JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY

VOLUME XXV · NUMBER 2
August 2023

ARTICLES

DISCUSSIONS

OPEN ACCESS FREE FOR AUTHORS AND READERS ISSN 1559-3061

JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY

VOLUME XXV · NUMBER 2
August 2023

ARTICLES

DISCUSSIONS

JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY
http://www.jesp.org

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.

Journal of Ethics and Social Philosophy
VOL. 25, NO. 2 · AUGUST 2023
https://doi.org/10.26556/jesp.v25i2.1547

IN DEFENSE OF MODERATION

CULPABLE IGNORANCE AND THE STRUCTURE OF EXCULPATION

James Goodrich

An infant begins to drown in a neighborhood pool. Anne—the local on-duty lifeguard—jumps in, pulls the infant out, and performs standard CPR. The infant dies. The infant would have survived had Anne performed a different CPR procedure—the one designed for infants. Why did Anne not perform the correct procedure? She did not know about it. And she did not know about the special procedure for infant CPR because, during her training, Anne left early to take a smoke. Anne knew that important information might be shared during the ten minutes she was gone, but took the risk anyway.[1] Anne was thus knowingly reckless and culpable for her ignorance. Anne's recklessness, moreover, at least partially explains why she failed to save the infant. Does this mean that Anne is culpable to some degree for failing to save the infant's life?

Let us clarify this question with a distinction.[2] *Benighting acts* are those acts in which the culpably ignorant agent culpably fails to remedy her ignorance or risks missing out on some morally important information that might help guide her future decisions. Anne's act of sneaking out for a cigarette was a benighting act. The *unwitting wrongful act* is the later, objectively wrong act that the culpably ignorant agent performs out of their ignorance. Anne's unwitting wrongful act was her failure to save the infant's life.

Culpably ignorant agents, by definition, are culpable for their benighting acts. But there is disagreement over whether culpably ignorant agents are also culpable for their unwitting wrongful acts. Liberals think they are not. Conservatives and moderates believe they are culpable to at least some degree for their unwitting wrongful acts. The difference is that while conservatives believe agents are fully culpable for their unwitting wrongful acts, moderates believe

1 This case is inspired by Smith, "Culpable Ignorance," 552.

2 This terminology is introduced in Smith, "Culpable Ignorance."

that they are culpable to some degree, though not necessarily fully culpable for their unwitting wrongful acts.

I like the moderate view. It best accords with my—and I suspect others'—intuitions.[3] However, liberal critics have mounted a difficult and hitherto unanswered challenge to the moderate view. Roughly: the moderate must explain why an agent can be culpable (to some degree) for their unwitting wrongful acts because they are culpable for their benighting acts. And this explanation must sit well with plausible accounts of culpability. It has proven more difficult than one might have thought to meet this challenge.

I will defend the moderate view against the liberal's challenge. I will begin by developing a novel account of three things: (1) the grounds of culpability, (2) the grounds of excuses, and (3) the way excuses function within a theory of culpability. On my view, culpability is grounded in facts about wrongdoing *tout court*. However, the culpability-grounding function of facts about wrongdoing can be disabled by undercutting defeaters. These undercutting defeaters are what we colloquially refer to as "excuses" and excuses are then grounded in facts about an agent's quality of will. If I am right, the liberal's challenge hinges upon unchecked philosophical assumptions about the nature and structure of culpability.[4]

In the first four sections, I clarify the nature of the problem that animates my search for a new theory of the relationship between culpability and excuse and outline my account. Sections 5 through 8 develop my new account and defend it against objections.

1. THE LIBERAL'S CHALLENGE

Here is the liberal's challenge: moderates must explain why culpably ignorant agents are only partially excused for their unwitting wrongful acts in light of standard accounts of culpability.[5] In particular, the moderate needs to explain

3 There is room for disagreement between moderates about particular cases, including the one I have presented here. I propose that we grant that the moderate believes Anne to be culpable to some degree for her unwitting wrongful act. After all, this case is relevantly like the cases Smith deploys. See, again, Smith, "Culpable Ignorance," 556. If the moderate can meet the liberal's challenge on the very set of cases for which the liberal thinks their challenge best applies, we will ensure that no questions are begged.

4 Throughout this essay, I will only be concerned with ignorance of descriptive facts.

5 See especially Smith, "Culpable Ignorance," and "Tracing Cases of Culpable Ignorance"; and Husak, *Ignorance of Law*, ch. 3.

how their view is compatible with a quality of will account of culpability (QWA).[6] Holly Smith puts the point as follows:

> The culpably ignorant agent cannot be held to blame for his unwitting act, since he fails one of the conditions of culpability. His act does not arise from a defective configuration of desires and aversions.[7]

The thought is this: for an agent to be culpable, the wrong action for which the agent is culpable needs to have been produced by a morally objectionable motive, intention, or desire. And this is meant to fall out of our best account of what it is for an agent to be culpable. Smith characterizes her version of this account (roughly) as follows:

> *Smith's* QWA: The fact <*S* is culpable for *A*> is grounded in the facts that
> 1. <*A* (or its attempt) is objectively wrong>,
> 2. <*S* had a reprehensible configuration of desires and aversions>, and
> 3. <This configuration gave rise to the performance of *A*>.[8]

To get a feel for why one might be attracted to this view, consider a toy case: Beth is a conscientious walker who trips on an uneven sidewalk, falls into a puddle, and thereby splashes muddy water onto a passerby. It seems inappropriate to blame Beth. She might apologize out of kindness, but would surely be right to say, "I didn't mean to!" After all, her action did not arise out of some motivation or intention to harm or do wrong to the passerby. In fact, it did not even seem like an action! It was just an accident. Indeed, the appeal to "I didn't mean to!" is an expression of the fact that the agent did not intend any harm (and perhaps that she therefore should not be blamed).

Reconsider Anne, the lifeguard. When Anne performs adult CPR on the infant, what does she intend to do? In the version of the story I have offered, it seems like she intends to save the child. This intention is good, even noble. The fact that Anne fails to save the child (or even hastens its death) is antithetical to her intended aims. In other words still, Anne's failure to save the infant did not arise out of a reprehensible configuration of desires. Her unwitting wrongful

6 For a survey of recent quality of will accounts, see Shoemaker, "Qualities of Will."

7 See Smith, "Culpable Ignorance," 559. For similar remarks see Smith, "Tracing Cases of Culpable Ignorance," 113.

8 See Smith, "Culpable Ignorance," 556. Note that Smith does not state her account in terms of grounding conditions, but in terms of truth conditions. This is plausibly due to the philosophical norms of the era. In her writing, Smith is clearly concerned with something like explanation, not with what makes a sentence true or false. Smith confirms this in correspondence.

action arose out of good motives. If it is a necessary condition on an agent being culpable that their wrong act arises from some objectionable motives, culpably ignorant agents are not culpable at all for their unwitting wrongful acts. And this contradicts the moderate's view.

The moderate could respond by rejecting QWAs of culpability *tout court*. Some may be attracted to this move. I am not. I would rather mount a defense of the moderate view that does not crucially turn on whether we should accept a QWA of culpability. To do so would be to tether the fate of the moderate view to the hope that no form of the QWA will win the battle of theoretical virtues. Moreover, for such a move to help the moderate, it would also need to be shown that similar challenges do not arise on other accounts of culpability. As someone who is a moderate first, I would rather not take that gamble.

Here is a different move the moderate could make: the moderate could appeal to the objectionable motives that gave rise to culpably ignorant agents' benighting acts. These earlier motives are clearly objectionable. Anne should not, for example, go out for a smoke during lifeguard training. Perhaps the moderate can then say that an agent's unwitting wrongful act did arise out of a morally objectionable motive in the following sense:

> *Transfer Model*: Morally objectionable motives can "morally transfer" across (the right sort of) causal relations to give rise to later actions.

If the moderate does adopt the Transfer Model, perhaps they can explain why culpably ignorant agents are at least partially culpable for their unwitting wrongful acts: their earlier morally objectionable motives are causally related to their unwitting wrongful act.

The liberal could object that the Transfer Model leaves everything to be explained. What theory-neutral reasons do we have to think that such causally distant motives are to be considered in determining the culpability of a given action? From the liberal's point of view, the Transfer Model may seem like little more than a restatement of the moderate's intuition suitably dressed for a QWA of culpability. What the moderate needs is a credible, theory-neutral rationale for something like the Transfer Model. Without such a rationale, the moderate is open to the criticism that the Transfer Model merely reasserts the intuition that the liberal is inclined to reject. Moreover, plausibly not all motives in the causal chain leading up to a particular action are relevant. The "right sort of" locution invites reasonable philosophical suspicion. Why think that there is a way to characterize the "right sort of" causal relations such that they amount to more than the causal relations that fit the moderate's intuitions?

The Transfer Model therefore does not explain everything that needs explaining. Why would causally or temporally distant motives count as "giving

rise to" actions in the morally relevant sense? And why would some causally or temporally distant motives count as giving rise to an action, but other motives would not? The moderate needs more than an intuition here. The liberal and moderate start from a clash of intuitions. The liberal seems to be winning by incurring fewer explanatory peculiarities. The moderate needs a response to these explanatory challenges that does not commit them to such peculiarities or to claims the liberal could insist we not make.

2. My Strategy for Defending the Moderate View

The liberal's challenge to the moderate view is that the moderate's explanation of why the culpably ignorant agent is culpable looks unmotivated when combined with a QWA of culpability. How should moderates respond?

I will accept the liberal's presupposition about cases like that of Anne the lifeguard. Namely, there is not a bad motive that plausibly gives rise to Anne's unwitting wrongful action. I think we can reply to the liberal's challenge while holding this assumption fixed.

However, I will argue that moderates can help themselves to an alternative version of the QWA for which the liberal's explanatory challenge does not arise. That is, the version of the QWA that leads to the liberal's challenge has some unchecked theoretical baggage. Once we see that this theoretical baggage is unnecessary, we will see how the moderate can answer the liberal's challenge. Therefore, there is a version of the QWA of culpability that, by the moderate's own lights, runs into no serious explanatory challenges. However, I will not argue that the liberal must accept my new version of the QWA. Rather, my point is that the liberal's challenge relies on an inference from the claim that the moderate view faces explanatory trouble on one plausible QWA to the claim that the moderate view faces explanatory trouble on all plausible QWAs. My point then is that this inference at the heart of the liberal's challenge is unwarranted.

My account of culpability, *contra* Smith's, does not require an appeal to the "gives rise to" relation. The liberal's challenge is about how to make sense of objectionable motives giving rise to objectionable actions when such motives are causally distant or perhaps not even plausibly part of the causal chain. If, however, we can have a theory that does not rely on this notion as necessary, then we incur no explanatory burden. That is, we should do without the Transfer Model because it gives the "gives rise to" relation a central explanatory role. To be clear: I leave it open whether the "gives rise to" relation does work in some cases. My point is only that it is not necessary.

However, I cannot simply stipulate that I am dispensing with the claim that the "gives rise to" relation is necessary for grounding culpability. If I could,

this essay would end here. Why? Because the "gives rise to" relation has great explanatory power and it is not obvious what a satisfying account of culpability that dispenses with its central importance looks like. I will therefore need to replace the centrally important "gives rise to" relation with some other, plausible-enough theoretical tools. And these tools had better not just be cognates for which the liberal's challenge arises all over again. Thus, to credibly dispense with the "gives rise to" relation, I must answer some more foundational questions. These questions are about the essential nature of culpability and its grounds. I will now turn to constructing this theory. It will be best to proceed in small steps, for the QWA often assumed by liberals is, in some ways, only subtly different than my own. But these subtle differences, taken in conjunction, make all the difference to the plausibility of the moderate view.

3. THE SPARSE THEORY

My alternative proposal relies on a division of theoretical labor. I will first offer a theory of what grounds the fact that a given agent is culpable for a given action; then I will offer a theory of how excuses can "swoop in" to get agents off the culpability hook. Being clear about the difference between the grounds of culpability and the explanatory structure of excuses is key in understanding how we can avoid invoking the "gives rise to" relation.

Here is my account of the grounds of culpability:

> *The Sparse Theory*: The fact <*S* is culpable for *A*> is sometimes explained by the fact that <*A* (or its attempt) is morally wrong>.

That is it. The primary explanatory fact in my account of culpability is only that an agent has performed a morally wrong action.[9]

What do I mean by "wrong"? I mean whatever sense of that term you believe is most important in normative ethics. Some place great weight on the "objective" or "fact-relative" sense of wrong. Others will hold that it is more important that we focus on "subjective," "belief-relative," or "evidence-relative" conceptions. Still others will be happy to adopt a kind of pluralism, accepting each concept of moral wrongness as equally important and perhaps accepting a distinct concept of culpability corresponding to each sense of wrongness. It matters little for the purposes of this essay which sense of wrongness we deploy.[10]

9 There is a variant of the Sparse Theory according to which it is not the fact that an action is morally wrong that grounds culpability, but rather, it is the facts that make an act wrong that ground culpability. For present purposes, we can be agnostic about which is superior.

10 For what it is worth, I prefer the third, pluralist understanding of wrongness and culpability. The pluralist understanding allows us to describe situations with maximal

Of course, the Sparse Theory would be quite implausible if it were the only thing we said about culpability. After all, were you to perform a wrongful act out of non-culpable ignorance, you would not be culpable. A theory of excuse is needed to explain this fact. The Sparse Theory is therefore far more plausible when supplemented with a theory of excuse. Excuses, on my view, are undercutting defeaters. I will say more about undercutting defeaters in due course, but it is worth first discussing the relationship between the idea that excuses are defeaters and the idea that wrongful actions are the grounds of culpability.

While the fact that some agent performed a wrong action is the grounds of culpability—that is, it is the operative explanatory factor of why a given agent is culpable for a given action—background conditions still need to be met in order for wrongful action to play its grounding role. Consider a common analogy: a match is lit because I struck it. The fact that I struck the match is a perfectly good explanation of the further fact that the match is lit. However, certain background conditions must be in place in order for the striking of the match to successfully light the match: there must be sufficient oxygen in the room, the match must be sufficiently dry, and so on. Though these additional background factors play some role in the fullest possible explanation of the match being lit, such conditions are not the operative ground in question. They instead do something to explain why the operative ground itself was indeed operative in the given context.

This may seem strange. Why should a QWA theorist find the claim that bad motives merely function as explanatory background conditions plausible enough for the purposes of this discussion? Is there not some sense in which the whole point of the QWA is to say that the quality of the agent's will is more like the striking of the match than like the oxygen in the room? However, keep in mind that, according to the Sparse Theory, it is possible that facts about an agent's quality of will do play an operative role in some token explanations of why an agent is culpable. For example, it could well be true that the malicious intent of a murderer plays a role in explaining why they are culpable for murdering someone. My point is rather that this need not be the only way for facts about an agent's quality of will to play a part in a complete explanation for why they are culpable. We can separate out a theory of excuse that involves plausible claims about how facts about an agent's quality of will can figure in as background conditions in the explanation of an agent's culpability.

Thus, one possibility is that the operative grounds of the murder's culpability are overdetermined. Both the wrong action itself (without a further quality

---

specificity. A given agent may act wrongfully in one sense and thus be culpable in the corresponding sense without it being true that they are culpable in any other important sense.

of will excuse to be discussed in the next section) and the malicious motives would be sufficient to ground the murderer's culpability. Insofar as that is true, we can also accept, as some would like to, that agents can be culpable without being culpable for a wrongful action.[11] Those are just cases when only the bad motives are an operative ground of culpability.[12]

4. THE SPARSE THEORY AND EXCUSES

Now, reconsider excuses. We can think of excuses and their role as defeaters in a similar way to the background conditions needed for the match to be lit. Excuses may, in a given context, do something to explain why the fact that a given action is morally wrong fails to ground the further fact that an agent is culpable. But again, we will flesh out this idea in greater detail in what follows. Let us now turn to the question of how thinking about the division of labor between the Sparse Theory and the idea that excuses are defeaters fits in with liberal's challenge to the moderate view.

The Sparse Theory is compatible with any number of different substantive theories of excuse. For our purposes, we will be interested in a quality of will (QOW) theory of excuse. That is, the class of facts that count as excuses, according to my QOW theorist, is characterized by facts about whether the agent under consideration had a reprehensible configuration of desires, intentions, aversions, etc. Though it should be obvious by now, this locates the explanatory importance of the quality of an agent's will in the excuse part of our theory, not in the fundamental grounds of culpability itself. This, I think, is good enough for my view to count as a kind of QWA. But it does differ in structure from what I have called Smith's QWA. I offer a QWA of excuses, not a QWA of the grounds of culpability. Smith offers a QWA of the grounds of culpability.

But how does the Sparse Theory, when combined with an adequate theory of excuse, help the moderate? Recall that, earlier, I said we should get rid of the central explanatory importance of the "gives rise to" relation. The Sparse Theory clearly does that. But how are we going to characterize the class of facts that count as excuses without helping ourselves to the "gives rise to"

11 For discussion, see, e.g., Capes, "Blameworthiness without Wrongdoing."

12 For what it is worth, considerations of parsimony tempt me to the view that wrongful action is always the only operative ground of culpability and facts about an agent's quality of will only come in on the excuse part of the theory. I think Capes's arguments can be handled by appeal to the right theory of subjective wrongness ("Blameworthiness without Wrongdoing"). However, giving that much weight to parsimony and responding to Capes in this way will both be controversial moves among QWA theorists, so they need not follow me on these further controversial claims. They are not central to the topic at hand.

relation? We are not. Here, we will appeal to the "gives rise to" relation. For some fact to excuse some agent for their wrongful action—and thereby prevent the successful grounding of culpability—that fact needs to appeal to some morally acceptable motivations that gave rise to the wrongful action under assessment.

One might ask: Why does my appeal to the "gives rise to" relation in my theory of excuse not just reintroduce the liberal's challenge under a different guise? There are two parts to my answer: (1) we are using facts about the quality of an agent's will not to explain why some agent is culpable, but to explain why they are excused; and (2) we can assess the morally acceptability of a given excuse without appealing all and only to the wrongful action. Let us take each of these points in turn.

