be an alternative that required an agent to forego her basic needs (e.g., food, water, shelter).¹³

Now, although an agent's beliefs do not make her alternatives unacceptable, it is not necessary that her alternatives actually be such in order for her to act nonvoluntarily. An agent may act nonvoluntarily when she mistakenly believes that her alternatives are objectively unacceptable.¹⁴ Her mistake would not lie in her belief about the criteria for what makes an alternative objectively unacceptable. Rather, her mistake would lie in her belief that the alternatives available to her match those criteria. Hence, an agent may act nonvoluntarily when she incorrectly believes that an alternative involves sacrificing her basic needs.

Finally, although an agent may believe that she only faces (objectively) unacceptable alternatives—and she may be right—that does not necessarily mean that her action is nonvoluntary. In order for an agent to be forced into doing something, she must do it *because* she believes there are no acceptable alterna-

dom, and Equality, chs. 3–4; Gibbard, "Natural Property Rights"; O'Neill, "Nozick's Entitlements"; Ryan, "Yours, Mine, and Ours."

9 Olsaretti, *Liberty, Desert and the Market*, 96.

10 Chapter 5 of *Liberty, Desert and the Market* contains the majority of Olsaretti's criticisms. Chapter 6 also contains such criticisms but is primarily devoted to defending her own account of voluntary action. Also see Olsaretti, "Freedom, Force, and Choice."

11 Olsaretti, "Debate," 114.

12 Olsaretti, *Liberty, Desert and the Market*, 153.

13 Olsaretti, *Liberty, Desert and the Market*, 119, 154.

14 Olsaretti, "Debate," 113–14.

tives.¹⁵ An agent may do x while believing that there are no acceptable alternatives, yet not have been forced to do x , if they would have chosen to do x even had there been acceptable alternatives. Hence, an agent may work at a particular job knowing that the alternative is starvation. But if the agent would still work at this job in the face of acceptable alternatives, then they work at this job voluntarily.

The intended upshot of all of this is that if Olsaretti is correct about what makes an action nonvoluntary, then libertarians should abandon their commitment to the free market. This is because the market often involves choices made by individuals, typically low-income, that are made because these individuals believe (correctly, in most cases) that there are no acceptable alternatives. As Olsaretti puts it:

Insofar as libertarians take the voluntariness of choice to be a necessary condition for the legitimacy of interference with individuals, so that only voluntary choices are justice-preserving and responsibility-grounding, they then have reason to support not the free market, in which the voluntariness of many choices is vitiated, but rather those circumstances in which every one is provided with a range of acceptable options and is thereby enabled to make voluntary choices.¹⁶

“Those circumstances,” according to Olsaretti, require both government regulations of the market and a government-funded safety net.¹⁷

2. A CRITIQUE OF OLSARETTI'S ARGUMENT

Olsaretti claims that if libertarians were to accept her account of voluntary action, they would be forced to give up their embrace of the free market in favor of government regulation and a social safety net. However, nothing Olsaretti argues counts against the claim that government regulation and redistribution, because they are coercive, violate the voluntariness requirement. Therefore, libertarians can consistently say, while accepting Olsaretti's account of voluntary action, that it is wrong to subject someone to taxation and regulation in order to give others “a range of acceptable options.”

Not only do government regulation and redistribution violate the voluntariness requirement, but the case of a person working just to subsist does *not* violate the voluntariness requirement. How so? Because the voluntariness requirement

15 Olsaretti, *Liberty, Desert and the Market*, 156.

16 Olsaretti, *Liberty, Desert and the Market*, 137.

17 Olsaretti, *Liberty, Desert and the Market*, 164.

is a rule governing when it is acceptable to interfere with someone. The fact that a person takes a job in order to subsist does not, as such, involve interference with them. Additionally, the sheer fact that a person cannot meet their basic needs is not a violation of the voluntariness requirement. Letting a person starve, however reprehensible, is not an interference with them. So were the libertarian to get past the coerciveness of government regulation and redistribution, they still would not be compelled to endorse a social safety net, despite accepting Olsaretti's account of voluntary action.