On the first point: according to Smith, we use facts about the moral acceptability of an agent's motivations to explain why some agent is culpable. On my view, we use facts about the moral acceptability of an agent's motivations to explain why some agent is excused. This means that when I appeal to the "give rise to" relation, I am not appealing to it in order to explain why an agent is culpable. So, if I think that culpably ignorant agents are culpable to some degree for their unwitting wrongful action, the ground of their culpability still stems from the wrongness of their action. It does not stem from some causally distant motives. It is therefore not incumbent upon me to explain why causally distant motives would ground the culpability of culpably ignorant agents. When I appeal to the "gives rise to" relation, I am doing so to characterize the class of facts that count as excuses. I therefore have not incurred the sort of explanatory burdens pointed to by the liberal. And this is in part because my explanation of why an agent is culpable does not appeal to the sorts of facts it which the liberal assumes the moderate is appealing.

On the second point: I have divided out the theory of excuse from the theory of what grounds culpability. When an agent does wrong, they are culpable unless some fact—an excuse—defeats the grounding relation given by the wrongful action. Excuses are facts about a morally acceptable set of motivations that gave rise to the wrongful action in question. If someone has an acceptable and morally sincere set of motivations, then they have an excuse for having done wrong. The question that divides the liberal and moderate then becomes "What constitutes a morally acceptable set of motivations?" Part of the answer the liberal and moderate will agree on: the content of the motivations. Desires to harm undeserving people are morally unacceptable. Desires to save people are, *ceteris paribus*, morally acceptable.

But the moderate and liberal disagree about what conditions can affect the moral acceptability of an excuse outside of the content of the motivation.

Moderates think the relations that an otherwise good motive stand in to other mental states may be relevant to whether an agent's otherwise good motive is exculpating. That is, the fact that someone has an otherwise good motive counts as an excuse only if that motive fails to stand in the appropriate relations to other morally objectionable motives. We might think of these other facts about the relations otherwise good motives stand in to other bad motives as defeaters for the facts that would otherwise count as an excuse. In other words still, we can help ourselves to the idea that there is a recursive structure to defeat that tells us something about excuses.

To illustrate this rather abstract point, reconsider Anne. Anne's failure to save the child is a *pro tanto* ground of culpability. This *pro tanto* ground could potentially be defeated by the fact that Anne had a sincere motive to save the child—an excuse. But this *pro tanto* excuse is itself (partially) defeated by the further fact that Anne's motivation to save the child is combined with a belief that had been given rise to by past bad motivations. The important thing to note is that Anne's culpable motivations for her benighting act do not explain why she is culpable for her unwitting wrongful act. They explain why her otherwise noble motives are insufficient as an excuse for wrongdoing.

We can summarize this view (roughly) as follows:

> QOW *Excuses*: Some fact *F* counts as an excuse for some wrongdoing (and thereby defeats the grounding relation) only if:
>
> 1. *F* is about a motive that both give rise to ϕ and has morally acceptable content; and
> 2. There is no further fact *G* that defeats the exculpatory force of *F*.

It is important for my account that 1 and 2 are clearly separated out, for *G* need not itself stand in any special relation to over and above the extent to which *G* stands in the right kind of relation to *F*. This means that the fact that the culpably ignorant agent has a bad motive at a prior point in time needs to be suitably related to the *pro tanto* excusing fact. It does not itself need to explain anything about why the agent is culpable in the first place. The moderate therefore fails to incur the explanatory challenge they would incur if they were to instead accept Smith's QWA. We will discuss this point in greater substantive detail later in this essay.

Even if this undefended outline of a view succeeds in avoiding a commitment to the particular "gives rise to" relation of Smith's QWA, the big picture is still radically incomplete. In particular, there are three notable gaps: (1) a full account of what makes some fact the right sort of fact to defeat a *pro tanto* excusing fact, (2) an account of how to get degrees of culpability out of the Sparse Theory, and (3) a clear cut analysis of what goes on in cases of culpable

ignorance. Something will need to be said about each of these if the conjunction of the Sparse Theory and QOW Excuses is actually to deliver on the promise of meeting the liberal's challenge. We will consider each of these issue in turn, but it will be helpful to first consider an objection.

5. THE NATURE OF CULPABILITY

I said that there are two parts to a theory of culpability. The first part was to give an account of what grounds culpability. I offered such an account and supplemented it with a somewhat rough theory of excuse. To answer some of the remaining challenges I mentioned at the end of the last section, I will need to say something about the second part of a theory of culpability. The second part of a theory of culpability is the part that tells us what it is for an agent to be culpable. While I would not aim at something quite so ambitious, we can think of this second part of a theory of culpability as the part that gets at the essence of the concept or the part that provides a real definition of culpability. This task is subtly different than explaining why or when a given individual is culpable.

We should distinguish between two big-picture models of what it is to be culpable. On the *actor-focused model*, facts about culpability are fundamentally facts about the moral status of the wrong-doer or their character; culpability is a kind of moral stain on one's soul or a bad, dark mark in the ledger of one's moral character. According to the *reactor-focused model*, culpability is fundamentally about how other agents should respond to the actions of the wrongdoer in question. That is, to be culpable is to be the appropriate object of another person's blame responses.[13]

There is room for both the actor-focused model and reactor-focused model in our best moral theory, for we can simply countenance more than one concept of culpability. This opens up a possible concession to the liberal. It could be that their view is more plausible on one model and that the moderate's view is more plausible another. The liberal would get something right if this were true, but so would the moderate. My own view is that the moderate's view sits well with the reactor-focused model. Insofar as that is true, there is some important concept of culpability to which the moderate can appeal in stating their view. That is good enough for the purposes of implementing my strategy.

13 These two models go by different names throughout the literature on culpability. The terms I introduce here are my own, which I prefer because they bring out the contrast between the two positions more clearly than other terminology does. For a helpful discussion of how various philosophers have thought about the two models, see the earlier chapters of McKenna, *Conversation and Responsibility*.

And this is so even if we conceded that the liberal's view looks more plausible on the actor-focused model, for it secures the claim that the moderate is on to something deep and important about the nature of moral culpability.

Alright, let us get a toy reactor-focused model on the table. We can start with:

> FA-*Culpable*: *S* is culpable for an action if and only if there is reason to have the appropriate reactive attitude toward *S* in virtue of.[14]

A few clarifications: I use the phrase "appropriate reactive attitude" merely to remain agnostic about whether the blame-constituting attitude is resentment, anger, disapproval, some combination of these, or some other attitude or combination of attitudes. Moreover, in the spirit of the Sparse Theory, I will assume that we have a *pro tanto* reason to have the appropriate reactive, blame-constituting attitude toward a given individual only if that person has acted wrongly.[15] The fact that someone has done something wrong plays the role of being a *pro tanto* reason for having the appropriate reactive attitude toward them in many if not most cases. However, this reason can be defeated. Therefore we should think that:

> *C-Defeat*: There is a *pro tanto* reason to have the appropriate reactive attitude toward an individual in virtue of an action if and only if is wrong and this reason is not undercut by other morally relevant considerations.

Though it is stated as a bi-conditional, we need not read this claim as a reductive or real definition. The recursive part of the claim therefore should not really bother us too much.

And finally we can plug this all directly back into the Sparse Theory:

> *The Sparse Theory**: The fact <there is a *pro tanto* reason to have the morally appropriate reactive attitude toward *S* in virtue of an act *A*> is explained by the fact that <act *A* is objectively wrong>.

14 This statement takes inspiration from the kind of neo-Strawsonian approach in Wallace, *Responsibility and the Moral Sentiments*. This statement may need to be amended further. For example, perhaps not just any old reason will do. If an evil demon threatens to blow up the world if you do not have a particular reactive attitude toward Felicity when she has done nothing wrong, you have reason to have that reactive attitude. However, intuitively, this does not mean that Felicity is culpable. It is possible, then, that we may need to add a further constraint like the reason in question makes a particular attitude "fitting." I take it that any reactor-focused model will require this sort of caveat, and this amendment and others like it are not an *ad hoc* fix for the moderate view *per se*. After all, reactor-focused liberals and conservatives would face cases like this as well. I thank an editor of this journal for pressing me on this point.

15 The appropriate attitude constituting praise or admiration plausibly has nothing to do with acting wrongly. Thus, acting wrongly is only relevant to culpability. Thank you to an editor for pushing me to clarify this point.

And we can understand QOW Excuses more precisely as follows:

> QOW *Excuses**: Some fact *F* is an undercutting defeater for our reasons to have the morally appropriate reactive attitude toward *S* in virtue of a wrongful act *A* only if
>
> 1. *F* is about some motive with morally acceptable content that give rise to *A*; and
> 2. There is no further fact *G* that defeats the exculpatory force of *F*.

With this version of the Sparse Theory and QOW Excuses in tow, we can consider the three remaining gaps in the Sparse Theory*: (1) an account of what makes some fact the right sort of fact to defeat a *pro tanto* excusing fact, (2) an account of how to get degrees of culpability out of the Sparse Theory, and (3) a clear-cut analysis of what goes on in cases of culpable ignorance. In the next several sections I address these three gaps in order.

Before we move on, however, let us consider why excuses are best understood as undercutting defeaters.[16] The main alternative within the framework I have been developing would be that excuses are rebutting defeaters. Rebutting defeaters can be thought of, roughly, as reasons that outweigh other reasons. If there is a rebutting reason against having the morally appropriate reactive attitude toward an agent in a given case, this would not imply that we lack a *pro tanto* reason for having the morally appropriate attitude toward an agent in virtue of their wrongful action. We would simply lack all-things-considered reason to blame the person in question. However, for many excuses, it seems false that I have any reason whatsoever to blame someone. To illustrate: suppose Erwin presses the doorbell of his friend's house, which he has pressed a number of times before. Little does Erwin know that the doorbell has recently been wired to trigger an explosion thousands of miles away, which will kill dozens. Erwin certainly is not acting from a bad QOW. How could he be? Erwin has no clue that the doorbell has been wired to trigger an explosion. No reasonable person could foresee this. Do we have any reason—even one that is outweighed—to blame Erwin? Plausibly not. In a purely objective sense, however, Erwin acts wrongly. The best thing for the advocate of the Sparse Theory* to say, then, is that whatever excuse is operative in Erwin's case does more than simply outweigh the reason we have to blame Erwin (in virtue of his objectively wrong action). Indeed, plausibly there is no such reason to blame Erwin. But how could this be?

16 My use of the distinction between "undercutting" and "rebutting" defeaters follows the generalization of the distinction from epistemic reasons to all normative reasons proposed in Schroeder, *Slaves of the Passions*, ch. 7. I thank an associate editor of this journal for pushing me to address my appeal to undercutting defeat.

Undercutting defeaters work differently than rebutting defeaters. Undercutting defeaters "remove" or "disable" our reasons that might otherwise exist were those defeaters not present. Erwin's excuse, plausibly, is an undercutting defeater. After all, this would explain why Erwin's act would meet the grounding condition of the Sparse Theory* and yet there is no reason to blame Erwin. The reason that would otherwise be grounded has been undercut or removed by Erwin's excuse. In what follows, I will often use the term "defeater" to mean "undercutting defeater" in particular.

6. THE CONCERN CONSTRAINT

I aim to defend the moderate against the liberal's challenge. I have sketched a theory of what grounds culpability and a theory of excuse. This theory of excuse, if it is to be useful for the moderate, needs to explain why some facts count as defeaters for *pro tanto* excuses and why others do not. If this cannot be done successfully, then the moderate is still in trouble. I will not attempt to settle the question of what distinguishes the relevant facts. I will merely attempt to show that the moderate has something plausible to say.

Philip Robichaud and Jan Willem Wieland have offered the beginnings of a response to the liberal's challenge. They divide the task into two papers.[17] In one, they argue that it does not follow from Smith's articulation of the liberal's challenge that blame fails to transfer across the morally relevant relations.[18] If by "follows" they mean "deductively follows," then surely they are right. But as far as I can tell, that fails to cut to the heart of the liberal's challenge. The liberal's challenge is explanatory. It therefore requires an abduction. And such forms of argument are perfectly good even if they are not deductive.

Of course, this still leaves their second essay. In that essay, Robichaud and Wieland defend the following principle:

> *Concern Constraint*: B_1 transfers to B_2 only if the benighting act expresses a deficit of concern for the same consideration in virtue of which the unwitting act is wrong.[19]

B_1 and B_2 represent the culpability for the benighting act and unwitting act respectively. The Concern Constraint embodies a pretty plausible way to restrict the proper scope of the Transfer Model. And Robichaud and Wieland argue as much.

17 See Robichaud and Wieland, "Blame Transfer" and "A Puzzle concerning Blame Transfer."

18 See Robichaud and Wieland, "Blame Transfer," 296.

19 See Robichaud and Wieland, "A Puzzle concerning Blame Transfer," 17.

But the Concern Constraint does not supply the sort of explanation needed in response to the liberal's challenge. The liberal wants a theory-neutral explanation of why the Transfer Model should be accepted. The worry is that, without such an explanation, the Transfer Model just looks like a dressed up restatement of the moderate intuition. What Robichauld and Wieland have provided us with is the version of the Transfer Model that is the least susceptible to counterexample. While helpful, this is not an explanation of why we should accept the Transfer Model in the first place. Once again, consider Smith:

> Of course, it is true that at an earlier time, the time of the benighting act, the agent had a reprehensible configuration of desires—a configuration that typically included a willingness to risk eventual wrong—doing of exactly the sort exemplified in the unwitting act. But the fact that he earlier had faulty motives does not show that he now has faulty motives.[20]

I do not read this passage as pinning the moderate with a counterintuitive implication about who is culpable. It is an attack on the quality of the moderate's explanation of their view. And even if I am wrong about the exegetical point, not much changes. I have simply misinterpreted the liberal challenge in a constructive way: the moderate now faces a new explanatory challenge to which Robichaud and Wieland do not respond.

Robichaud and Wieland might reply that Smith does provide counterexamples. This is true, but we must be careful. Smith employs counterexamples, in my reading, to show the inadequacy of a variety of proposed explanations of the Transfer Model. She does not provide counterexamples to the Transfer Model as such. Her objection to the Transfer Model—and therefore the moderate view—is the aforementioned explanatory objection.[21]

Robichaud and Wieland might reply to this point by claiming that what they did was offer an explanation of the Transfer Model that was not open to counterexample. They therefore did provide a response to Smith. However, they would be wrong to make this response. The Concern Constraint is aptly named. It reads like a constraint on candidate blame transfers, not an explanation of why blame would transfer from an earlier event to a later event. Moreover, when Robichaud and Wieland offer an intuitive gloss of the Concern Constraint, they actually assume what needs explaining:

> If blameworthiness is to transfer from B_1 to B_2, then there must be a match between the kinds of reasons for which the agent shows

20 See Smith, "Culpable Ignorance," 559.

21 Smith confirms this in correspondence.

> diminished concern in the benighting act and the kind of reasons that underwrite the wrongness of the unwitting act.[22]

It is the antecedent of their conditional that the liberal is attacking, not the consequent or the conditional itself. The question is why we should be committed to the idea that culpability transfers. Again, we cannot answer this question by simply pointing to our intuitions about culpably ignorant agents because the issue of whether culpably ignorant agents are culpable for their benighting acts is what is at issue.

Here is a different response that Robichaud and Wieland could make. They could follow Daniel Miller's recent suggestion that

> an agent's degree of blameworthiness for some action (or omission) depends at least in part upon the quality of will expressed in that action, and an agent's level of awareness when performing a morally wrong action can make a difference to the quality of will that is expressed in it.[23]

Miller's suggestion seems to make progress insofar as it allows us to point to a mental state possessed by culpably ignorant agents at the time they perform their unwitting wrongful actions. Given that Anne is at least nonoccurrently aware that she may have missed some lifesaving information when she performs adult CPR on the infant, then perhaps this counts as a lack of concern.[24]

It would be strained to claim that Anne's nonoccurrent awareness in any meaningful sense "gives rise to" her giving the infant adult CPR. While the state is present in some dispositional sense, it hardly seems to be giving rise to much of anything. So it is unclear whether the mere dispositional presence of this lack of awareness fully explains why the liberal challenge has been met. The general problem is that the liberal has a view about the explanatory role a mental state needs to play. The problem is not about the type of mental state or its content.

Miller might push back. He might say that the non-occurrent awareness is expressed in an "indirect" way.[25] Robichaud and Wieland, similarly, distinguish between "distal" and "direct" motives, writing:

22 Robichaud and Wieland, "A Puzzle concerning Blame Transfer," 15, emphasis added.

23 See Miller, "Circumstantial Ignorance and Mitigated Blameworthiness," 34. Thank you to an anonymous reviewer for bringing this worry to my attention.

24 One salient concern is what we should say about agents who forget about their benighting acts. Miller offers some response in "Circumstantial Ignorance and Mitigated Blameworthiness," 38–39. It is unclear, though, whether his response would apply to Anne were she to forget about sneaking out of the lifeguard training.

25 See Miller, "Circumstantial Ignorance and Mitigated Blameworthiness," 37–38.

> We agree that transfer is problematic if, following Smith, *S* is blameworthy for the unwitting *A* only if *A* expresses a deficit of concern on *S*'s part at the time of *A* (which constitutes a direct motive).... [The Concern Constraint] merely requires that the benighting act that led to *A* expresses a deficit of concern (which constitutes a distal motive).[26]

Thus, Robichaud and Wieland—and plausibly Miller—may be unimpressed by my table pounding about the explanatory impotence of Anne's non-occurrent awareness. So long as there is a distal motive expressed in a culpably ignorant agent's unwitting wrongful act, their non-occurrent awareness need not give rise to that unwitting wrongful act.

Which motives count as "distal" versus merely in the past and unrelated? Robichaud and Wieland seem to think that it is (1) those motives that meet their concern constraint and (2) those suitably related to the expression of a deficit of concern in question. But I cannot see how Anne's attempt to save the drowning infant's life "expresses" a deficit of concern. It is not intuitively obvious and I am not willing to take it on faith that such an expression occurs. Perhaps there is some non-question-begging reasons to affirm that such an expression occurs. However, Robichaud and Wieland do not offer any.[27] Any attempt to provide such reasons seems to reintroduce the liberal's challenge: How do we get an "expression" of the intuitively relevant kind of concern without the "gives rise to" relation?

In summary, Robichaud and Wieland's defense of the Concern Constraint is an admirable contribution to those of us who wish to defend the moderate view. I will appeal to the Concern Constraint myself later on. However, it is not by itself an answer to the liberal's challenge. In fairness to Robichaud and Wieland, it is unclear whether they would claim to have offered a full reply to the liberal's challenge. Perhaps they never intended to. In that case, we may well be allies for the purposes of this paper since we both would then recognize that the moderate needs to do more work.

The Concern Constraint may help us do that work. We can adapt it so that it fits in with the conjunction of the Sparse Theory* and QOW Excuses. Here is a stab at it:

> *Concern Constraint**: A fact *F* defeats a putative excusing fact for a given unwitting wrongful act *A* only if *F* is about a benighting act that

26 Robichaud and Wieland, "A Puzzle concerning Blame Transfer," 24. Thank you to an anonymous reviewer for pressing this response on behalf of Robichaud and Wieland.

27 Robichaud and Wieland do argue that one reason Smith provides for accepting a timing constraint can be diffused; see "A Puzzle concerning Blame Transfer," 24. However, this does not constitute an explanation of why they are licensed to claim that the right kind of expression occurs in Anne's action.

> expressed a deficit of concern for the same consideration in virtue of which *A* is wrong.

And we might then reformulate QOW Excuses* as follows:

> CC *Excuses*: Some fact *F* is a defeater for our reasons to have the morally appropriate reactive attitude toward *S* in virtue of a wrongful act *A* only if
>
> 1. *F* is about some motive with morally acceptable content that gives rise to *A*; and
> 2. There is no further fact *G* about a benighting act that (i) expresses a deficit of concern for the same consideration in virtue of which *A* is wrong and (ii) at least partially in virtue of *F* now obtains.

The moderate may wish to eventually generalize 2 so that it sounds a bit less *ad hoc*. By invoking the term "benighting act," the condition admittedly sounds as if it is crafted to help with cases of culpable ignorance in particular. I, of course, doubt that 2 actually is *ad hoc*. All 2 tells you is how to distinguish between which facts are and which facts are not second-order excuse defeaters—defeaters for putative excusing facts. One would just want an economical way of eliminating "benighting act" language.

One might be worried that clause ii of condition 2 reintroduces the "gives rise to" relation, which I have claimed is the source of the Liberal's Challenge. This is not so. The "gives rise to" relation was a relation between a motive and a wrongful act. The "in virtue of" relation I discuss here is a relation between two kinds of facts. One kind of fact, *F*, is a putative excuse. The other kind of fact, *G*, is about a motive relating to *F*. All *G* needs to do is partially explain why *F* obtains in a metaphysical sense while citing a deficit of concern. And this is because *G* merely explains why *F* fails to be a good excuse. *G* does not explain why a given agent is culpable for *A*. The wrongness of *A* primarily explains why the given agent is culpable, and—along with the fact that there is no relevant *F*—it thereby fully explains why the given agent is culpable. Nothing about this story requires that the moderate believes that *G* gives rise to *A*. Since such facts do not enter the moderate's explanatory story, the moderate needs not explain such facts. In other words still, *G* explains why *F* is a bad excuse, not why a given agent is culpable for *A*.

I take it that CC Excuses fills our first gap. It tells us what makes some fact the right sort of fact to defeat a putative excusing fact. Or at least, it is a reasonably plausible first stab. It shows that my strategy to evade the liberal's challenge by offering an alternative package of views about the nature and grounds of culpability is not dead on arrival.

7. DEGREES OF DEFEAT

Let us now turn to our second gap. The moderate needs an account of how to get degrees of culpability out of the Sparse Theory*. But why does the moderate need an account of how to get degrees of culpability out of the Sparse Theory*? Well, because they believe that culpably ignorant agents are often partially, but not fully, culpable for their unwitting wrongful acts. And for all I have said so far, the Sparse Theory* and CC Excuses do not look to accommodate degrees of culpability. We will therefore need to offer some further revisions to the theory. (Notice, of course, that conservatives could appeal to everything I have said in defense of the moderate thus far, but not care about offering an account of degrees of culpability. If no good theory of gradable defeat is found, we therefore are not forced to the liberal view.)

We can supply a gradable theory of culpability by offering an account of partial defeat. There is more than one way to do this. We could, as others have, invoke some form of non-monotonic logic in order to capture the gradable structure of defeat.[28] But since our task is sufficiently simple, we need not deal with the intricacies of non-monotonic logics. Instead, we can build up our theory of partial defeat with a bit of simple math and the idea that degrees of culpability ebb and flow with the strength of the reasons we have to blame others for their wrongful actions.[29]

We can start by determining the strength of a given reason, *R*, to have some morally appropriate reactive attitude toward an agent in virtue of their wrongdoing. This reason can be formally represented by a tuple:

$$\text{Reason} = \langle F, A, T \rangle.$$

F is a fact that stands in favor of an agent, *A*, performing action of act type *T*. (It is worth noting that formally representing reasons with this tuple need not commit us to the claim that reasons are tuples. That is a metaphysical thesis that is likely false and we do not have the space to discuss.) The strength of a reason is a function of the members of said tuple. Let us represent this with:

$$\text{Strength of Reason} = S_{\langle F, A, T \rangle}.$$

We might think that $S_{\langle F, A, T \rangle}$ is also in some way a function of the degree of wrongness or badness of the action performed by the agent. Next, we can multiply $S_{\langle F, A, T \rangle}$ by a defeat function, $D_{\langle F, A, T \rangle}$ for the given reason where

28 See especially Horty, *Reasons as Defaults*; and Bonevac, "Defaulting on Reasons."

29 For an alternative model of defeat that relies only on orderings, see Schroeder, *Slaves of the Passions*, ch. 7.

$1 \geq D_{<F,A,T>} \geq 0$. This allows us to then claim that the degree to which an agent is culpable for an action T can be represented as follows:

$$\text{Degree of Culpability} = S_{<F,A,T>} \times D_{<F,A,T>}.$$

When a defeater fully defeats the given reasons, $Di = 0$. If we have defeaters for defeaters—as I have proposed we should in the case of culpably ignorant agents—we can understand Di as the following function:

$$\text{Strength of Defeater} = D_{<F,A,T>} = 1 - D_{D<F,A,T>},$$

where $1 \geq D_{D<F,A,T>} \geq 0$. Here $D_{D<F,A,T>}$ is the strength of a defeater for the defeater $D_{<F,A,T>}$. We can determine the strength of $D_{D<F,A,T>}$ by the same function. We can continue to iterate this embedded function *ad infinitum* if need be; the model therefore embodies a kind of recursive structure whereby the strength of each defeater depends on, *inter alia*, the strength of the further defeaters. Consider an example. Suppose $D_{D<F,A,T>} = 0$. If this is true, then the strength of all of the defeaters for $D_{<F,A,T>}$ will equal 1. That is, $D_{<F,A,T>}$ will equal 1. But if $D_{D<F,A,T>}$ equals 0.3, this is because the defeaters for $D_{<F,A,T>}$ equal 0.7. We can therefore see how this function captures the recursive structure of defeaters.

Let us throw one last widget into our model just to show how malleable it is. We might think that sometimes more than one fact is relevant to an agent's decision on whether to perform some action. Moreover, it could be that the agent is ignorant of several facts. And she could be culpable for her ignorance of each fact to different degrees. How would we represent that? We can represent this simply by modifying the view as follows to account for multiple defeats for a reason, $<F,A,T>$:

$$\text{Degree of Culpability}^* = S_{<F,A,\Phi>} \times \prod_{D_i \in G} D_i.$$

G is the set of all defeaters for $<F,A,T>$ and D_i represents each member defeater of G. The function takes the product of the strength of all such defeaters and then multiplies this product, which will be between one and zero, by the strength of the reason under consideration. The recursive structure of each defeater in the set is maintained as long as we claim that the strength of each defeater in G is determined as we suggested before:

$$D_i = 1 - D_{D_i}$$

There are surely further complications to consider and further ways to tweak the model. For the purposes of this essay, we only need a simple version of the model. The important point is that we can get a gradable structure out

of thinking about culpability in terms of the strength of reasons and partial defeaters that affect these strengths.

Of course, plenty of further, more substantive questions remain. Most importantly: With which gradable properties do the degrees of defeat ebb and flow? This may well be a subject for an entire book. The liberal, conservative, and moderate all need an account of how the strength of excuses might affect how culpable an agent is. That is, some excuses are better than others and therefore exculpate to a greater degree than others. Since everyone needs an account of how to determine the strength of excuses (and therefore first-order defeaters for culpability), I will set this question to the side. But the moderate needs something more: a substantive account of how to determine the strength of those defeaters that undercut excuses. If there were no plausible story on offer, then the moderate would be in trouble.

Luckily, there are plausible candidates. Reconsider Anne. Suppose Anne justifiably had a credence of 0.001 that important, lifesaving information would be shared during the time that she sneaked out for a smoke. Anne would be culpable to some degree (according to the moderate), but it is plausible to think that she would have been more culpable had her credence been 0.5. To wit, the strength of the defeater in this case may be proportional to the degree of undue risk Anne ran by performing her benighting act. Why would the strength of a defeater be proportionate to the undue moral risk run by the agent in performing their benighting act? We might think that our reasons to blame Anne for her benighting act are proportional to the risk she ran. As her actions become riskier, we will have stronger reasons to blame her. And it is then plausible to generalize from our point about risk to say:

> CC *Defeater Strength*: The strength of a given defeater for a putative excuse is proportionate to the strength of the reasons we had to blame the agent for their benighting act.

More work would need to be done in order to show that this would work as a general principle, but it does point us in a nice direction. Determining the requisite strength of a defeater by appealing to facts about the strength of reasons we have to blame agents for their benighting acts looks to be nicely principled.

I take it that by now I have done enough to fill the second gap in my alternative QWA of culpability. I have shown how we can think about the structure of partial defeaters in our theory of culpability. This brings us considerably closer to a sufficiently plausible theory of culpability—and one that side-steps the explanatory problems the moderate would incur were Smith's QWA the only account on offer. We can now turn to the final gap to fill in the theory.

8. THE APPRAISAL SYMMETRY

The third and final gap that needs filling is a clear-cut analysis of what goes on in cases of culpable ignorance. One might think that this is easy enough to supply given all of the machinery I have laboriously laid out in the preceding pages. The simple version of the story goes like this. Anne performed a wrong action by failing to save the drowning infant. This would usually make her act culpable. However, she had a putatively acceptable motive for her unwitting wrongful act: she wanted to save the child's life. This fact about Anne's motivations is a putative excuse and therefore—absent other defeaters—undercuts our reasons to blame Anne. However, there is another undercutting defeater in play: Anne's unwitting wrongful act is due, in part, to her taking an undue moral risk that was wrong for the same reasons that her unwitting wrongful act was wrong. This second-order defeater partially undercuts the exculpatory force of Anne's QOW excuse. Anne's QOW excuse then only partially defeats our reasons to blame Anne in virtue of her unwitting wrongful act. Anne is therefore culpable to only some degree and the moderate's view has been secured by my model.

Of course, we can analyze the case in this way within the framework I have developed. But what reason is there to think that this is a good way of analyzing the case of culpably ignorant agents? I think that the analogues of the Sparse Theory* and QOW Excuses* give us a better explanation of widely shared intuitions about a class of praiseworthy agents. Insofar as we think that our theories of culpability and praiseworthiness are more plausible when structurally symmetrical, this will lend non-negligible support to the view I have developed in this essay.

Consider Saintly Jack. Jack began his life with the desire to do the most good he could. Whenever someone was in trouble and he knew how to help them, Jack would rush to their aid. His only regret was that he could not do more good. So Jack set out to find the best way to help others. After a ton of high-quality research, Jack came to the conclusion that he will reliably help the most people if he instills a disposition in himself to always act from selfish motives. And Jack was right. As a result, Jack very reliably performs the right action at every juncture, though from entirely selfish motives. Further suppose that Jack one day comes upon Jill who is drowning in a pond. Jack thinks to himself, "While I would ruin my new shoes, I could get a substantial reward for saving her!" He then jumps in and pulls Jill out of the pond, saving her life. Is Jack praiseworthy?

According to my intuitions: yes—Jack is praiseworthy to at least some degree. I think this intuition can be felt further if you compare Jack to Tom. Tom just acts from selfish motives. He always has and always will. Jack strikes me as being more praiseworthy for saving Jill than Tom would be for saving

Jerry in similar circumstances. Now this need not be by a huge degree. So long as one gets the intuition at all, I am in business.

This makes perfectly good sense on the symmetrical version of my view. Jack acted rightly. *Ceteris paribus*, this should give us reasons to praise him. However, Jack acts rightly for the wrong reasons. The fact that <Jack performs the right action for the wrong reasons> undercuts our reasons to praise him so long as no second-order defeaters are in play. But, in this case, there is a second-order defeater: the fact that <Jack instilled a disposition to act from selfish motives so that he could help more people>. Moreover, we can imagine that the reasons that justify Jack in instilling a disposition to act from selfish motives are the reasons that justify him in saving Jill's life. This second-order defeater at least partially defeats the first-order defeater given by Jack's selfish motives. And this in turn means that our reasons to praise Jack for saving Jill have not been fully defeated. So a sparse QWA theorist about praiseworthiness seems to capture our intuitions nicely in this case.

Let us now contrast the analogue of my view for praiseworthiness with an analogue of Smith's view and see whether it can capture the relevant intuition:

> *Smith's Praise* QWA: The fact <*S* is praiseworthy for performing act *A*> is explained by the facts that
>
> 1. <Act *A* is objectively right>,
> 2. <*S* had a noble configuration of desires>, and
> 3. <This configuration gave rise to the performance of *A*>.

It should be immediately clear that Jack fails the second condition. He does not have a noble configuration of desires at the time he is acting. Or, if one prefers to understand Jack as having some kind of noble second-order desires, Jack would still fail the third condition since this second-order desire did not give rise to his action. By stipulation, a selfish motivation gave rise to his action. Jack cannot therefore be praiseworthy to any degree on such a view. I think this will run afoul of most people's intuitions more starkly than analogous cases of culpable ignorance.

But let us be careful: my point is not that the liberal now faces some new challenge. My point is more simply that the framework developed in this essay can be modified to easily analyze cases of praiseworthy agents like Jack. Insofar as a symmetrical account of praise and blame is attractive, my model seems to provide the tools for a plausible, general analysis for appraising all kinds of agents: culpable agents who are culpable due to the origins of their ignorance and praiseworthy agents who are praiseworthy due to the origins of their motives. And with that, we have filled the third and final gap.

I have thus defended the moderate view against its most prominent challenge. I have done so by offering a novel account of how facts about an agent's

QOW can sometimes function in explanations for why agents are excused. This view is compatible with the liberal, moderate, and conservative accounts of culpable ignorance. This suggests a change in the dialectic: if the liberal wishes to argue against the moderate, they need a new objection or to show that the view developed here suffers from some fatal flaw. At least for the time being, the moderate is off of the explanatory hook they have been hanging on for the better part of the last forty years.

University of Wisconsin–Madison
jpgoodrich@wisc.edu

REFERENCES

Bonevac, Daniel. "Defaulting on Reasons." *Noûs* 52, no. 2 (June 2018): 229–59.

Capes, Justin. "Blameworthiness without Wrongdoing." *Pacific Philosophical Quarterly* 93, no. 3 (September 2012): 417–37.

Horty, John. *Reasons as Defaults.* Oxford: Oxford University Press, 2012.

Husak, Douglas. *Ignorance of Law.* Oxford: Oxford University Press, 2016.

McKenna, Michael. *Conversation and Responsibility*. Oxford: Oxford University Press, 2012.

Miller, Daniel. "Circumstantial Ignorance and Mitigated Blameworthiness." *Philosophical Explorations* 22, no. 1 (2019): 33–43.

Robichaud, Philip, and Jan Willem Wieland. "Blame Transfer." In *Responsibility: The Epistemic Condition*, edited by Philip Robichaud and Jan Willem Wieland, 281–98. Oxford: Oxford University Press, 2017.

———. "A Puzzle concerning Blame Transfer." *Philosophy and Phenomenological Research* 99, no. 1 (July 2019): 3–26.

Schroeder, Mark. *Slaves of the Passions*. Oxford: Oxford University Press, 2007.

Shoemaker, David. "Qualities of Will." *Social Philosophy and Policy* 30, nos. 1–2 (January 2013): 95–120.

Smith, Holly. "Culpable Ignorance." *Philosophical Review* 92, no. 4 (October 1983): 543–71.

———. "Tracing Cases of Culpable Ignorance." In *Perspectives on Ignorance from Moral and Social Philosophy*, edited by Rik Peels, 95–119. New York: Routledge, 2016.

Wallace, R. J. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press, 1994.

Journal of Ethics and Social Philosophy
VOL. 25, NO. 2 · AUGUST 2023
https://doi.org/10.26556/jesp.v25i2.1561

WHO DO YOU SPEAK FOR? AND HOW?

ONLINE ABUSE AS COLLECTIVE SUBORDINATING SPEECH ACTS

Michael Randall Barnes

> Internet trolls, predominantly anonymous posters, realized they could work together to try to destroy the lives of people who disagreed with them.
>
> —Ian Sherr and Erin Carson, "GamerGate to Trump"

THIS PAPER is about online abuse. I have two goals in directing our attention here. First, I want to show that this is a serious but neglected area of subordinating speech and that social philosophers of language have good reason to pay more attention to the specific harms of online discourse. Second, I will argue that accounting for the realities of online abuse shows that speaker authority—the thing that makes harmful speech harm in the way it does—is dynamic and emergent and often depends on the broader community of both audiences and other "speakers" in ways that current theories are ill equipped to explain.[1] I argue that much of online abuse is best understood as a type of *collective subordinating speech act*, where this collective is an *ad hoc* group that constitutes itself through speech, and it is (partly) this group that gives online abuse the subordinating force that it has. Overall, my hope is to show that attention to online abuse is useful both for illuminating the harmfulness of that important phenomenon itself and also for clarifying features it shares with "in real life" (IRL) hate speech that regularly go underemphasized in the existing literature.