The case of a person working in order to subsist is only relevant to the voluntariness requirement if a contract or exchange is involved. Herein lies a change to the libertarian's view that they would be compelled to make if they adopted Olsaretti's account of voluntary action. The libertarian would have to revise their stance on contracts. In particular, they would have to treat as unenforceable all contracts and exchanges entered into because one of the parties believed they had no acceptable alternative. This would include all contracts and exchanges involving parties who are extremely poor (even ones where *all* the parties are extremely poor). Yet to not enforce such contracts or exchanges would create a major barrier to the extremely poor ever lifting themselves out of poverty.¹⁸

Now consider a trope that Olsaretti often uses. The trope is that of a worker who works a hazardous job for low wages because the alternative to not working is starvation.¹⁹ What would the revised libertarian view (the libertarian view that includes Olsaretti's account of voluntary action) say about this trope?

If we imagine that this worker is an at-will employee, then it is unclear how justice is being disrupted, even on the revised libertarian view. The very nature of employment-at-will is that the employee is not locked into any agreement—they can quit at any time. Only the employer incurs an enforceable obligation—namely, the obligation to pay the employee for any labor provided.²⁰ The em-

18 According to Nobel laureates Abhijit V. Banerjee and Esther Duflo, the lack of stable, timely, and low-cost contract enforcement in developing nations explains why moneylenders charge poor borrowers high interest rates ("The Economic Lives of the Poor," 155–56). This lack of access to cheap credit makes it virtually impossible for the extremely poor to start businesses. This is especially tragic given the entrepreneurial spirit manifested, albeit on a small scale, by so many poor people in developing nations. In addition to the piece by Banerjee and Duflo, see Collins et al., *Portfolios of the Poor*; Rutherford, *The Poor and Their Money*.

19 Some places where this trope can be found include Olsaretti, *Liberty, Desert and the Market*, 119–20, 124–25, 127–28, 156, 160, and "Freedom, Force, and Choice," 58–59, 61–62, 72, 74.

20 If the employee is paid before providing any labor, then, absent providing such labor, they have an enforceable obligation to (at least) return the money they were paid. This enforceable obligation is consistent with the revised libertarian view because voluntary consent is not required for the enforceable obligation to respect a person's libertarian rights, and

ployer does have the right to fire the employee. But firing the employee only involves withholding a benefit from them and, as such, does not violate the voluntariness requirement.

Now, suppose that our worker and all of their co-workers organize a union and successfully negotiate employment contracts providing for better pay and benefits, as well as better hours and working conditions. Suppose further that part of this contract includes the provision that every employee must be a member of the union and must pay union dues (taken out of each worker's paycheck). Our worker's new employment contract is justice disrupting on the revised libertarian view. Our worker has nonvoluntarily incurred an enforceable obligation to pay union dues, therefore the voluntariness requirement is violated. (We are still assuming that the worker has no acceptable alternative to working the job they have.) Interestingly, our worker's new employment contract would not be considered justice disrupting on the traditional libertarian view.²¹

3. POSSIBLE REPLIES

I have argued that the libertarian would not be forced to accept government regulation and redistribution were they to adopt Olsaretti's account of voluntary action. This is based on two claims: (1) government regulation and redistribution violate the voluntariness requirement, and (2) the case of a person working in order to subsist does *not* violate the voluntariness requirement.

In reply to the first claim, Olsaretti could argue that, although government regulation and redistribution are coercive mechanisms, they are not necessarily ruled out by the voluntariness requirement. This is because, on her account, a person acts voluntarily even when they have no acceptable alternative, as long as they do not do what they do *because* they have no acceptable alternative. So just because citizens face legal penalties if they do not pay their taxes or comply with certain regulations does not mean that they do these things nonvoluntarily. Their payment of taxes and compliance with regulations is still voluntary if what motivates them is not fear of punishment but a sense of justice.²²

Yet someone could use this same reasoning to argue against the necessity of

the employee would be violating the employer's libertarian rights by not working and then keeping the money.

21 That is, if we assume that the government did not force the employer to bargain with the union, and the union did not use violence, or the threat of violence, to intimidate the employer or any worker.

22 Olsaretti could make a similar point about union workers who have no acceptable alternative to working and paying union dues.

government regulation and redistribution by saying that just because we know that a destitute person has no acceptable alternative to working a low-paying, hazardous job does not mean that their choice to work at that job is not voluntary. In response, Olsaretti might say that we can assume that any destitute person working at such a job is not doing so voluntarily. Besides, even if they were, the point of regulation and redistribution is to *ensure* that everyone is making voluntary choices.