It is not controversial to say that a lot of harmful speech now occurs online. Yet much of the philosophical work in this area has focused on offline life. This immediately raises two questions: (1) Can current accounts of oppressive speech adequately capture digital hate? (2) How does the (perceived) anonymity of (many) online harassers contribute to the force of their speech? To answer

1 A quick note on the term "speaker," which is a bit ill fitting for online contexts. A more accurate term may be "poster" or "user." However, throughout, I mainly use "speaker," as the alternatives are not perfect either, and because I aim to make a contribution to the speech act theory tradition, which tends to use "speaker."

these questions, I argue that the combination of anonymity and shared language offers online abusers a path to a type of group authority that lends more power to their speech than they might first appear to have. While most abusive messages online—tweets, emails, direct messages (DMs), and the like that harass, threaten, or otherwise potentially harm their targets—are uttered by individual users acting for myriad reasons, I claim that the cumulative effect of receiving dozens, hundreds, or even thousands of these messages impacts the force of these speech acts in a significant way, backing them up with a unique type of authority and making them unlike offline hateful speech. Thus, I argue that online abuse is best understood as a type of collective subordinating speech act. In other words, online abusive speech is a form of subordinating speech where the "speaker" of these messages is better conceived as a collective, though often an *ad hoc* one.[2]

To make this argument, I explore the popular model that claims that speakers can gain subordinating authority through processes like *licensing* and *accommodation*. The basic idea is that while a hate speaker can lack the necessary authority to subordinate before they make their utterance, because of the silence of bystanders, the audience fails to block the speaker's speech act, imbuing it with subordinating force. This approach has proven quite fruitful at explaining the outsized harm a seemingly powerless individual can achieve through their speech. Yet I argue that it fails to explain the dynamics of online abuse and that this failure reveals a more widespread tension in the concept.

I begin in section 1 by outlining some ways in which internet speech is different from noninternet speech. This will, in many ways, be fairly familiar to most readers, but it is worth making explicit as these features shape our speech acts in profound ways but too often fall from view. After this general overview of the distinctiveness of online speech, I then describe, in section 2, some of the key features of my narrower topic: online abuse. In section 3, I explain how these features pose a problem for existing accounts of subordinating speech, particularly around the notion of authority. This leads me to develop an alternate conception of the subordinating authority at work in online abuse. Section 4 is devoted to developing this idea, focusing on (1) the role of anonymity and

2 My use of the term "collective subordinating speech" is a bit different from Anthonie Meijers's use of "collective speech acts" (in "Collective Speech Acts"). This discrepancy is worth clearing up right away. In short, Meijers follows a broadly Searlian framework, and for that reason explains collective speech acts in terms of collective intentions. For my part, I am interested in uncovering the authority conditions that affect the force of particular speech acts, rendering them harmful—hence my inclusion of "subordinating" in the term. Because of this harm-centric approach, I am more concerned with identifying speech acts that an audience might, for various reasons, take to be representative and backed up by a group of agents, and give them uptake that reflects this perception. But this need not be tied to collective intentions, so I do not adhere to Meijer's analysis.

(2) the use of shared language in constructing and sustaining this distributed and collective speaker authority.[3] Here I include some considerations about how this conclusion could impact the types of mechanisms social media platforms use to mitigate the harms of abuse on their platforms.

Throughout, I argue that much online abuse challenges existing accounts of subordinating speech. It, therefore, represents in some ways a distinct phenomenon. At the same time, though, I believe this analysis can also shed some light on IRL subordinating speech. That is, I aim to show how online speech makes explicit many features that it shares with offline hate speech but that tend to be ignored or de-emphasized in existing accounts. Despite the internet offering bigots and abusers new tools and strategies, for the victims, the experience of being targeted by such abuse can be remarkably similar. The examination of online abuse, therefore, helps reveal key features of subordinating speech across mediums. I make these connections explicit in my conclusion.

1. INTERNET SPEECH: IT'S DIFFERENT

To state the obvious, online speech is different from offline speech. Terms like IRL, "meat space," and others make plain what we all know at a moment's reflection: what occurs online and through our screens is different and distinct from what happens outside of those parameters. This is not to claim, though, that "Twitter isn't real life." Far from it, my position is that what we do online is just as real and significant as our offline actions but that we must appreciate the differences the medium presents.

For starters, unlike standard in-person speech, online speech is mediated by an immense infrastructure of cables, wires, servers, satellites, modems, internet service providers, electricity grids, data networks, computers and smartphones, and so much more that is at the same time incredibly obvious as well as somewhat hidden from view. This infrastructure plays a role in determining *who* is able to perform online speech, as well as how early users often set the tone for acceptable behavior long after a much larger and more diverse group of users comes online. The fact that we can trace a line from the early history of "trolling" to current tactics in online harassment suggests a lineage from the sociological history of the internet to some of the problems we now face.[4]

3 For consideration of how similar features are at play in offline contexts for some types of propagandistic hate speech, see Barnes, "Presupposition and Propaganda." For consideration of the protest speech of social movements, see Barnes, "Positive Propaganda and the Pragmatics of Protest."

4 For accounts of early trolling, see, for example, Phillips, "The Oxygen of Amplification"; Bartlett, *The Dark Net*; and Quinn, *Crash Override*. And for brief philosophical analyses

And beyond the physical infrastructure of the internet, along with its economic history, we must also acknowledge that the platforms that currently host the bulk of online speech—Meta Platforms, Google, Microsoft, Amazon, and Apple—make decisions that shape the contours of online speech. Perhaps most significant is the invisible and opaque algorithmic amplification and moderation that each platform employs.[5] However, more mundane aspects like the default settings about public and private profiles, who can send DMs to whom, message-length restrictions, image capabilities, limits on sharing, forwarding, or replies, and much more are all features that have concrete impacts on what speech acts are possible in online environments.

At this point, it must be admitted, though, that the internet is a big place and that different platforms offer different affordances.[6] So, with an admission that none of the following is universally true for all online speech, let us consider some further distinguishing features of *much* of how we now communicate over the internet.

First, a lot of online speech, as written text, is in an important way less embodied than offline speech—or at least *differently* embodied. Our texts, tweets, emails, and the like usually occur within a small screen that we interact with mainly via our thumbs and fingers. This fact is both banal and significant. It has the consequence that when reading the words of another, we can experience it as a voice within our head, perhaps in our own voice, rather than as speech directed at us from the actual lips of another agent. Talking with another becomes, in some cases, talking with oneself.

Additionally, most online speech is asynchronous, at least to an extent. This sits along a spectrum, with some formats (such as email and message boards) on one end and other nearly but not quite real-time formats on the other. But even supposedly instantaneous platforms (e.g., WhatsApp, Zoom) admit to delays, outages, and buffering that manage to interrupt what we might think of as the "normal" flow of a conversation. The effect is that entirely different norms take hold when we cannot rely on the immediate feedback of our interlocutor, even when using supposedly "live" chat applications. Simple things like how long it is appropriate to wait before following up are norm-governed practices impacted by features like read receipts and time stamps.

---

of trolling, see Barney, "[Aristotle], On Trolling"; and Cherry, "Twitter Trolls and the Refusal to Be Silenced."

5 See, for example, Tufekci, *Twitter and Teargas*; Noble, *Algorithms of Oppression*; and Lynch, *The Internet of Us*.

6 For a recent account of the notion of affordances, see Davis, *How Artifacts Afford*.

Anonymity or pseudonymity is an often-cited feature of online communications.[7] This too is best conceived along a spectrum, and one with multiple axes. While at times we may be speaking anonymously to the other participants on a forum, this does not imply that we are anonymous to the site's host. It comes in degrees, from the relatively rare total anonymity one might have in certain parts of the web, to the anonymity of a screen name that does not easily lead back to one's offline life. And I include here the kind of anonymity one finds in a crowd even when they use their real name.

There is also the ambiguous state of the audience that is typical of so much online speech. Social media posts are often characterized by a genuine uncertainty about to whom one can be said to be speaking. One's tweets, for example, might be read by only a handful of one's followers, or perhaps by thousands of strangers with whom this could be one's only interaction ever. For most users, it is simply unknown exactly to whom they are talking when they hit "send."

And, as many cases of sudden online infamy show, we can be drastically wrong about who our actual audience ends up being, like when a larger public gives uptake to utterances meant only for semi-private consumption. That this occurs demonstrates how our online speech acts are not in our control. As we speak online, our communicative goals can be seemingly outstripped by the medium, where the broader community's norms may play a greater role in determining what exactly we meant, and what we did with our words, than our own intentions.[8]

One reason this can occur is that platforms take some effort to hide from us the algorithmic architecture that renders this speech situation entirely unnatural. Facebook may ask you "what's on your mind," and Twitter might goad you to tell it "what's happening," but this is merely in support of their underlying goal to incentivize you to produce more (free) content for them. The fact is that our seemingly ephemeral expressions are cataloged in their servers where the data is mined to sell advertisements. And our willingness to share is fed by the rush of endorphins caused by carefully crafted notification systems and user-interface designs.[9]

In a classic article on the topic, John Suler notes that similar features lead to what he calls the "online disinhibition effect."[10] People, he noted, acted *differently* online than offline. He was careful to note that there is both benign and

7 See Levmore, "The Internet's Anonymity Problem"; Levmore and Nussbaum, *The Offensive Internet*.

8 Online shaming offers an instructive case. See Ronson, *So You've Been Publicly Shamed*; Norlock, "Online Shaming"; and Adkins, "When Shaming Is Shameful."

9 For one articulation of this idea, see Lanchester, "You Are the Product."

10 Suler, "The Online Disinhibition Effect."

toxic disinhibition, and more significantly that this was not meant to suggest that one's online self was somehow more *real*. Similarly, I do not mean to imply that these features of online speech make it somehow more artificial, more constrained, or less genuine than offline speech. Adaptation to online, mediated, text-and-image-based environments has been swift and full of ingenuity far beyond what platform designers could predict. A whole language of emojis and GIFs sits at our fingertips. My point here has simply been to remind us of these differences as they point to a noteworthy architecture that scaffolds our communicative acts. Fundamental questions like who or what should count as a "speaker," or how retweets, "likes," tagging, and emojis should fit into an account of utterances all need to be reexamined, as does my current question: How has the internet changed *harmful* speech?[11]

The philosophical literature on subordinating speech has seen steady growth for a few decades. And while the internet has been around almost as long, much of the philosophical work on hate speech, propaganda, and subordinating speech in general has focused on offline life.[12] In-person hate speech, like what you might see in public spaces, propaganda as it is disseminated in print or on the radio, and, more recently, microaggressions as they occur in settings like a workplace or college classroom, are the main examples.[13] This has remained the case even as more and more of our lives have migrated online.[14]

But online speech raises many new issues for social philosophers of language. The overall context of online communication—the total speech situation, as Austin would call it—is radically different from that of offline communication. To begin to explore these differences, let us briefly consider the internet's impact on propaganda—including notable subcategories like "fake news," or mis- and disinformation. One initial thought might be that all the internet has done is make it easier to spread propaganda to more people more quickly. And

11 For retweets, see Marsili, "Retweeting."

12 For a quick and nondecisive example, consider that the index for the 2012 anthology *Speech and Harm* has no entries for the terms "internet," "website," "online," or other specifically online communication mediums (see Maitra and McGowan, *Speech and Harm*). There are, of course, a few noteworthy exceptions, some of which I note below.

13 This is a large and growing literature. For important contributions, see Maitra, "Subordinating Speech"; Langton, "The Authority of Hate Speech" and "Blocking as Counter Speech"; McGowan, "On 'Whites Only' Signs and Racist Hate Speech" and *Just Words*; Stanley, *How Propaganda Works*; Tirrell, "Genocidal Language Games"; Rini, "How to Take Offense"; Saul, "Beyond Just Silencing"; and Liebow, "Microaggressions."

14 The anthology *Free Speech in the Digital Age*, edited by Susan Brison and Katharine Gelber, is an important recent entry in this area.

that would be problem enough.[15] However, the reach and speed of the internet is but one concern. Beyond these issues, further complications arise.

Regina Rini argues that social media posts can be considered a "bent" form of testimony whose features exacerbate preexisting problems. That is, our unstable norms around sharing information online—e.g., a retweet ≠ endorsement—enable old tensions to flourish in new ways. For Rini, fake news is not limited to online communications, but there is, as she says, "a strong contingent relationship between fake news and social media," making the one ripe for the other.[16] As she says:

> Perhaps people are less inclined to subject ridiculous stories to scrutiny *because* we have unstable testimonial norms on social media. A friend posts a ridiculous story, without comment, and *maybe* they don't really mean it. But then other friends "like" the story, or comment with earnest revulsion, or share it themselves. Each of these individual communicative acts involves some ambiguity in the speaker's testimonial intentions. But, when all appear summed together, this ambiguity seems to wash away.[17]

Rini's analysis shows how fake news can spread organically, where little to no malicious intent is needed, *because* of the distinct features of online communication, specifically social media. Other theorists, such as Zeynep Tufekci and Michael Lynch, worry that the personalization algorithms used on Facebook, YouTube, and other platforms make a hard problem—what to believe in our saturated information environment—even harder.[18] And, as Tufekci adds, "social media's business model financed by ads paid out based on number of pageviews makes it not just possible but even financially lucrative to spread misinformation, propaganda, or distorted partisan content that can go viral in algorithmically entrenched echo chambers."[19] The worry, therefore, is not simply that social media permits the rapid spread of propaganda, but that it has also incentivized new forms of propaganda to emerge, reach their targets, and further entrench themselves in communities.[20]

15 For an analysis of the "instantaneousness" of online hate speech, see Brown, "What Is So Special about Online (as Compared to Offline) Hate Speech?"

16 Rini, "Fake News and Partisan Epistemology," 45.

17 Rini, "Fake News and Partisan Epistemology," 49.

18 See Tufekci, "It's the (Democracy-Poisoning) Golden Age of Free Speech"; Lynch, *Know-It-All Society*.

19 Tufekci, *Twitter and Teargas*, 241. See, also, Nguyen, "Echo Chambers and Epistemic Bubbles."

20 For a particularly dramatic example of the potential developments at the intersection of technology and harmful speech, consider "deepfakes," that is, videos made using

At the extreme, the crossover between online hate and real-life violence is hard to deny. After the New Zealand mosque shootings, *New York Times* writer Charlie Warzel wrote:

> It's becoming increasingly difficult to ignore how online hatred and message board screeds are bleeding into the physical world—and how social platforms can act as an accelerant for terroristic behavior. The internet, it seems, has imprinted itself on modern hate crimes, giving its most unstable residents a theater for unspeakable acts—and an amplification system for an ideology of white supremacy that only recently was relegated to the shadows.[21]

This pattern has repeated itself in other locales, most explicitly in Buffalo, New York, where a shooter once again posted their manifesto online and attempted to livestream their acts on an online platform.

It is undeniable, therefore, that harmful speech enabled by emerging technology poses new sorts of problems of urgent concern. Violence arguably caused by online propaganda and misinformation has been reported in many countries including the US, Myanmar, Germany, India, and Canada. The role of Facebook, YouTube, and other platforms in exacerbating regional conflict is a contested debate.[22]

It is noteworthy, however, that the bulk of this debate addresses online subordinating speech as it functions in its *propagandistic* mode—as outreach or as a source of hateful beliefs that later cause harm—rather than on cases where speech is directly targeting particular individuals.[23] This is apparent in the focus on online speech's ability to manipulate beliefs and otherwise poison the information environment.[24] This sometimes leads to discussions of the "potential" harms of online hate, disinformation, or deepfakes that focus on abstract values like "democracy" or "civility" as its main victims. However, it ignores those who have *already* been victimized by online hate. In what follows, I examine online abuse as a topic worthy of serious philosophical investigation.

---

machine-learning algorithms to create the illusion that someone has said or done something they never did. For analyses, see Rini, "Deepfakes and the Epistemic Backstop"; and Rini and Cohen, "Deepfakes, Deep Harms."

21 Warzel, "Mass Shootings Have Become a Sickening Meme."

22 See Barnes, "Online Extremism, AI, and (Human) Content Moderation."

23 Note that a single speech act can play both roles at once. For an overview of the distinction, see Langton, "Beyond Belief."

24 This is perhaps the result of the fact that social epistemologists have been most active in this area.

I aim to bring out its structural elements while also drawing attention to how it is experienced by those targeted.

2. ONLINE ABUSE

The previous section provided a broad overview of a few features that make online speech distinct from IRL speech, as well as some reasons to worry about novel types of *harmful* speech online. At the general level, I believe we need to understand the peculiar features of these speech acts—including the material, structural, and design affordances that enable them—in order to assess any threats they may pose and consider how we might mitigate their harms. To demonstrate this, the remainder of the paper will focus on the narrower topic of *online abuse*. Offline models of subordinating speech do not easily accommodate the online features of this type of harmful speech, so it calls for reconsideration. In this section, I will lay out some notable aspects of online abuse; in the next I will show how these pose a challenge for standard philosophical accounts on offer.

To begin, we need a better idea of what online abuse includes.[25] Media studies professor Emma Jane articulates the breadth of the problem well in her (aptly titled) paper, "'Your a Ugly, Whorish, Slut': Understanding E-bile." Jane coins the term "e-bile" to capture what she describes as the "extravagant invective, the sexualized threats of violence, and the recreational nastiness that have come to constitute a dominant tenor of Internet discourse."[26] Jane stresses how this e-bile is found in nearly all corners of the internet and displays impressive flexibility in terms of functional use, but also that its effect varies depending on factors like who is targeted and what particular speech acts are being performed. That is, noting first how common this type of vitriolic speech is and how and when it combines with other factors can help to pinpoint when it rises to the level of online abuse.

On its commonness, and flexibility, Jane writes that

25 For some first-person accounts that touch upon the varied features of online abuse in detail, see Koul, *One Day We'll All Be Dead and None of This Will Matter*; La, "Here's How Trolls Treat the Women of CNET"; Quinn, *Crash Override*; Valenti, *Sex Object*; West, *Shrill*. For journalistic pieces on the topic, see Bernstein, "In 2015, the Dark Forces of the Internet Became a Counterculture" and "The Unsatisfying Truth about Hateful Online Rhetoric and Violence"; and Jeong, *The Internet of Garbage*.

26 Jane, "'Your a Ugly, Whorish, Slut,'" 532. Note that the topic under discussion goes by a few names: "e-bile," "cyberbullying," "online harassment," and more. I go with "online abuse," partly to follow internet safety activist Zoë Quinn, who suggests that "the term 'online abuse' is far more accurate because it perpetuates the dynamics of real-life abusive situations" (*Crash Override*, 50).

> hyperbolic vitriol—often involving rape and death threats—has become a *lingua franca* in many sectors of cyberspace. It is a commonsensical, even expected, way to, among other things: register disagreement and disapproval; test and mark the boundaries of online communities; compete and create; ward off boredom; prod for reaction; seek attention; and/or simply gain enjoyment.[27]

And yet, despite being put to many uses in so many contexts, Jane notes that "the rhetorical constructs of individual e-bile texts are strikingly similar in terms of their reliance on profanity, *ad hominem* invective, and hyperbolic imagery of graphic—often sexualized—violence."[28] She concludes that e-bile is found in nearly all corners of the internet and is used to perform a wide variety of speech acts, but at the same time has a uniformity across these usages, with expressions of sexual violence being a prominent trope.

Interestingly, Jane says that in many cases "e-bile appears to be a pleasurable—albeit competitive—game, in which players joust to produce the most creative venom, break the largest number of taboos, and elicit the largest emotional response in targets." It is a sort of commonplace online derogatoriness, and for this reason, she suggests that "what looks like hate speech might better be classed as 'boredom speech' or 'gaming speech.'"[29] However, as Jane is quick to note, while this may reflect the intentions behind many of these utterances, this does not capture the range of effects the *targets* of e-bile may experience, which can, in some cases, be very serious. She says that some of those "who have been targeted by e-bile generally report … emotional responses ranging from feelings of irritation, anxiety, sadness, loneliness, vulnerability, and unsafeness; to feelings of distress, pain, shock, fear, terror, devastation, and violation."[30] This is particularly the case, moreover, when what the target receives is not a mere one-off message, but an abundance of vitriolic and violent utterances. Their email inbox, Twitter mentions, DMs, etc., become flooded with horrendous comments and threats from a large number of strangers.

And here is where we can begin to narrow from the more general rhetorical patterns common to e-bile down toward the phenomena of online abuse. What in some contexts may be a type of expected, consensual—though misogynist—mutual banter, in other contexts can constitute a type of verbal attack. That these utterances share similar rhetorical styles—and that they are undeniably common in online communities—should not distract us from the fact that

27 Jane, "'Your a Ugly, Whorish, Slut,'" 542.

28 Jane, "'Your a Ugly, Whorish, Slut,'" 533.

29 Jane, "'Your a Ugly, Whorish, Slut,'" 534.

30 Jane, "'Your a Ugly, Whorish, Slut,'" 536.

their power to harm varies relative to the types of actions they are used to perform. Below, I will explain more fully how the utterances of online abuse function to harm their targets. Here, I simply aim to delineate the topic, noting that online abuse is partly characterized by its sheer scale and volume. This widespread circulation is what we refer to when we mean something has "gone viral." While I maintain that a particularly determined individual can inflict online abuse through "cyberstalking" or "cyberbullying," my focus will be on cases where the harassing language comes from multiple speakers.

Moreover, there is also the plain fact that if one is a member of an oppressed group offline, then that identity affects how likely they are to suffer abuse online and, of course, what form that abuse will take. Research from the Women's Media Center Speech Project confirms that women are more likely to be victims of online abuse, and the content of that abuse is overtly misogynistic.[31] Men and women of color often receive racist comments in response to mundane posts, especially if they are public figures.

For targets, these messages often form a pattern, and that pattern maps onto and is a part of broader structures of oppression. It is these two features that raise these individual pieces of e-bile from one-off oddities to become the harmful, indeed abusive, speech acts they are. However, before explaining how these utterances harm in the way they do—and how that poses a challenge to philosophical accounts of harmful speech—I first want to address the issue of motivations behind these utterances in more detail as this helps clarify my approach.

That is, as Jane and many others note, the functions and motives behind abusive rhetoric are more diffuse than might be expected. Even when directing messages toward out-group members as part of what we may call an overall abusive campaign, individual posters of vile content may do so for wildly varying reasons. This leads some commentators to suggest that they are not *really* engaging in a type of hate speech, so it is best to just "ignore the trolls." However, it is worth highlighting that many emotions besides hate motivate hate speech. As Jeremy Waldron puts it, "hatred is relevant not as the motivation of certain actions, but as a possible *effect* of certain forms of speech," that is, what this speech aims at or is likely to incite.[32]

So, while it is true that the motives and superficial purposes of online abuse might vary—one-upping, building solidarity, etc.—a more insidious function plausibly sits just below the surface: the intimidation of outsiders in order to exclude, and the reification of existing hierarchies of domination. And

31 For a brief overview of relevant survey data, see https://womensmediacenter.com/speech-project/research-statistics.

32 Waldron, *The Harm in Hate Speech*, 35. See also MacKinnon, "Pornography as Defamation and Discrimination," 808; and Smith, "Fighting Hate Is a Losing Battle."

this function, I argue below, is achieved partly through the *group activity* that online abuse becomes. It is by recognizing the red herring that is the individual speaker's—or "shitposter's"—underlying psychology, and in particular the irrelevance of their (stated) motives, that we are led to put the focus back on the act the speech *performs*, along with its expected *effects*—that is, its illocutionary and perlocutionary dimensions, to use the speech act theory terms of Austin.[33] In the next section, I turn to the philosophical literature on subordinating speech in part to demonstrate why it is not up to the task of assimilating online abusive speech into its (offline) apparatus before describing how online abusive speech attains its subordinating force.

3. THE AUTHORITY PROBLEM FOR ONLINE ABUSE

If subordinating someone through speech is a type of power that only some speakers have, then a natural question to ask is who holds this power and how do they acquire it. This is the authority problem for subordinating speech, and the question of what authority conditions enable different types of subordinating speech acts is a topic that has received sustained analysis.[34] Many compelling answers to this authority problem have been developed, including the claim that, in fact, speakers do not require any special authority to subordinate with their words or, if they do, all that is needed is a type of informal authority within a given domain. Other models show how speakers can come to gain the authority they lacked prior to speaking through processes like licensing and accommodation.[35]

This last approach has proven quite powerful, and it will be my focus as I leave the others largely aside.[36] The basic idea of accommodation is that while a speaker can lack the necessary authority to subordinate before they make their utterance, their speech act can nonetheless contain a *presupposition* of authority. If their audience fails to block the speaker's speech act by remaining silent, then this presupposition of authority is successfully added to the speech

33 Austin, *How to Do Things with Words*.

34 For helpful articulation of the problem as well as some of the main moves in the debate, see Maitra, "Subordinating Speech"; Witek, "How to Establish Authority with Words"; Bianchi "Asymmetrical Conversations."

35 See McGowan, "On Covert Exercitives"; Langton, "Speech Acts and Unspeakable Acts" and "The Authority of Hate Speech"; Barnes, "Speaking with (Subordinating) Authority."

36 I do so partly because accommodation is, in my estimation, the most popular account on offer, but also because I believe considering its faults leads us toward a better account. Quickly, I will note that an account that relies on an informal conception of authority—e.g., one that picks up on parameters of privilege like race, gender, and class—will have a harder time in online contexts, in part because of the prevalence of anonymous speakers and others whose only physical presence might be a cartoon avatar on the target's screen.

situation, understood as the "score" (following a Lewisian framework) and/or the "common ground" (following a Stalnakerian framework). The thought, explained by Rae Langton, is that speech acts, "including directives generally, and hate speech specifically, can acquire authority by an everyday piece of social magic: authority gets presupposed, and hearers let it go through, following a rule of accommodation."[37]

But online speech poses problems for accounts of licensing and accommodation. In particular, the role of *silence* in online spaces is not straightforwardly analogous to offline spaces. For this reason, Alexander Brown argues that "it can be harder to infer assent, licensing, or complicity from silence in the face of hate speech when that hate speech occurs online as opposed to offline."[38] The upshot of his analysis is that the standard story of how speech (or speakers) may be licensed to achieve subordinating authority is importantly incomplete for online speech. If bystander silence is required for licensing, but *online* bystander silence is notably different from offline silence, licensed authority may be harder (or impossible) to come by for hate speakers.

Moreover, according to the standard picture of accommodating authority, blocking—where an audience member rejects or challenges the speaker's utterance, including its presupposition—should be sufficient to cancel the authority from being accommodated. As Langton describes it: "A hearer who blocks what is presupposed, also blocks the *speech act* to which the presupposition contributes. . . . That is why blocking a presupposition can make the speech act fail."[39] It is worth emphasizing that Langton is here referring primarily to the *illocutionary* success of a speech act, not its perlocutionary effects (though it can affect this too), and this is because blocking prevents—or rather undoes, by her account—the acquisition of authority.[40] "A successful blocker," she says, "changes a past utterance from the unactualized way it would have been to the way it actually is. If a speaker's presupposed authority is blocked by a hearer . . . that blocking changes the past."[41]

37 Langton, "Blocking as Counter Speech," 152. For a more full account of the specific harm that bystander silence can contribute, see Ayala and Vasilyeva, "Responsibility for Silence."

38 Brown, "The Meaning of Silence in Cyberspace," 221.

39 Langton, "Blocking as Counter Speech," 145.

40 To see both sides of this, Langton says that "besides interfering with persuasion—with 'perlocutionary' success, in Austin's terms—blocking can interfere with the speech act itself, its 'illocutionary' success" ("Blocking as Counter Speech," 149). And later: "Blocking prevents illocutionary accommodation, tracked by score, *and* perlocutionary accommodation, tracked by common ground, achieving the latter because it achieves the former" (155).

41 Langton, "Blocking as Counter Speech, 156." As she further explains this: "Blocking can disable, rather than refute, evil speech. It can make speech *misfire*, to use Austin's label for a speech act gone wrong. It offers a way of 'undoing' things with words (to twist his

But in cases of online abuse, this does not seem to be what happens, or so I argue. Consider how, in cases of online abuse, a target might receive hundreds of messages, including *some* that are supportive and do the job of challenging the speech of harassers, right alongside comments that encourage suicide or worse. Here, there is no single, linear conversation to map the score or common ground onto. I believe this goes some way to explaining why counterspeech standardly fails to render these speech acts nonsubordinating. While it can help, it does not do the job of "blocking" or "canceling" a move in a language game as Langton hopes.

Why is this the case? I argue it is because the conversational dynamics of online speech are very unlike IRL conversations, where—paradigmatically—there are two parties who engage in a back and forth. Even when we add more participants, the image is still of a single, continuous thread where each new contribution builds upon and is constrained by what preceded it. Blocking makes sense in this context, as it is itself a contribution that future moves must acknowledge. But if you have ever looked at the replies under someone's viral tweet, you will know that this is not what is going on. Some comments get traction while others are ignored. Multiple, overlapping conversations all occur at once, playing out in a manner whose progression is hard to track. And when you add reply or quote functionality, the ability to call back to a specific moment in the exchange is enhanced. This all leads to a sort of branching of multiple conversations—if we even want to call them that—whose IRL parallel is hard to find and that do not share a single, easily traceable common ground.

Another answer to why blocking moves typically fail online emerges from considering the speech acts being performed here in more detail. As Jane notes about e-bile, "the point is rarely about winning an argument *via* the deployment of coherent reasoning, so much as a means by which discursive volume can be increased—e-bile is utilized, in other words, to out-shout everyone else."[42] Seen in this way, it becomes clearer why more speech—blocking speech—often will not work. Recognizing that its point is not to add new content to the conversational score—content that might be contested—but instead to inundate its targets with a barrage of hurtful words and imagery, shows the limits of this standard approach when the assailants number in the dozens, hundreds, or

title)—and this 'undoing' has, I shall suggest, a *retroactive* character, which Austin himself described. It offers a ticket to a modest time machine, available to anyone willing and able to use it" (145–46). For a different account that explores the potential to "undo" the past, see Caponetto, "Undoing Things with Words."

42 Jane, "'Your a Ugly, Whorish, Slut,'" 534.

even thousands.[43] Seeing this speech for what it is thus explains the question about blocking online—that is, why counterspeech cannot effectively do the blocking work it is supposed to do.

To be clear, this is not to say that counterspeech is pointless or serves no purpose.[44] It is simply to show its limits and how those limits expose conceptual problems within the accommodation framework. That is, this also demonstrates how the accommodation model relies on an overly rational, psychological picture of how content gets added to the common ground. Challenging messages that (in theory) *ought* to undo the initial speech act(s) often fail to do so (in practice), and the harm and subordination remain. We see this in how targets of abuse can still experience legitimate harms despite the presence of "blocking" utterances from others, as well as how harassers often do not acknowledge that any counterspeech even occurred and instead carry on *as if* it had not.

So, considering the apparent inability of the accommodation account to explain the force of abusive speech performed online, I believe we need to look elsewhere. Specifically, what is needed is an alternative account that can explain how seemingly powerless and often anonymous speakers can attain subordinating authority, even in the face of counterspeech. Rather than the somewhat passive model accommodation offers, I believe a much more active process is in play. Online abuses, I will argue, are best understood as cases of *collective subordinating speech acts*, as they are backed up by a collective authority attained by a chorus of speakers. In the following section, I explain how the sort of anonymity of the crowd made possible in online spaces, along with coalescence around shared language, enables a mass of speakers to attain a type of authority that impacts the force of their speech acts.

4. Online Abuse and the Construction of Collective Authority

When considering the type of online abuse I am directing us toward, it can seem obvious—trivial even—that much of the power that lies behind these utterances emerges from sheer numbers. This is part of the story, to be sure. The impact of a large number of speakers directing their hostility at a single target is not something that can be ignored. And online abuse harms in the way

43 The important role of graphic sexual and violent imagery in online abuse is, unfortunately, one aspect I mainly leave aside for this paper.

44 As Lynne Tirrell says (about IRL speech): "Challenges tend to push the game backward—they cannot undo the move but they can revoke a license. . . . Over time, enough challenges or challenges of the right kind might kill the viability of the move, depending on how local or global the challenge becomes" ("Toxic Speech," 143).

it does in part because of how awful it can be to find oneself in an unwanted spotlight, particularly when this means one is bombarded by racist, sexist, and/or transphobic commentary. However, there are additional features beyond mere numbers that come into play and give online abusive speech the particular force it has. That is, there is more to the authority conditions that enable online abuse than simply scale. Below, I describe two features that each contribute to the authority that underlies abusive speech online and, in doing so, explain the subordinating force it has.

4.1. Anonymity and the Force of (Veiled) Threats

As we have already seen, threats of physical and sexual violence are not rare online. Indeed, one of the common tropes of e-bile highlighted above is the ubiquity of violent misogyny:

> E-bile targeting women commonly includes charges of unintelligence, hysteria, and ugliness; these are then combined with threats and/or fantasies of violent sex acts which are often framed as "correctives." Constructions along the lines of "what you need is a good [insert graphic sexual act] to put you right" appear with such astounding regularity, they constitute an e-bile meme. Female targets are dismissed as both unacceptably unattractive man haters and hypersexual sluts who are inviting sexual attention or sexual attacks.[45]

And while direct threats do occur, more common is violent aggression expressed in the form of "hostile wishful thinking, such as 'I hope you get raped with a chainsaw.'"[46] While this indirect phrasing allows abusers to avoid legal trouble and skirt terms of service, it does not make these statements any less threatening to their targets. It is often, I claim, an escalation, as it seems to imply a coordinated group effort with a division of labor.

That is, veiled threats of this sort are only properly understood when we consider them in their full context, where they tend to imply a larger network of harassers. First, if the threat comes from an anonymous or unknown account—a nonfollower, for instance—that might suggest that it was directed

45 Jane, "'Your a Ugly, Whorish, Slut,'" 533.

46 Jane, "'Your a Ugly, Whorish, Slut,'" 533. Sarah Jeong calls this "colorably threatening harassment," which is: "Harassment that is not overtly threatening, but is either ambiguously threatening such that an objective observer might have a hard time deciding, or is clearly intended to make the target fearful while maintaining plausible deniability" (*The Internet of Garbage*, 33).

there by others, as the coordination of abusive campaigns is more common than many realize.[47] As Sarah Jeong reports,

> [The] examination of sustained harassment campaigns shows that they are often coordinated out of another online space. In some subcultures these are known as "forum raids," and are often banned in even the most permissive spaces because of their toxic nature. In the case of the harassment of Zoë Quinn, Quinn documented extensive coordination from IRC chat rooms, replete with participation from her ex-boyfriend.[48]

Even if there is no explicit coordination, there is often an implicit type that works just as well. One common pattern in online harassment is for an account with a large number of followers to quote tweet—a type of retweet where the retweeter can add further commentary—another user, mock them, and subtly suggest that their own followers pile on. The dynamics of social media, which reward engagement, can often lead to an escalation in harassment as users encourage each other in their shared goal of belittling the person singled out. As legal scholar Danielle Citron puts it, "online harassment can quickly become a team sport, with posters trying to outdo each other. Posters compete to be the most offensive, the most abusive."[49]

Second, and a bit more subtly, the way these utterances are given *uptake* reveals something important about how speakers accrue authority. As Lynne Tirrell argues, "our speech acts also undertake a meta-level *expressive commitment* about the very saying of what is said. Expressive commitments are commitments to the viability and value of particular ways of talking."[50] These expressive commitments can shift the boundaries of what counts as acceptable discourse in a community. And, in the case of online abuse, given that harassing speech in this medium often receives "likes" from other users, these commitments to the value of this discourse take *tangible* form. This helps shift the boundaries of permissibility.[51] Alexander Brown gestures toward this idea

47 Again, I adopt a low threshold for what counts as anonymity as I am mostly concerned with how these speakers appear to their audience. For this reason, I consider the perceived anonymity of crowds to be sufficient for anonymity in this sense.