But then should we not avoid market regulations and redistributive taxation so as to make sure the voluntariness requirement is satisfied? Also, if we were sure that all citizens are sufficiently moved by a sense of justice to pay taxes and comply with regulations, then there would be no need for legal penalties to serve as incentives. This is just another way of saying that there would be no need for taxes and regulations.²³ Thus, any situation where we can be sure that government regulation and redistribution satisfy the voluntariness requirement is also a situation where government regulation and redistribution are not needed.

Moving on to the second claim (that the case of a person working in order to subsist does *not* violate the voluntariness requirement), Olsaretti could concede the point. However, she could argue that if we go back to what motivates the voluntariness requirement in the first place—namely, the right of self-ownership and the attendant idea that people’s actions should be the result of their voluntary choices—then by adopting her account of voluntary action, libertarians would be forced to endorse government regulation and redistribution.

Now if the right of self-ownership includes the right that our choices be voluntary, then taxation is ruled out for reasons already mentioned. On the other hand, things look more promising for Olsaretti with respect to regulations, specifically regulations governing wages and working conditions. As Olsaretti sees it, an employer who offers a destitute person low-paying, hazardous work because they know that person has no acceptable alternative is “intentionally and avoidably contributing to another person’s being forced.”²⁴ In virtue of im-

23 In order to better coordinate their efforts, such citizens may want to have the government run welfare programs and set recommendations for wage rates and working conditions. As long as there are no legal penalties for not giving money to these programs or for not complying with these recommendations, then these programs and recommendations would be permissible on the traditional libertarian view.

24 Olsaretti says this in the context of addressing a criticism by Gordon Barnes. Olsaretti claims that some offers are unjust because they involve the subordination of one person’s will to another (*Liberty, Desert and the Market*, 146). As a counterexample, Barnes presents the following thought experiment:

You and I are wandering in the desert, and we run into each other. You are lost, and you have no idea how to get out of the desert. However, I know the territory, and so

pinging on a person's freedom to lead their life according to their own voluntary choices, such an offer violates the right of self-ownership.

I think one can dispute the claim that an employer who makes such an offer is "intentionally and avoidably contributing to another person's being forced."²⁵ However, I will not press that objection here. I will simply say that if the libertarian were to rule out such offers (not just the enforcement of contracts and exchanges that result from them), then we still have a view that looks worse for the poor than the traditional libertarian view. It is still the case that all contracts and exchanges between the extremely poor would be deemed unenforceable, and that government redistribution of wealth would be ruled out. The only thing that looks good for the poor now is that if they get job offers, those jobs will have to pay a certain minimum, and the working conditions will have to meet certain standards. But if businesses cannot afford to pay higher wages and provide better working conditions, then those jobs will not exist, and there will be no safety net to fall back on.²⁶

4. CONCLUSION

Libertarians support a society where all people's choices are freely made. Libertarians also support the free market. Olsaretti claims that they cannot have it

I know my way out. I offer to let you follow me out of the desert. Since you have no alternative, it seems that your will is now subordinate to mine, at least if you want to get out of the desert. However, it is just as clear that there is no injustice in this situation. (*Why Is Coercion Unjust?* 463)

In response (this is where the quote in the text is from), Olsaretti says:

We could hold that someone's *intentionally and avoidably contributing* to another person's *being forced* is necessary for a forced exchange to be justice-disrupting. . . . On this view, there is an important difference between Barnes' desert case . . . and the case in which I offer to let you follow me out of the desert only if you sign a contract with which you bind yourself to work for me at subsistence wages for the next decade. In the latter but not in the former case, I am intentionally contributing to making it the case that you make a forced choice. In the latter but not the former case there is an injustice, because *I set the terms of the offer* in such a way that your choice to accept them is forced. (*Coercion and Libertarianism*, 298)

- 25 In her example mentioned in the previous note, it is not clear how Olsaretti is guilty of "intentionally and avoidably contributing" to you being forced. What contributes to you being forced is that you are lost in the desert (that is why you have no acceptable alternative). Olsaretti does not say in her example that she is responsible for you being lost in the desert.
- 26 My thanks to an anonymous reviewer for encouraging me to address the two replies I have discussed in this section.

both ways. Given her account of voluntary action, libertarians should see that a society where all people's choices are freely made is not a free market society, but, rather, is a society that involves government regulation of the market and a government-funded safety net.