48 Jeong, *The Internet of Garbage*, 74. While this is only one instance, further evidence suggests this practice is not as uncommon as some presume. For further examples, see Tufekci, *Twitter and Teargas*; Gray-Donald, "Canada's Right-Wing Rage Machine vs. Nora Loreto"; and Phillips, "The Oxygen of Amplification."

49 Citron, *Hate Crimes in Cyberspace*, 5.

50 Tirrell, "Toxic Speech," 144.

51 For another account on the shifting bounds of permissible speech, see Saul, "Racial Figleaves."

when he says that "the process of licensing hate speakers online could require more in the way of positive engagement with the hateful content … [such as] clicking the heart icon … or adding a supporting comment *via* the 'Reply' function."[52] I agree, and I aim to make this explicit. As I am putting it, in online contexts we can often see the shifts in the normative terrain—resulting from speech acts backed up by subordinating authority—by noting the numeric value in the "likes" and retweets harassment receives.

So, what might at first glance seem like a one-off message from a single individual can, in fact, reveal a message from a group of like-minded people. It is in this sense that it is a mistake to view the speech acts typical of online abuse through an individualistic lens. As Citron says, "when cyber mobs attack victims, individuals each contribute little to the attacks. The totality of their actions inflicts devastating harm, but the abuse cannot be pinned on a particular person."[53] This poses a problem for criminal law—Citron's focus—but, in general, taking this perspective is not too difficult; it simply amounts to listening to those who have experienced this harm. As Jeong says, "targets of harassment, particularly members of marginalized groups, may view a single comment differently than an outsider might, because they recognize it as part of a larger pattern."[54]

For those targeted by such speech, then, what is noteworthy is that online abuse can be read as a glimpse into the in-group speech of others, where marching orders are being given, are well-received, and might then be carried out by any one of the many anonymous figures on the other end of the internet. This takes a very real toll on its targets. As Lindy West says of her own experience with online harassment, questions like "Am I safe? Is that guy staring at me? Is he a troll?" easily flood your mind in public spaces.[55]

So, while anonymity poses challenges for the description of online abuse—namely, by foreclosing some standard explanations for the authoritative force of subordinating speech—it, in fact, provides a powerful tool for those who wish to inflict harm on their targets. It is the combination of anonymity and apparent group solidarity—"likes" instead of condemnation—that is a dangerous mix for targets of abuse, and, I claim, an important source of the authority these speech acts rely on to subordinate their targets.

This is evident in Quinn's description of her own experience with online abuse: "I read many of the threats in my ex's voice.… But this was somehow

52 Brown, "The Meaning of Silence in Cyberspace," 125.

53 Citron, *Hate Crimes in Cyberspace*, 24.

54 Jeong, *The Internet of Garbage*, 32.

55 West, *Shrill*.

more insidious—he wasn't just continuing his abuse; *he was crowdsourcing it*."[56] This vivid account is supported by media researcher Eden Litt's suggestion that "without being able to know the actual audience, social media users create and attend to an *imagined audience* for their everyday interactions."[57] That is to say, when we cannot directly perceive our audience, we create it in our minds. This calls back to one of the defining features of online communication I described earlier, and here we see how it impacts the force of online abuse. With this in mind, Kathryn Norlock notes that advising someone to "ignore the trolls" is beyond pointless. . . . The advice to ignore the social community as it lives in one's head is more than ineffective—*it's missing the force*."[58]

I believe this is exactly correct, that the force of online abuse—which is determined in part by the authority that sustains it—is dependent on the unique features of online communication. By seeing how anonymous avatars can become a monolith in one's mind, we can recognize a conception of subordinating speaker authority that, in fact, requires something like anonymity. It is in leveraging the target's own cognitive resources—namely, their capacity for imaginal relationships, which are necessary given text-based communication—that large-scale online abuse campaigns become more than the sum of their parts. Beyond affecting the force of individual messages, anonymity creates the semblance of cohesion where there might not, in fact, be any, thereby uniting different speakers who might not have anything in common aside from their hostile speech directed at the same individual.

Moreover, it is *through* this speech that they become united (at least in the mind of the target). It is for this reason that I refer to these as *collective subordinating speech acts*, whose subordinating authority—its capacity to harm in the distinctive way it does—is constituted by the active participation of an *ad hoc* community of speakers. Through repetition and endorsement, signaling support and solidarity, individual speech acts acquire authoritative standing in relation to a target, enabling them to harm. Each new utterance adds to the strength of the overall practice. Like accommodation, then, audience uptake secures authority for speech that, absent that uptake, would have a different pragmatic force. But as I have emphasized, in these cases, the practices that do the heavy lifting here are *active*, not passive.[59] In all these cases, speech plays an active role in solidifying the collective authority that strengthens their words,

56 Quinn, *Crash Override*, 51.

57 Litt, "Knock, Knock," 333.

58 Norlock, "Online Shaming," 194.

59 For a different but related adaptation of the concept of accommodation, see Adams, "Authority, Illocutionary Accommodation, and Social Accommodation."

turning it into the genuinely subordinating speech that it is. This is done in part through the construction of in-groups and out-groups. Herbert and Kukla point out that the recognition of insider status is something that comes into being through social practices that, in fact, constitute that status. They say that

> being recognized as an insider by insiders is not just the recognition of a separate fact; rather, this recognition plays a constitutive role in having that insider status. Part of being an insider is being recognized as one. Crucially, the relevant sort of recognition is not mere passive, conscious acknowledgment, but the kind of recognition that is built into practice.[60]

In online abuse, this takes the form of harassers cheering on other harassers through "likes," "retweets," and one-upping one another, along with other practices like coordinating on targets and sharing information.

So far, I have argued that, in cases of online abuse, anonymity—or at least, the anonymity one finds in the crowd—can contribute to the active construction of a group identity that may be wielded to inflict great harm. But anonymity is only part of the explanation I want to offer; shared, insider language is the other. I turn to this next.

4.2. Shared Language and Solidarity

To start, it is useful to note that the affordances of social media make it clear how a user's speech act is always tied to their (ever-shifting) socially constituted position—even when it is anonymous. Whether via a profile picture, short bio, hashtag, or emoji, social media brings new means of signaling identity. I want to emphasize, however, how this just amplifies what has always been the case offline. Mary Louise Pratt articulates this thought well when she writes:

> Once you set aside the notion of speech acts as normally anchored in a unified, essential subject, it becomes apparent that people always speak from and in a socially constituted position, a position that is, moreover, constantly shifting, and defined in a speech situation by the intersection of many different forces. On this view, speaking "for oneself," "from the heart" names only one position among the many from which a person might speak in the course of her everyday life.[61]

On social media, these implicit features of offline life are made fully explicit, often purposely so. Including a rose emoji or #MAGA, for example, can instantly situate a speaker as part of a wider community and communicate their broader

60 Herbert and Kukla, "Ingrouping, Outgrouping, and Peripheral Speech," 584.

61 Pratt, "Ideology and Speech-Act Theory," 63.

allegiances. These aspects of online speech allow individuals to actively construct and manage the version of themselves they present. This allows for a lot of variety, freedom, and play, including the inconsistencies that Pratt describes—e.g., concealing or emphasizing distinct parts of oneself for different platforms.

What is relevant for my purposes is how, on social media, this type of signaling often occurs by parroting the speech acts of another. "Speaking for oneself," in this context, often means speaking with the voice of another. While this is not an uncommon feature of speech, it is heightened and made explicit online, most obviously so through the use of hashtags, which allow one to visibly connect their own utterance to those of (usually) many others.[62] This feature of social media has proven powerful, as large social movements can galvanize around a hashtag that, in essence, consists in joining with the voices of others.[63] This can bring out both good avenues for effective solidarity and bad ones, as practices like the co-opting and appropriation of the words and voices of the more marginalized are all too common. For example, the phrase and hashtag "Black Lives Matter" has been taken up, twisted, and put to use for all sorts of ends, including by opposing forces.

So, while I am talking about a more general phenomenon, here I want to focus on how this can contribute to the group authority at issue in online abuse. Namely, hashtags (and related rhetorical constructions) help unify the voices of many into an *ad hoc* collective. As I will describe it, hashtags are *explicit ventriloquisms* and are a vivid example of language's role in constituting a group identity. That is, as Quinn puts it, how "the same techniques that people have used to organize important grassroots movements like Black Lives Matter can be used by people trying to destroy someone."[64]

In the course of building his account of slurs, the linguist Geoffrey Nunberg describes ventriloquisms:

> In a particular context, a speaker pointedly disregards the lexical convention of the group whose norms prescribe the default way of referring to *A* and refers to *A* instead via the distinct convention of another group

62 For a pragmatic analysis of hashtags as well as other unique features of online speech, see Kukla, "'Don't @ Me!'"

63 For an analysis of the impact of social media and other digital communication technologies on progressive activism, as well as how repressive regimes have learned to clamp down on these groups, see Tufekci, *Twitter and Teargas*. And for the story of how social media played a key role in the growth of the Black Lives Matter movement, see Khan-Cullors and bandele, *When They Call You a Terrorist*.

64 Quinn, *Crash Override*, 52.

> that is known to have distinct and heterodox attitudes about *A*, so as to signal his affiliation with the group and its point of view.[65]

That is, when a speaker uses a ventriloquism, they are disregarding the standard term that convention dictates and are instead mimicking the voice of another. In doing so, they signal their allegiance to a specific community, at least in that moment. Nunberg uses the example of a university dean using *ain't* in place of *isn't* to implicate that the knowledge being communicated was more folksy than academic.[66]

While Nunberg's main goal is to argue that slurs are cases of ventriloquisms, I aim mainly to get at an interesting feature of language, and I believe this account helps get us there. As he summarizes his view: "In a nutshell: racists don't use slurs because they're derogative; slurs are derogative because they're the words that racists use."[67]

Crucially it is not just shared attitudes that are implicated, but shared group membership:

> As [Langston] Hughes tells it, the force of [the n-word] goes beyond anything the speaker believes or feels about blacks. . . . It also evokes the things such people have *done* to blacks—with the speaker pointedly affiliating himself with the perpetrators. The word can turn a bigot from a hapless, inconsequential "I" into an intimidating, menacing "we."[68]

Without committing to this account of slurs, I do want to suggest that this analysis clarifies the pragmatic force of online abuse. Namely, the conception of ventriloquisms on offer demonstrates the potential of constructing a collective identity through shared language and tropes, as well as the ability for such a collective identity to undergird harmful speech. As I see it, hashtags and other shared rhetorical constructions function as explicit ventriloquisms, and in doing so serve to strengthen shared group identity for harassers. Hashtags are the most visible in part because they are literally visible, and their pragmatic function is to tie one utterance to many others. At their most extreme, they generate utterances with a first-person plural speaker—resulting in speech acts

65 Nunberg, "The Social Life of Slurs," 267.

66 As Nunberg explains the case: "A dean at an Eastern university [said]: 'Any junior scholar who stresses teaching at the expense of research *ain't* gonna get tenure.' In the dean's mouth, the use of the demotic *ain't* rather than *isn't* implied that his conclusion wasn't based on expert knowledge or a research survey; it was as if to say, 'You don't need an advanced degree to see that; it's obvious to anyone with an ounce of sense'" ("The Social Life of Slurs," 265).

67 Nunberg, "The Social Life of Slurs," 244.

68 Nunberg, "The Social Life of Slurs," 286. Note that this, according to Nunberg, distinguishes his view from a similar one offered by Camp, "Slurring Perspectives."

spoken by a collective "we." They perform the function, in other words, of what Hughes (as told by Nunberg) said slurs were capable of, and as a result can bring a similar subordinating authority to bear on targets.

Renée Jorgensen Bolinger develops a similar pragmatic account of slurs that can add to this story. While not a perfect parallel, Bolinger's *contrastive choice* account of slurs can add to the idea of ventriloquism by explaining how marked expressions can carry important signals about their speakers *for their audience*. As Bolinger puts it, "When we use slurs, we communicate information about ourselves and our attitudes towards the targets."[69] This information is signaled, moreover, through a speaker's decision to choose a particular term over a non-marked alternative. As she explains:

> For signals based in *contrastive choice*, the relevant behavior is the free selection of a marked expression, and performance signals that the speaker endorses a cluster of attitudes associated with the term (or, more precisely, a high probability that the speaker shares some or all of the attitudes in this cluster).[70]

Using a hashtag, it is worth pointing out, involves choice. It is literally marked—in blue, generally—and in some situations, it communicates the choice of *affiliation* or *association* with other users. But I do want to suggest that this thought applies beyond hashtags as well, which, as I said above, were simply the most visible version of this phenomenon. Some phrases, I claim, play a similar role as hashtags—and so, function as ventriloquisms—without being so explicit. Most often this occurs when a hashtagged phrase gains so much prominence that it enters the lexicon as marked in this peculiar way. Some examples likely include BlackLivesMatter, MeToo, MAGA, GamerGate, and even longer phrases like "it's about ethics in journalism," which was a common trope in GamerGate.

Or consider the use of the term "SJW," particularly as it occurs online. This is, in most cases, used pejoratively, referring commonly to individuals who promote socially progressive views like feminism and anti-racism. Importantly, this term is used almost exclusively by those who *oppose* these goals. In using this term, then, whether prefixed by a hashtag or not, speakers pragmatically convey information about their own group membership to their audience.

Again, as Bolinger helpfully explains:

> The information content of signals based in contrastive choice is linked to how marked the term is: if α is a term that is used almost exclusively by speakers who embrace ϕ, and this fact is well-known, then a contrastive

69 Bolinger, "The Pragmatics of Slurs," 439.

70 Bolinger, "The Pragmatics of Slurs," 447.

> preference for a is a high-information signal, raising the probability of the speaker's endorsing ϕ nearly to 1. The more well-known the association between a and ϕ is, the higher the information content of the signal, and thus the more strongly the contrastive choice signals the speaker's endorsement of ϕ.[71]

Since "SJW" is used mainly by its detractors, and since this is well-known, using it carries a high-probability signal that the speaker endorses these views too. The act of signaling this information performs an important function for both insiders and outsiders. In short, what terms like this do when repeated so much as to be marked in this way is to express and solidify group membership. This is a dynamic process performed primarily through speech acts. And it is through this process, moreover, that the targets of online abuse come to recognize that they are being addressed not by a single speaker, but by a mob.

This interpretation makes sense, moreover, since it is often exactly what is occurring. And as Jeong reports, it is this interpretation that makes sense of the "really bizarre phenomenon" of "all the low-level mobbers, who have little-to-no real investment in going after the target, and would not manifest any obsessions with that particular target without the orchestrator to set them off." As she explains:

> Here they resemble the zombie nodes of spam botnets, right down to the tactics that have been observed to be deployed—rote lines and messages are sometimes made available through Pastebin, a text-sharing website, and low-level mobbers are encouraged to find people to message and then copy/paste that message.[72]

Here again we see how in online abuse the implicit is often made fully explicit. Speakers are literally copying and pasting their utterances from one another, and in doing so adding strength to the subordinating force of each speech act. Shared language, along with the technological features of online communication, make this possible.

More importantly, this shows vividly why an individualist approach to online abuse is inapt for describing the force of these speech acts. It is only when we see these speakers as part of a collective, and a collective, moreover,

71 Bolinger, "The Pragmatics of Slurs," 447. Moreover, on this view, this is not reducible to speaker intentions: "Signaling on this framework is factive: a speaker signals some content ϕ when her use of an expression satisfies the conditions, regardless of whether she intended to communicate ϕ, and independent of whether hearer uptake occurs" ("The Pragmatics of Slurs," 447).

72 Jeong, *The Internet of Garbage*, 68.

that is constructed in part through the active use of shared speech acts, that we capture the pragmatic impact of these speech acts. They are, as I put it, backed up by a distinctively collective subordinating authority and so are *collective subordinating speech acts.*

Seeing online abuse as a sort of group activity encourages us not only to reject a lingering individualistic lens but also, I claim, is necessary for devising solutions to the harm they present. Reporters Max Fisher and Amanda Taub put this succinctly when they write:

> It is becoming increasingly common for groups of people, whipped into a rage by influential people on social media, to single out targets for mass campaigns of online harassment and threats. . . . *The main problem seems to be that social media companies' guidelines tend to focus on content in isolation*. Because the accounts that instigate the hatred and rage don't necessarily participate in the mass harassment directly—often their followers are the ones who send the death threats or do the doxxing—this problem is a poor fit for that approach.[73]

As this shows, tackling this problem properly requires addressing the collective from which the speech draws its power, whether it is an organic *ad hoc* group, or a preexisting community with a clear (if informal) hierarchy. Seeing this bigger picture is helpful in explaining the damage it can do to a community and it paints the way toward effective solutions. Social media companies can track this behavior—like they do all of our behavior—and, rather than basing their moderation decisions on individual pieces of content examined in isolation, they could focus on these patterns: swarming, copy-pasting, mass movements in attention across platforms, and other group-based practices rather than content.