I have argued, however, that Olsaretti is wrong about the implications of a libertarian view that includes her account of voluntary action. Even *if* this revised version of libertarianism countenanced certain kinds of market regulations (and I remain skeptical), it would condemn a government-funded safety net and attempts by the poor to better their own condition by entering into legally enforceable contracts and exchanges. If libertarians are to be convinced that they should reject the free market in favor of something like social democracy, then the thesis of self-ownership, the voluntariness requirement, or their view on property rights will need to be challenged. Keeping those features of libertarianism in place, while adopting Olsaretti's account of voluntary action, leads to a view that is far harsher to the poor than the kind of libertarianism she set out to critique.²⁷

Eastern Michigan University
pbornsch@emich.edu

REFERENCES

- Banerjee, Abhijit V., and Esther Duflo. "The Economic Lives of the Poor." *Journal of Economic Perspectives* 21, no. 1 (Winter 2007): 141–68.
- Barnes, Gordon. "Why Is Coercion Unjust? Olsaretti vs. the Libertarian." *Analysis* 72, no. 3 (July 2012): 457–65.
- Christman, John. *The Myth of Property: Toward an Egalitarian Theory of Ownership*. New York: Oxford University Press, 1994.
- Cohen, G. A. *Self-Ownership, Freedom, and Equality*. Cambridge: Cambridge University Press, 1995.
- Collins, Daryl, Jonathan Morduch, Stuart Rutherford, and Orlanda Ruthven. *Portfolios of the Poor: How the World's Poor Live on \$2 a Day*. Princeton: Princeton University Press, 2009.

²⁷ I would like to thank Ryan Fischbeck, Jacob Sparks, and an anonymous reviewer for this journal for their helpful comments on earlier drafts of this paper. An earlier version of this paper was presented to an audience at Albion College. I would like to thank those who were in attendance for their helpful feedback.

- Garnett, Michael. "Coercion: The Wrong and the Bad." *Ethics* 128, no. 3 (April 2018): 545–73.
- Gibbard, Allan. "Natural Property Rights." *Nous* 10, no. 1 (March 1976): 77–86.
- Machan, Tibor R. *Individuals and Their Rights*. La Salle, IL: Open Court, 1989.
- Mack, Eric. "Moral Individualism and Libertarian Theory." In *Liberty for the Twenty-First Century*, edited by Tibor R. Machan and Douglas B. Rasmussen, 41–58. Lanham, MD: Rowman & Littlefield, 1995.
- Narveson, Jan. *The Libertarian Idea*. Peterborough, ON: Broadview Press, 2001.
- Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.
- Olsaretti, Serena. "Coercion and Libertarianism: A Reply to Gordon Barnes." *Analysis* 73, no. 12 (April 2013): 295–99.
- . "Debate: The Concept of Voluntariness—A Reply." *Journal of Political Philosophy* 16, no. 1 (March 2008): 112–21.
- . "Freedom, Force, and Choice: Against the Rights-Based Definition of Voluntariness." *Journal of Political Philosophy* 6, no. 1 (March 1998): 53–78.
- . *Liberty, Desert and the Market*. Cambridge: Cambridge University Press, 2004.
- . "Rescuing Justice and Equality from Libertarianism." *Economics and Philosophy* 29, no. 1 (March 2013): 43–63.
- O'Neill, Onora. "Nozick's Entitlements." *Inquiry* 19, nos. 1–4 (1976): 468–81.
- Otsuka, Michael. *Libertarianism without Inequality*. Oxford: Oxford University Press, 2003.
- Rothbard, Murray. *For a New Liberty: The Libertarian Manifesto*. Auburn, AL: The Ludwig von Mises Institute, 2006.
- Rutherford, Stuart. *The Poor and Their Money*. New Delhi: Oxford University Press, 2000.
- Ryan, Cheyney C. "Yours, Mine, and Ours: Property Rights and Individual Liberty." *Ethics* 87, no. 2 (January 1977): 126–41.
- Steiner, Hillel. *An Essay on Rights*. Cambridge, MA: Blackwell Publishing, 1994.
- Vallentyne, Peter, and Hillel Steiner, eds. *Left-Libertarianism and Its Critics: The Contemporary Debate*. New York: Palgrave, 2000.
- . *The Origins of Left-Libertarianism: An Anthology of Historical Writings*. New York: Palgrave, 2000.
- Vallentyne, Peter, Hillel Steiner, and Michael Otsuka. "Why Left-Libertarianism Is Not Incoherent, Indeterminate, or Irrelevant: A Reply to Fried." *Philosophy and Public Affairs* 33, no. 2 (Spring 2005): 201–15.
- Van Parijs, Philippe. *Real Freedom for All*. New York: Oxford University Press, 1995.