5. Conclusion

In this paper I have examined what I take to be some key features of online abuse. I have emphasized the role of anonymity in cultivating the appearance of coordination and a division of labor in online abuse—even if in fact there is none—and shown how shared language plays an important role here as well. That is, I argued that anonymity plays a key role in building a type of collective authority for online abusive speech acts and, moreover, that the construction and endorsement of this group identity through shared rhetorical constructs like hashtags further adds to the targets' sense that they are being addressed

73 Fisher and Taub, "Social Media Has a Mob Violence Problem."

by a collective rather than individuals. I argued that this combination of anonymity and apparent group solidarity—shared phrases and hashtags, likes, and retweets—is a dangerous mix for targets of abuse, and an important source of the authority these speech acts rely on to subordinate. It is in this way, I argued, that online abuse becomes more than the sum of its parts.

These and other features build collective authority for seemingly isolated speech acts and, as I will now suggest in closing, reveal aspects of IRL subordinating speech that are often overlooked. In other words, I believe that greater attention to features similar to those I have highlighted in the online case can help bring out underemphasized aspects of offline hate speech. Racist graffiti spray painted on college campuses, slurs yelled from passing cars, white-nationalist flyers displayed in public, all invoke a sort of anonymity and group activity in a similar way to create an overall environment of exclusion. Across mediums, the force of any individual subordinating speech act draws on many other instances of similar utterances made by similar speakers, and this should be made more explicit in our accounts of its pragmatic functions. This follows from the more general observation that accounting for the realities of subordinating speech—both online and IRL—demonstrates that speaker authority is dynamic and emergent, and often depends on the wider community in more ways than simple accommodation suggests. Passive bystanders play an important role, to be sure, but greater attention must be paid to those who *actively* back up the subordinating speech of others. As I argue, this sort of contribution leads us away from an individualistic understanding of oppression, as is necessary. Online abuse makes this vivid, but I claim this is a feature shared by IRL forms of subordinating speech, and one that must be kept in mind.[74]

Australian National University
michael.barnes@anu.edu.au

REFERENCES

Adams, N. P. "Authority, Illocutionary Accommodation, and Social Accommodation." *Australasian Journal of Philosophy* 98, no. 3 (July 2020): 560–73.

Adkins, Karen. "When Shaming Is Shameful: Double Standards in Online Shame Backlashes." *Hypatia* 34, no. 1 (Winter 2019): 76–97.

Austin. J. L. *How to Do Things with Words*. Cambridge, MA: Harvard University

74 I want to thank the editors and referees for this journal, as well as audiences at numerous conferences and workshops. And special thanks are owed to Matthew Shields, Heather Stewart, and Quill Kukla for written comments at earlier stages of this project.

Press, 1962.
Ayala, Saray, and Nadya Vasilyeva. "Responsibility for Silence." *Journal of Social Philosophy* 47, no. 3 (Fall 2016): 256–72.
Barnes, Michael Randall. "Online Extremism, AI, and (Human) Content Moderation." *Feminist Philosophy Quarterly* 8, nos. 3/4 (2022).
———. "Positive Propaganda and the Pragmatics of Protest." In *The Movement for Black Lives: Philosophical Perspectives*, edited by Michael Cholbi, Brandon Hogan, Alex Madva, and Benjamin Yost, 139–59. Oxford: Oxford University Press, 2021.
———. "Presupposition and Propaganda: A Socially Extended Analysis." In *Sbisà on Speech as Action*, edited by Laura Caponetto and Paolo Labinaz. London: Palgrave Macmillan, 2023.
———. "Speaking with (Subordinating) Authority." *Social Theory and Practice* 42, no. 2 (April 2016): 240–57.
Barney, Rachel. "[Aristotle], On Trolling." *Journal of the American Philosophical Association* 2, no. 2 (Summer 2016): 193–95.
Bartlett, Jamie. *The Dark Net: Inside the Digital Underworld*. New York: Melville House, 2014.
Bernstein, Joseph. "In 2015, the Dark Forces of the Internet Became a Counterculture." *Buzzfeed*, December 23, 2015. https://www.buzzfeed.com/josephbernstein/in-2015-the-dark-forces-of-the-internet-became-a-countercult?utm_term=.jcmLwKXlW#.hbPV3edYy.
———. "The Unsatisfying Truth about Hateful Online Rhetoric and Violence." *Buzzfeed News*, November 23, 2018. https://www.buzzfeednews.com/article/josephbernstein/does-hateful-speech-lead-to-violence.
Bianchi, Claudia. "Asymmetrical Conversations: Acts of Subordination and the Authority Problem." *Grazer Philosophische Studien* 96, no. 3 (September 2019): 401–18.
Bolinger, Renée Jorgensen. "The Pragmatics of Slurs." *Noûs* 51, no. 3 (September 2017): 439–62.
Brison, Susan J., and Katharine Gelber, eds. *Free Speech in the Digital Age*. Oxford: Oxford University Press, 2019.
Brown, Alexander. "The Meaning of Silence in Cyberspace: The Authority Problem and Online Hate Speech." In Brison and Gelber, *Free Speech in the Digital Age*, 207–23.
———. "What Is So Special about Online (as Compared to Offline) Hate Speech?" *Ethnicities* 18, no. 3 (June 2018): 297–326.
Camp, Elisabeth. "Slurring Perspectives." *Analytic Philosophy* 54, no. 3 (September 2013): 330–49.
Caponetto, Laura. "Undoing Things with Words." *Synthese* 197, no. 6 (June

2020): 2399–2414.
Cherry, Myisha. "Twitter Trolls and the Refusal to Be Silenced." In *The Real World Reader: A Rhetorical Reader for Writers*, edited by James S. Miller, 221–25. New York: Oxford University Press, 2015.
Citron, Danielle Keats. *Hate Crimes in Cyberspace*. Cambridge, MA: Harvard University Press, 2014.
Davis, Jenny L. *How Artifacts Afford: The Power and Politics of Everyday Things*. Cambridge, MA: MIT Press, 2020.
Fisher, Max, and Amanda Taub. "Social Media Has a Mob Violence Problem. Could Soccer Hooliganism Prevention Offer a Model for Solving It?" *New York Times*, June 6, 2019. https://static.nytimes.com/email-content/INT_14028.html.
Fogal, Daniel, Daniel W. Harris, and Matt Moss, eds. *New Work on Speech Acts*. Oxford: Oxford University Press, 2018.
Gray-Donald, David. "Canada's Right-Wing Rage Machine vs. Nora Loreto." *Briarpatch*, April 15, 2018. https://briarpatchmagazine.com/blog/view/canadas-right-wing-rage-machine-vs-nora-loreto.
Herbert, Cassie, and Rebecca Kukla. "Ingrouping, Outgrouping, and the Pragmatics of Peripheral Speech." *Journal of the American Philosophical Association* 2, no. 4 (Winter 2016): 576–96.
Jane, Emma A. "'Your a Ugly, Whorish, Slut': Understanding E-bile." *Feminist Media Studies* 14, no. 4 (2014): 531–46.
Jeong, Sarah. *The Internet of Garbage*, rev. ed. New York: The Verge, 2018.
Khan-Cullors, Patrisse, and Asha Bandele. *When They Call You a Terrorist: A Black Lives Matter Memoir*. New York: St. Martin's Press, 2018.
Koul, Scaachi. *One Day We'll All Be Dead and None of This Will Matter*. Toronto: Doubleday Canada, 2017.
Kukla, Quill R. "The Pragmatics of Technologically Mediated Speech Acts: Don't @ Me." In *The Oxford Handbook of Applied Philosophy of Language*, edited by Luvell Anderson and Ernie Lepore. Oxford University Press, forthcoming.
La, Lynn. "Here's How Trolls Treat the Women of CNET." CNET, July 11, 2017. https://www.cnet.com/news/cnet-women-hate-troll-comments/.
Lanchester, John. "You Are the Product." *London Review of Books* 39, no. 16 (August 2017). https://www.lrb.co.uk/the-paper/v39/n16/john-lanchester/you-are-the-product.
Langton, Rae. "The Authority of Hate Speech." In *Oxford Studies in Philosophy of Law*, vol. 3, edited by John Gardner, Leslie Green, and Brian Leiter, 123–52. New York: Oxford University Press, 2018.
———. "Beyond Belief: Pragmatics in Hate Speech and Pornography." In

Maitra and McGowan, *Speech and Harm*, 72–93.
———. "Blocking as Counter Speech." In Fogal, Harris, and Moss, *New Work on Speech Acts*, 144–64.
———. "Speech Acts and Unspeakable Acts." *Philosophy and Public Affairs* 22, no. 4 (Autumn 1993): 293–330.
Levmore, Saul. "The Internet's Anonymity Problem." In Levmore and Nussbaum, *The Offensive Internet*, 50–67.
Levmore, Saul, and Martha C. Nussbaum, eds. *The Offensive Internet: Speech, Privacy and Reputation*. Cambridge, MA: Harvard University Press, 2010.
Liebow, Nabina. "Microaggressions: A Relational Analysis of Harms." In *Autonomy and Equality*, edited by Natalie Stoljar and Kristin Voigt, 195–219. New York: Routledge, 2021.
Litt, Eden. "Knock, Knock. Who's There? The Imagined Audience." *Journal of Broadcasting and Electronic Media* 56, no. 3 (September 2012): 330–45.
Lynch, Michael Patrick. *The Internet of Us: Knowing More and Understanding Less in the Age of Big Data*. New York: Liveright Publishing Corporation, 2016.
———. *Know-It-All Society: Truth and Arrogance in Political Culture*. New York: W. W. Norton & Company, 2019.
MacKinnon, Catharine. "Pornography as Defamation and Discrimination." *Boston University Law Review* 71, no. 5 (1991): 793–818.
Maitra, Ishani. "Subordinating Speech." In Maitra and McGowan, *Speech and Harm*, 94–120.
Maitra, Ishani, and Mary Kate McGowan, eds. *Speech and Harm: Controversies Over Free Speech*. New York: Oxford University Press, 2012.
Marsili, Neri. "Retweeting: Its Linguistic and Epistemic Value." *Synthese* 198, no. 11 (November 2021): 10457–83.
McGowan, Mary Kate. *Just Words: On Speech and Hidden Harm*. Oxford: Oxford University Press, 2019.
———. "On Covert Exercitives: Speech and the Social World." In Fogal, Harris, and Moss, *New Work on Speech Acts*, 185–201.
———. "On 'Whites Only' Signs and Racist Hate Speech: Verbal Acts of Racial Discrimination." In Maitra and McGowan, *Speech and Harm*, 121–47.
Meijers, Anthonie. "Collective Speech Acts." In *Intentional Acts and Institutional Facts: Essays on John Searle's Social Ontology*, edited by S. L. Tsohatzidis, 93–110. Dordrecht: Springer, 2007.
Nguyen, C. Thi. "Echo Chambers and Epistemic Bubbles." *Episteme* 17, no. 2 (June 2020): 141–61.
Noble, Safya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press, 2018.

Norlock, Kathryn J. "Online Shaming." *Social Philosophy Today* 33 (2017): 187–97.

Nunberg, Geoff. "The Social Life of Slurs." In Fogal, Harris, and Moss, *New Work on Speech Acts*, 237–95.

Phillips, Whitney. "The Oxygen of Amplification: Better Practices for Reporting on Extremists, Antagonists, and Manipulators Online." *Data and Society*, May 22, 2018. https://datasociety.net/output/oxygen-of-amplification/.

Pratt, Mary Louise. "Ideology and Speech-Act Theory." *Poetics Today* 7, no. 1 (1986): 59–72.

Quinn, Zoë. *Crash Override: How Gamergate (Nearly) Destroyed My Life, and How We Can Win the Fight against Online Hate*. New York: Public Affairs, 2017.

Rini, Regina. "Deepfakes and the Epistemic Backstop." *Philosopher's Imprint* 20, no. 24 (2020): 1–16.

———. "Fake News and Partisan Epistemology." *Kennedy Institute of Ethics Journal* 27, no. 2 supplement (June 2017): 43–64.

———. "How to Take Offense: Responding to Microaggression." *Journal of the American Philosophical Association* 4, no. 3 (Fall 2018): 332–51.

Rini, Regina, and Leah Cohen. "Deepfakes, Deep Harms." *Journal of Ethics and Social Philosophy* 22, no. 2 (July 2022): 143–61.

Ronson, Jon. *So You've Been Publicly Shamed*. New York: Riverhead Books, 2015.

Saul, Jennifer M. "Beyond Just Silencing: A Call for Complexity in Discussions of Academic Free Speech." In *Academic Freedom*, edited by Jennifer Lackey, 119–34. New York: Oxford University Press, 2018.

———. "Racial Figleaves, the Shifting Boundaries of the Permissible, and the Rise of Donald Trump." *Philosophical Topics* 45, no. 2: (Fall 2017) 91–116.

Sherr, Ian, and Erin Carson. "GamerGate to Trump: How Video Game Culture Blew Everything Up." CNET, November 27, 2017. https://www.cnet.com/news/gamergate-donald-trump-american-nazis-how-video-game-culture-blew-everything-up/.

Smith, David Livingstone. "Fighting Hate Is a Losing Battle." *Boston Globe*, August 29, 2017. https://www.bostonglobe.com/opinion/2017/08/29/smith/j5F9Mf4ZqPohu40xC6stTP/story.html.

Stanley, Jason. *How Propaganda Works*. Princeton: Princeton University Press, 2015.

Suler, John. "The Online Disinhibition Effect." *CyberPsychology and Behavior* 7, no. 3 (July 2004): 321–26.

Tirrell, Lynne. "Genocidal Language Games." In Maitra and McGowan, *Speech and Harm*, 174–221.

———. "Toxic Speech: Toward an Epidemiology of Discursive Harm."

Philosophical Topics 45, no. 2 (Fall 2017): 139–61.

Tufekci, Zeynep. "It's the (Democracy-Poisoning) Golden Age of Free Speech." *Wired*, January 16, 2018. https://www.wired.com/story/free-speech-issue-tech-turmoil-new-censorship/.

———. *Twitter and Teargas: The Power and Fragility of Networked Protest*. New Haven: Yale University Press, 2017.

Valenti, Jessica. *Sex Object: A Memoir*. New York: Dey Street Books, 2016.

Waldron, Jeremy. *The Harm in Hate Speech*. Cambridge, MA: Harvard University Press, 2014.

Warzel, Charlie. "Mass Shootings Have Become a Sickening Meme." *New York Times*, April 28, 2019. https://www.nytimes.com/2019/04/28/opinion/poway-synagogue-shooting-meme.html.

West, Lindy. *Shrill: Notes from a Loud Woman*. New York: Hatchette Books, 2016.

Witek, Maciej. "How to Establish Authority with Words: Imperative Utterances and Presupposition Accommodation." In *Logic, Methodology and Philosophy of Science at Warsaw University*, edited by Anna Brożek, 145–57. Warsaw: Warsaw University, 2013.

Journal of Ethics and Social Philosophy
VOL. 25, NO. 2 · AUGUST 2023
https://doi.org/10.26556/jesp.v25i2.1998

ARE ALL DECEPTIONS MANIPULATIVE OR ALL MANIPULATIONS DECEPTIVE?

Shlomo Cohen

MANIPULATION and deception are intriguing concepts in the sense that they both raise important and complex ethical issues that are not primarily speculative, theoretical, or the subject of extraordinary scenarios, but are rather the stuff of everyday concern for common people. The ability to reflect fruitfully about both concerns is, however, hindered by the significant lack of agreement on what both "manipulation" and "deception" precisely mean. There is serious risk of arguing past each other in normative debates when there is implicit disagreement on the precise meaning and contour of the phenomenon that is being evaluated.[1]

An immediate effect of the difficulty in defining both concepts is the problem of delineating the border between them. One attestation of the magnitude of disagreement and confusion can be found in the fact that we encounter two diametrically opposed positions on the relationship between the concepts: on the one hand is the view that manipulation is a subset of deception (all manipulations are deceptions), while on the other hand is the view that deception is a subset of manipulation (all deceptions are manipulations). The latter view is a direct conclusion of the thought that manipulations cause faulty mental states or deliberations in the other, and the trivial premise that false beliefs are (a paradigm of) faulty epistemic states that hinder successful deliberation.[2] Recently, Vladimir Krstić and Chantelle Saville argued explicitly that deception is a subset of manipulation, characterizing deception elegantly as "covert manipulation."[3] The opposite view, that all manipulations are deceptions, has

1 A good way to gain an initial appreciation of the difficulties involved in determining the meanings of "deception" and "manipulation" is to consult the respective encyclopedia entries for both concepts; see Mahon, "The Definition of Lying and Deception"; and Noggle, "The Ethics of Manipulation."

2 See, e.g., Noggle, "Manipulative Actions"; Barnhill, "What Is Manipulation?"; and Hanna, "Libertarian Paternalism, Manipulation, and the Shaping of Preferences."

3 Krstić and Saville, "Deception (under Uncertainty) as a Kind of Manipulation."

also had adherents.[4] The intuition here is clearly related by Shlomo Sher: "We connect deception with manipulation so strongly that it is sometimes thought that deception is a necessary aspect of manipulation."[5] The view that all manipulations are instances of deception has been explicitly and vigorously defended recently by Radim Bělohrad.[6] Bělohrad's is the most sustained argument for this view; I will therefore naturally spend some time engaging his arguments.

The core objective of this paper is to argue against both positions. If successful, this will illuminate the true relation between "manipulation" and "deception"—namely, that there is but a partial overlap between them, that none encompasses the other. It is possible that various thinkers have indeed assumed this view on the relationship between the two concepts, but to my knowledge, it has never been properly shown or systematically argued for. Hence, the two extreme views are still popular. Beyond the core project of arguing for the partial-overlap view, the discussion below will suggest some steps for delineating the borders between the two phenomena, highlighting some aspects of the relations between them, and pointing to a basic normative upshot.