DIVINE COMMANDS ARE UNNECESSARY FOR MORAL OBLIGATION

Erik J. Wielenberg

DIVINE COMMAND THEORY is experiencing something of a renaissance, inspired in large part by Robert Adams's 1999 masterpiece *Finite and Infinite Goods*. The basic ethical framework presented in that book has significantly influenced later work on divine command theory (DCT) by William Wainwright, David Baggett and Jerry Walls, C. Stephen Evans, and John Hare.¹ DCT's central claims are that (i) S is morally obligated to do A just in case God commands S to do A and (ii) all moral obligations are somehow grounded in or generated by divine commands. DCT-ists spell out this grounding relation differently.² Adams, for example, suggests that human moral obligations are *constituted* by the commands of a good and loving God.³

One prominent challenge to DCT alleges that divine commands are not always *sufficient* for moral obligation. For instance, it has been claimed that even if God were to command the torture of innocent children, torturing innocent children would not be morally obligatory.⁴ Here I develop a challenge to DCT that has received much less attention: I argue that divine commands are not always *necessary* for actions to be morally obligatory.⁵ One way of developing such an

1 Wainwright, *Reason and Morality*; Baggett and Walls, *Good God*; Evans, *God and Moral Obligation*; and Hare, *God's Command*.

2 See Murphy, "Theological Voluntarism."

3 Adams, *Finite and Infinite Goods*, 250, 281.

4 See Morriston, "What If God Commanded Something Terrible?"; and Pruss, "Another Step in Divine Command Dialectics."

5 Arguments for a similar conclusion have been advanced by Wainwright, *Reason and Morality*, 80–83; Wielenberg, *Value and Virtue in a Godless Universe*, 53–67; Manis, *Virtue, Divine Commands and the Debt of Creation*, 106 and 117–26; Wolterstorff, *Justice*, 272; and Murphy, *God and Moral Law*, 116–32. The argument here differs from those arguments in that I seek to leverage the typical commitments of contemporary DCT-ists in order to undermine their position that divine commands are necessary for moral obligations. The argument of this paper is also distinct from the "prior obligations" objection, which focuses on the obligation to obey God's commands (Evans, *God and Moral Obligation*, 98–101). The argument here is most similar to the argument of Manis; the main difference is that Manis's argument is tied

argument would be to defend a moral theory that is an alternative to DCT—utilitarianism, for example—and show that the alternative theory implies that there are some moral obligations not grounded in divine commands. We might call that an external critique of DCT. I will instead develop what could be called an internal critique of DCT, making the case that the DCT-ist's own commitments put pressure on her to concede the existence of some moral obligations that in no way depend on divine commands. Focusing on Robert Adams's theistic framework for ethics, I will argue that Adams's views about good, evil, reasons, and the nature of moral obligation suggest that there can be moral obligations that exist independently of any divine commands. My argument proceeds through the development of an example in which there is a moral obligation that is not even partially grounded in a divine command. I focus primarily on Adams's view because, as I noted above, many contemporary DCT-ists work in a broadly Adamsian framework, so the argument I advance here poses a challenge for many contemporary versions of DCT.

1. THE MAIN ARGUMENT

Adams develops what he calls a "theistic framework for ethics," of which DCT is just one part.⁶ He offers distinct accounts of three central ethical phenomena: (i) the Good, which is a transcendent paradigm of goodness, (ii) the goodness of finite things, and (iii) moral obligation. Adams proposes that the Good is God, finite goodness is resemblance to the necessarily existing divine nature, and moral obligation is being commanded by God. Note that these are claims not about the *meaning* of the words "good" or "obligatory" but rather about the *nature* of goodness and moral obligation. Adams claims that the meaning of a given term does not always reveal the full nature of the thing to which the term refers.⁷ An implication of Adams's view is that there are many facts about good and evil that do not depend on any divine commands. Relatedly, Adams's view implies that there are some reasons that exist independently of divine commands—reasons not in the psychological sense, as actual psychological motives, but rather in the objective or justifying or normative or third-person sense, as features that count in favor of a course of action whether they actually motivate anyone or not. For example, Adams asserts that the goodness of the consequences of a given action

specifically to the ideas of Kierkegaard whereas the argument here focuses primarily on the view of Adams. My argument can perhaps be understood as a fleshing out of a challenge for DCT that Murphy gestures toward in section 3.3 of *God and Moral Law*.

6 Adams, *Finite and Infinite Goods*, 4.

7 Adams, *Finite and Infinite Goods*, 15–16.

provides a reason to perform that action and also that the fact that an act is “good, either intrinsically or by serving a pattern of life that is very good” provides a reason to perform the act.⁸ Since good and evil can exist independently of God’s commands, there are reasons—indeed, moral reasons—for and against various courses of action independently of any divine command.⁹

Suppose, then, that I make a solemn promise to my wife to end my practice of staying out late drinking—a practice that, let us suppose, has significantly damaged our relationship, a relationship that is valuable and that I in fact value deeply. Keeping this promise would have good consequences and would serve a pattern of life that is very good, and so, by Adams’s lights, there are good reasons for me to keep the promise. Adams suggests that

valuing one’s social bonds gives one, under certain conditions, a reason to do what is required of one’s associates or one’s community. . . . I value the relationship which I see myself as actually having, and my complying is an *expression* of my valuing and respecting the relationship.¹⁰

Adams here speaks of a case in which one party has made a demand of another, but it is plausible that something similar holds in the case of a promise like the one described above. I value the relationship with my wife, and my keeping my promise is an expression of my valuing and respecting the relationship; that is a (moral) reason to keep the promise.

Accordingly, independently of any divine commands, there are strong moral reasons for me to keep my promise to refrain from staying out late drinking. We can bring further reasons to keep the promise into the picture by drawing on Adams’s views about evil. Adams suggests that evil can be understood as being related to the divine nature in a way other than resemblance or similarity. One important kind of badness is constituted by attacking or opposing something good. And one particularly nasty form of attacking something good is what Adams calls “morally horrible.” Some acts in this category are “rape, murder, and maiming, torturing, or brainwashing a human being.”¹¹ According to Adams, what makes these sorts of acts particularly bad is that they are all violations of human persons, who are sacred images of God. Such violations are acts that attack persons seriously and directly; their foreseeable effects are so damaging to

8 Adams, *Finite and Infinite Goods*, 241, 255.

9 That these reasons are properly classified as *moral* reasons on Adams’s view is suggested by the fact that on his view, these are the sorts of reasons for action generated by moral obligations.

10 Adams, *Finite and Infinite Goods*, 242.

11 Adams, *Finite and Infinite Goods*, 104.

their victims that fully intending them constitutes hostility toward their victims. Adams says that without any divine commands, murder and torture would not be morally wrong “but they would still be bad, and specifically horrible.”¹²

Suppose now that I find myself at the bar early one evening. I have had a couple of drinks but I have not yet stayed out late drinking, so I have not broken my promise. Suppose further that things are turning ugly; the group I am with is turning toward drinking hard liquor, and I can foresee that if I stay at the bar and participate in the drinking of hard liquor, there is a good chance that I will be a participant in a murder later in the evening. Now I have an additional compelling moral reason to keep my promise to my wife: if I do not keep the promise on this occasion, there is a good chance that I will participate in a morally horrible act. In this situation, there are several powerful moral reasons in favor of me keeping my promise and very few (and no powerful) reasons in favor of me breaking my promise. Overall, then, it seems that I have moral reasons for keeping my promise that are both powerful and *decisive*.¹³

I suggest that in the example I have described, I am morally obligated to keep my promise and that this obligation exists independently of any divine commands. As I just noted, I have powerful and decisive moral reasons to keep the promise.¹⁴ Keeping the promise has many of the features that Adams and Evans (another prominent DCT-ist) identify as defining features of moral obligations: It is something I should take seriously and care about.¹⁵ If I give the various reasons described above the appropriate weight, then I will be strongly motivated to keep the promise.¹⁶ If I break the promise and am responsible for doing so, then I am guilty and blameworthy.¹⁷ There is someone (my wife) who has the right to expect me to keep my promise and who may rightly be disappointed in

12 Adams, *Finite and Infinite Goods*, 105, 107–8.

13 See Parfit, *On What Matters*, 33.

14 Adams offers an argument against the principle that the morally obligatory is “what we have most reason (from a moral point of view) to do,” so I should note that my argument does not depend on any such principle (*Finite and Infinite Goods*, 238). My argument is rather an argument from counterexample; I build a case that seems to have the various marks of moral obligation identified by Adams independently of any divine command. Adams describes a case in which “the preponderance of moral reasons favors your not walking on the lawn, but also favors not worrying about it very much and not feeling guilty if you do it,” suggesting that in such a case there is no moral obligation not to walk on the lawn (*Finite and Infinite Goods*, 238). But in my example, there are strong reasons for me to worry about keeping my promise and to feel guilty if I break it.

15 Adams, *Finite and Infinite Goods*, 235; Evans, *God and Moral Obligation*, 31.

16 Adams, *Finite and Infinite Goods*, 235; Evans, *God and Moral Obligation*, 31.

17 Adams, *Finite and Infinite Goods*, 235.

me and blame me if I fail to keep my promise.¹⁸ Adams writes approvingly of Mill's claim that "we do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience."¹⁹ Not only do I have powerful reasons not to remain at the bar, but there is also reason for me to feel guilty and hence be punished by my own conscience if I remain at the bar, and there is reason for others—particularly my wife—to punish me by their opinions of me.

The DCT-ist must deny that all of this is sufficient to give me a genuine moral obligation to keep my promise. But why? What is it that a divine command adds to the situation that is supposed to make the difference between my not being morally obligated to keep my promise and my being so obligated? At least part of the answer seems to be that a divine command brings additional powerful reasons into the situation. So, suppose that God commands me to keep my promise. I now have further reasons to keep the promise, reasons grounded in the goodness of my relationship with God and in the goodness of God himself.²⁰ Evans acknowledges that there may be "strong, even overriding reasons" to perform a given action independently of any divine command. He suggests that "God's command introduces a new character to the situation that the act alone without the command would not possess" and that "we acquire powerful new reasons for performing (or not performing) various acts when God issues commands."²¹ That seems plausible, but it is not plausible to suppose that piling on more reasons somehow makes the difference between my not having and my having a moral obligation to keep the promise. Even without the divine command, there are powerful and decisive moral reasons for me to keep the promise, and the command does not add to the situation any of the defining marks of moral obligation identified by Adams or Evans.

One feature that is missing from the promise-keeping example is a *demand*. In my example, I have made a promise, but no one has demanded anything of anyone else. Adams declares that "having an obligation to do something consists in being required . . . by another person or group of persons, to do it."²² But Adams seems not to provide any support for the claim that it is essential to the nature of moral obligation that it include a literal demand from someone. Of

18 Evans, *God and Moral Obligation*, 27.

19 Mill, *Utilitarianism*, 48–49. See Adams, *Finite and Infinite Goods*, 233–34.

20 Adams, *Finite and Infinite Goods*, 252–53.

21 Evans, *God and Moral Obligation*, 36, 69.

22 Adams, *Finite and Infinite Goods*, 242. For a helpful critical discussion of this alleged "social character" of moral obligation, see Murphy, *God and Moral Law*, 124–32.

course, one may stipulate that they will use “moral obligation” in a particular way, and so it is open to the DCT-ist to declare that when they use “moral obligation,” they mean something like “a moral requirement constituted by a demand from another person.” But then the debate between the DCT-ist and her opponent becomes a purely semantic one upon which nothing of substance hangs. Adams’s view of good, evil, and reasons suggests that, independently of any divine commands, there can be actions that a person has powerful and decisive moral reasons to perform, that should be taken seriously, that one can be strongly motivated to perform, that one is guilty and blameworthy for (responsibly) failing to perform, and that one deserves punishment for (responsibly) failing to perform. There is no good reason to insist that such acts are *genuinely* morally obligatory only if someone demands that they be performed.²³

Adams suggests that moral obligation “grounds reasons for action.”²⁴ These remarks suggest the view that what constitutes *S*’s moral obligation to do *A* is the facts that generate moral reasons for *S* to do *A* (and perhaps to feel guilty if *S* neglects to do *A*, and for others to blame *S* for responsibly failing to do *A*). Adams says that what constitutes *S*’s moral obligation to do *A* is the command of a good and loving God to *S* to do *A*. However, strictly speaking, it seems that on Adams’s view moral obligations are also partially constituted by certain other facts, such as *S* having a valuable relationship with God, God having done various things for *S*, and *A* itself being good.²⁵ Accordingly, in the promise-keeping example described above, what constitutes my moral obligation to refrain from staying out late drinking is my promise to my wife—or more precisely, the promise together with our relationship and various other relevant features of the situation. Other moral obligations in other circumstances are constituted by different reason-giving features of the situation. While DCT offers what we might call a monist account of what constitutes moral obligations (moral obligations are always constituted by divine commands), a plausible alternative is a pluralist account where moral obligations are constituted by different sorts of facts in different situations—an account along broadly Rossian lines.²⁶

23 Perhaps one thought behind DCT is that obligations are moral demands; on DCT, God is morality, and morality *literally* demands things of us! But this line of thought seems to be driven by an excessively literal understanding of “moral demands.”

24 Adams, *Finite and Infinite Goods*, 241. See also Evans, *God and Moral Obligation*, 15.

25 Adams, *Finite and Infinite Goods*, 252–53.

26 Ross, *The Right and the Good*, 16–47.

2. CONCLUSION

I have argued that DCT goes wrong in its central claim that all moral obligations are somehow grounded in divine commands. This is not to deny that if God exists and issues commands, his commands generate moral obligations. Divine commands may, for instance, turn acts that would otherwise be merely morally permissible—such as loving one’s neighbors with all of one’s heart—into acts that are morally obligatory.²⁷ The main lesson of the present discussion is that it is going too far to claim that *only* a divine command can give rise to a moral obligation; the correct view is that divine commands are one among many possible sources of moral obligation.²⁸

DePauw University
ewielenberg@depauw.edu

REFERENCES

- Adams, Robert. *Finite and Infinite Goods: A Theistic Framework for Ethics*. Oxford: Oxford University Press, 1999.
- Baggett, David, and Jerry L. Walls. *Good God: The Theistic Foundations of Morality*. Oxford: Oxford University Press, 2011.
- Evans, C. Stephen. *God and Moral Obligation*. Oxford: Oxford University Press, 2013.
- Hare, John. *God’s Command*. Oxford: Oxford University Press, 2015.
- Manis, R. Zachary. *Virtue, Divine Commands and the Debt of Creation: Towards a Kierkegaardian Christian Ethic*. PhD diss., Baylor University, 2006.
- Mill, John Stuart. *Utilitarianism*. Edited by George Sher. 2nd ed. Indianapolis: Hackett, 2001.
- Morrison, Wes. “What If God Commanded Something Terrible? A Worry for Divine-Command Meta-Ethics.” *Religious Studies* 45, no. 3 (Summer 2009): 249–67.
- Murphy, Mark. *God and Moral Law: On the Theistic Explanation of Morality*. Oxford: Oxford University Press, 2011.
- . “Theological Voluntarism.” *Stanford Encyclopedia of Philosophy* (Summer 2019). <https://plato.stanford.edu/archives/sum2019/entries/voluntarism-theological/>.

27 Quinn, “Divine Command Theory.”

28 I am grateful to Wes Morrison for helpful discussion of the argument of this paper.

- Parfit, Derek. *On What Matters*. Vol. 1. Oxford: Oxford University Press, 2011.
- Pruss, Alexander. "Another Step in Divine Command Dialectics." *Faith and Philosophy* 26, no. 4 (October 2009): 432–39.
- Quinn, Philip. "Divine Command Theory." In *The Blackwell Guide to Ethical Theory*, 2nd ed., edited by Hugh LaFollette, 81–102. Oxford: Blackwell, 2013.
- Ross, David. *The Right and the Good*. Edited by Philip Stratton-Lake. 2nd ed. Oxford: Oxford University Press, 2003.
- Wainwright, William. *Reason and Morality*. Burlington, VT: Ashgate, 2005.
- Wielenberg, Erik. *Value and Virtue in a Godless Universe*. Cambridge: Cambridge University Press, 2005.
- Wolterstorff, Nicholas. *Justice: Rights and Wrongs*. Princeton: Princeton University Press, 2008.

JOURNAL of ETHICS & SOCIAL PHILOSOPHY
<http://www.jesp.org>
ISSN 1559-3061

The *Journal of Ethics and Social Philosophy* (JESP) is a peer-reviewed online journal in moral, social, political, and legal philosophy. The journal is founded on the principle of publisher-funded open access. There are no publication fees for authors, and public access to articles is free of charge. Articles are typically published under the CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-NODERIVATIVES 4.0 license, though authors can request a different Creative Commons license if one is required for funding purposes.



Funding for the journal has been made possible through the generous commitment of the Gould School of Law and the Dornsife College of Letters, Arts, and Sciences at the University of Southern California.