1. Are All Manipulations Deceptions?

A *prima facie* observation may well suggest that many kinds of manipulation, while admittedly "tricky," do *not* amount to deception. In a previous paper, I provided an overview of such kinds of manipulations.[7] The following is an instructive example. It speaks of a pharmaceutical company that, being cognizant of people's tendency to associate the color blue (more than, say, the color orange) with tranquility, manufactures blue tranquilizer pills. "Predictably, marketing blue tranquilizer pills causes the public to buy more of them than the rival company's orange pills—coming to view them, falsely, as more potent." I concluded: "Since nothing in marketing blue pills deviates in any way from standards of veracity, there is no deception. And yet judgment was surely manipulated."[8] Beyond mere reliance on intuition, I argued that false beliefs in the consumers are triggered by a psychological mechanism that associates the color blue with tranquility. They are not caused by expressing a proposition—not even an implicit proposition—hence this manner of creating false beliefs

4 See, e.g., Goodin, *Manipulatory Politics*; Beauchamp, "Manipulative Advertising"; and Bruderman, "The Nature of Aesthetic Manipulation in Consumer Culture."

5 Sher, "A Framework for Assessing Immorally Manipulative Marketing Tactics," 104.

6 Bělohrad, "The Nature and Moral Status of Manipulation."

7 Cohen, "Manipulation and Deception."

8 Cohen, "Manipulation and Deception," 485.

is non-propositional. "Being non-propositional, these cases of manipulative communication have *no truth value*. This undergirds the intuition that they cannot possibly qualify as deceptions."[9]

Radim Bělohrad has recently suggested a way in which to oppose all such analyses and reassert the view that "manipulation essentially involves deception."[10] He believes that this view can be upheld if only we think more carefully on the *intentions* of the manipulator. Here is a reconstruction of his all-manipulations-are-deceptive argument:

1. All manipulations involve lack of transparency regarding the manipulator's true intentions.
2. Lack of transparency regarding intentions creates false beliefs in the other—viz., about the agent's state of mind—and by virtue of this qualifies as deception.

From 1 and 2 we conclude that all manipulation is deception.

While Bělohrad's is the most developed defense of this view, it represents a common intuition. This intuition is clearly articulated by Nathaniel Klemp:

> Manipulation always involves some level of insincerity. In fact, manipulative actions are the antithesis of sincere ones. When speakers lie, conceal relevant information, or distract listeners by appealing to irrational tendencies, they act with a lack of genuineness and with hidden ulterior motives. Such actions are in direct opposition to the "honesty," "genuineness," and "straightforwardness" defining sincerity.[11]

Below I attempt to refute both claim 1 and claim 2 independently (in sections 1.1 and 1.2, respectively). This analysis will, in turn, help delineate the scope of manipulation that *is* deceptive.

1.1. Refuting Claim 1

The counterclaim to 1, above (all manipulations involve lack of transparency regarding the manipulator's true intentions), is that lack of transparency of intentions is *not necessary* for manipulation. Two elements make up and support this view: that there is an extensive set of examples of manipulations that seem not to involve lack of transparency, and that—in contrast to Bělohrad's argument—not all these examples can be explained away as cases of coercion. In 1.1.1 and 1.1.2 I review these in turn.

9 Cohen, "Manipulation and Deception," 487.

10 Bělohrad, "The Nature and Moral Status of Manipulation," 459.

11 Klemp, "When Rhetoric Turns Manipulative," 74.

1.1.1. Manipulation without Lack of Transparency

Examples of transparent manipulations seem very easy to come by. This in itself serves as a presumption against the thought that all manipulations involve lack of transparency regarding the manipulator's intentions. Common metaphors used to describe manipulation are those of "pushing one's buttons" or "pulling one's levers." Very often we are all too aware that such a manipulative interpersonal dynamic is taking place, but we nonetheless consider such kinds of transparent cases not only to be manipulations but *paradigms* thereof. Importantly, this last point renders the presumption against the thesis of universal lack of transparency *very strong*. What kinds of cases are we talking about? Salient examples usually involve invocation of either positive or negative feelings that serve to motivate the other. Prominent instances include:

> *Playing on emotions of guilt in inappropriate ways*: The mother tells her daughter: "After all the hardship I went through in raising you, how can you do this to me?" The smart daughter understands that "this" refers to a decision that concerns a trifling matter that her mother happens not to like, and which only concerns her own personal life and is none of her mother's business—she easily sees through the manipulation. Yet she reluctantly admits that these guilt-evoking manipulations, when they come from her mother, have a way of working on her.
>
> *Directing social pressure against someone in a way that corners him and makes him feel uncomfortable not to conform*: Your partner wants to go on a family camping trip but you do not. While you are discussing it, your partner calls out to your children "Hey kids! Who wants to go on a camping trip?" The children cheer. You correctly judge that it is better to go on the camping trip (despite the drawbacks) than to disappoint your children.[12] (Assume that, *a priori*, both partners agree that decisions regarding trip destinations are to be made by them, not the children.)
>
> *Influencing someone by stroking their vanity*: The best chance of getting John to agree is to flatter him in the right way. Dan, who feels lazy at the moment, exploits this; he tells John: "This math problem is a bit too difficult for me. Take a look at it—I am sure you can figure it out in no time." While John is aware of this weak spot of his, Dan's playing on John's sense of pride in his ability nonetheless proves (again) to be the winning move.

12 Barnhill, "What Is Manipulation?" 54.

> *Influencing someone with the help of seduction*: Whenever Delilah wants Samson to be more open with her about things he prefers to keep discreet, she makes sure she is wearing her sexy robe before asking. While her tactic is obvious to Samson, he admits that it mostly works.

All the above are arguably paradigmatic cases of manipulation (the first two by eliciting negative emotions, the last two by eliciting positive ones); yet they are (and surely at least can be) fully transparent. The phenomenon of transparent manipulation is perhaps nowhere as straightforward and common in our culture as in advertisements. Ads are often, if not always, manipulative; and while the game they play is fully transparent (often ads are explicitly *declared*), advertising works—indeed, sometimes phenomenally.

In an effort to defend the universal nontransparency thesis of manipulation, Bělohrad offers the following argument: "I agree that sometimes the victim of manipulation may see through the intentions of the manipulator. But the question we must ask is not whether manipulation can be disclosed and still be effective, but whether the manipulator can be truly explicit about her intentions."[13] To this there are two responses. (1) It is not clear why we should think that the latter (i.e., the manipulator actually verbalizing her manipulative intention) rather than the former (the victim seeing through the intention) is the crucial parameter. The reason I believe this claim is wrong is the following: in order to manipulatively induce the intended emotion—say, guilt—in the other, the manipulator obviously has to *act out* a guilt-inducing behavior; and the point is that this could have the intended conditioning effect even if the person being affected is aware of what is happening. But if, on the other hand, the potential manipulator verbalizes that she could so act out, without actually acting it out, then *this* obviously would not contain the crucial element for exerting psychological influence on the other, and could not therefore amount to manipulation.[14] (2) It could rightly be argued that verbalizing the intention would interfere with the successful acting out of the manipulative behavior (as it nonsurprisingly does in the example of lying, and normally would in cases of inducing guilt). This lack of explicitness, however, just is not invariably the case. A beautiful example, which has been increasingly investigated in recent years, is that of "open-label placebo." In open-label placebo, the prescription of placebo pills is not done deceptively but is honestly *explained* to the patient;

13 Bělohrad, "The Nature and Moral Status of Manipulation," 458.

14 Bělohrad provides an example to support his diagnosis, but his example is infelicitous, since it is an example of *lying*, which is the least helpful kind of example to give if one wants to prove that in manipulation *generally* one cannot verbalize one's intentions and still succeed.

nonetheless, accumulating evidence shows that the open-label placebo manipulation works![15]

1.1.2. Manipulation and Coercion

Given the implausibility of denying the observation that some manipulative influences seem transparent, perhaps the natural move to take is to argue that transparent "manipulations" (i.e., where the intentions of the influencer are overt and clear to the target) are all in fact cases of *coercion*. This is precisely the route taken by Bělohrad. The intuition here is, presumably, this: if one sees that one is being manipulated and still cannot resist succumbing to the influence, then such irresistible influence is best understood (not as manipulation but) as coercion. If one shows that all nondeceptive (transparent) "manipulations" are in fact instances of coercion, then this clears the way to defending the view that all manipulations are indeed deceptive.

In lieu of expounding a theory about how to distinguish manipulation from coercion (which would require a full-blown paper), I will here take a paradigmatic example of transparent manipulation, and explain why it is not coercion. Recall the example of transparent manipulative seduction (or temptation) above. We ask: Is it reasonable to reclassify all transparent manipulative seductions as, in effect, cases of coercion? I will now offer four clear and simple intuitions why such a move would be exceedingly unreasonable.

(a) Reclassifying all transparent manipulative seductions as, in effect, cases of coercion would improbably get all those who succumb to overt seduction off the moral and legal hook. When one acts under coercion, one is (at least typically, if not always) neither morally blameworthy nor legally guilty. If all transparent manipulative seductions are coercions, then it would be enough for anyone charged with, for example, committing adultery, to simply convince us that he was overtly seduced (without having solicited it, being negligent or reckless, etc.) and this would deflect all moral blame or legal responsibility. This would obviously be a laughable line of defense in, say, an alimony lawsuit in the wake of infidelity.

(b) If transparent manipulative seduction amounts to coercion, then engaging in sexual intercourse as a result of overt seduction would be considered *prima facie* as rape (since rape is defined as nonconsensual sexual intercourse, and "nonconsensual" and "coerced" amount here to the same thing). This is clearly absurd.

15 For helpful general discussions see, e.g., Kaptchuk, "Open-Label Placebo"; Kaptchuk and Miller, "Open-Label Placebo"; Schaefer, Sahin, and Berstecher, "Why Do Open-Label Placebos Work?"

(c) Consider an example where a woman who wants her husband to stay at home seduces him sexually, knowing that he finds it hard to resist.[16] Since the husband obviously perceives and understands what she is doing, we are supposed to view this as a case of coercion. Now imagine that the husband, who is no less manipulative than his wife, realizes that he can either stay home and get nothing or stay home while being pleasantly seduced; he therefore begins to feign needing to go out. While his smart wife realizes what is now happening, she finds that, in this new predicament, it is the all-things-considered best option to continue playing this game. What we have here is a reciprocally (transparent) manipulative relationship—surely not a rare phenomenon as such. According to the definition of transparent manipulation as coercion, however, we are supposed to view this tangle as a two-way coercion. But can an interaction one enters and remains in voluntarily be defined as coercion? This might be possible, some think, in cases of "coercive offers," but then even they never argued that such a type of interaction can possibly be *reciprocally* coercive.

(d) Transparent manipulative seductions can be quite reliably effective, even when the seduction is *very mild*. Consider in this respect the interaction between physicians and pharmaceutical sales representatives. These representatives are chosen often because they are very attractive; their task is to manipulate doctors into prescribing their companies' drugs. While the doctors know precisely what the true intentions of the representatives who "present medical information" to them are, this simple ruse is nonetheless effective (the companies would not continue investing in this practice were it not profitable). The manipulation here works through a very mild type of seduction: no sex is involved, merely the eliciting of a pleasant feeling through being showered with positive personal attention by a very attractive person. It defies common sense to argue that the doctors who are seduced by the sales reps in this very mild sense are thereby *coerced* by them. (Aristotle's words are fitting here: "It is absurd to make external circumstances responsible, and not oneself, as being easily caught by such attractions, and to make oneself responsible for noble acts but the pleasant objects responsible for base acts.")[17]

These simple examples are enough to show, I believe, how improbable it is to try to salvage the all-manipulations-are-deceptions view by rebranding all transparent manipulations as instances of coercion. Similar demonstrations as those I brought with respect to seduction can be easily constructed with respect to other examples of transparent manipulations.

16 Taken from Rudinow, "Manipulation."

17 Aristotle, *Nicomachean Ethics* III.1, 1110b14.

Finally, let us remember that the fact that some influence is transparent normally serves to *increase*, rather than decrease, the agency of the person subject to it. Hence, the argument that transparent manipulations are coercive would normally imply that they would be *a fortiori* coercive if they worked nontransparently (and there is surely no reason to think that seducing, inducing guilt feelings, and so on stop working when the manipulative intentions behind them remain in the dark). This further shows the deeply problematic nature of the idea that transparent manipulations are necessarily coercive.

1.2. Refuting Claim 2

If, as we saw, not all manipulations include concealed intentions—if, that is, some manipulations are transparent to their victims—then a necessary condition for the thesis that all manipulations are deceptive does not hold, and so the thesis fails. While my argument could stop here, I will nonetheless proceed to show how the second premise of the all-manipulations-are-deceptive argument also fails, as this will expose further valuable insights into the relations between manipulation and deception. That second premise, let us recall, says: "Lack of transparency regarding intentions creates false beliefs in the other—viz., about the agent's state of mind—and by virtue of this qualifies as deception." Since the analysis above arguably demonstrated that not all cases of manipulation involve lack of transparency regarding intentions, we now focus on and inspect only the subgroup of manipulations that do in fact lack transparency of intentions.

We should also note that not all cases of lack of transparency of intentions amount to manipulation. (This is trivial; e.g., that I do not disclose to the vendor my intentions in buying the product does not by itself amount to manipulating the vendor.) Hence, what is interesting to show is not merely that it is not the case that all lack of transparency of intentions qualifies as deception—i.e., the rejection of premise 2—but the rejection of the stronger, more specific thesis that not all lack of transparency of intentions *in the context of manipulation* qualifies as deception. (Since we struck down premise 1, this condition is not anymore given, and needs to be added.) I will accordingly amend premise 2 in a way that would make it more specific and precise—and concomitantly less vulnerable to criticism—and will make *this* the target of my attack.[18]

The amended (more defensible) version of the second premise of the all-manipulations-are-deceptive argument is this:

18 To clarify the formal aspect of this move: to show a counterexample to a subgroup of x is more demanding than to show a counterexample to x, since the former satisfies the latter, but not vice versa.

2*. Whenever there is lack of transparency of intentions, and it is sufficient to qualify as manipulation, then it is also sufficient for deception.

This stronger thesis is the one I will now oppose. The basic idea then is this: manipulation lacking transparency of intentions, even when it is expected to cause false beliefs (regarding the agent's state of mind) and indeed does cause them, is *not sufficient* for deception. Manipulators can be nontransparent about intentions without this making their (misleading) influence of others deceptive.

The structure of the argument below will be the following. First, I will explain that causing false beliefs in others by nontransparency, and especially by nontransparency of intentions, does not as such amount to deception. Second, I will claim that causing false beliefs in others by manipulating them does not as such amount to deception. Third and finally, I will argue that the combination of the previous two claims—namely, causing false beliefs in others by means of manipulations with nontransparent intentions—also does not as such amount to deception. This, if true, will refute 2* (and, *a fortiori*, 2).

The idea that intentional nontransparency (that causes false beliefs) does not as such amount to deception is very intuitive inasmuch as "reticence is not necessarily deceptive."[19] This is expressed in the deception literature in the distinction between deception and "keeping in the dark." James Edwin Mahon writes: "If *A* prevents *B* from acquiring a true belief, then *A* keeps *B* in ignorance. However, *A* does not deceive *B*."[20] Deception causes its target to be mistaken, while "keeping in the dark" can cause its target to merely remain ignorant—these two are qualitatively different. Keeping someone in the dark *can* of course amount to deception, but only if certain conditions hold. Thomas Carson elaborates: "withholding information can constitute deception if there is a clear expectation, promise, and/or professional obligation that such information will be provided."[21] In the absence of such conditions, nontransparency is merely a withholding of information, which does not as such invariably (or even usually) amount to deception.

Against this general baseline, we are here interested in the particular case of withholding information about one's *intentions*. To assess this, let us first quickly articulate the theoretical context. According to a very common view, a necessary condition for deception is that "truth is warranted" in the communicative context. This is often explained by the idea that not communicating the

19 Mahon, "Kant and Maria von Herbert," 417.

20 Mahon, "A Definition of Deceiving," 187.

21 Carson, *Lying and Deception*, 56.

truth involves a *breach of trust*.[22] Now the important question for us is about the *scope* of this warrant of truth, and therefore of breach of trust. Some thinkers assume that, strictly speaking, it applies only to assertions.[23] Others believe it applies to conversational implicatures just as much as to assertions.[24] Yet others are explicit that this warrant must be extended to nonlinguistic deceptions (e.g., gestures) too.[25] The question for us here is whether the norms regarding warrant of truth extend also to the communicator's *intentions*. The norms in question are clearly not metaphysical; they are conventional norms of human communication.[26] What is required of us, therefore, is to consult our intuitions about the limits of the application of the norms of communicative trust. Now it is quite clear that the norms regarding warrant of truth, and hence regarding trustworthiness, often (or at the very least sometimes) do not extend to the intentions of communicators. The reason for this, however, is never made explicit. Our reluctance to view such nontransparency as deception is not arbitrary. Rather, viewing such nontransparency as deception (and hence *pro tanto* morally wrong) would spell a (*pro tanto*) moral obligation to reveal one's inner world to others to an extent that would breach basic norms of privacy and thereby harm the dignity of persons. The fundamental dignitary interest in privacy is by no means suspended by the sheer fact of participating in communication. Thus, the moral imperative of respecting human dignity serves as a boundary to expanding the notion of deception to a wholesale, or even a default, inclusion of a requirement to reveal intentions. Communicators are therefore under no default obligation to make their intentions public domain. There is consequently no prevailing norm of warranting the truth of intentions in communication.

Next comes the question of whether causing false beliefs in others by manipulating them is sufficient for deception. This question is addressed precisely in the following example:

> Paul intends to manipulate Mary emotionally (for example, into liking Paul). Paul's actions cause Mary to develop certain false beliefs, although

22 This dominant view can be found in different variations in Chisholm and Feehan, "The Intent to Deceive"; Williams, *Truth and Truthfulness*; Strudler, "The Distinctive Wrong in Lying"; Faulkner, "Lying and Deceit"; among many others.

23 See, e.g., Augustine, "Against Lying"; and Chisholm and Feehan, "The Intent to Deceive."

24 See, e.g., Williams, *Truth and Truthfulness*; and Saul, *Lying, Misleading, and What Is Said*.

25 See, e.g., O'Neil, "Lying, Trust, and Gratitude"; and Cohen, "The Moral Gradation of Media of Deception."

26 While I assume that these norms apply to humans more or less universally, it is enough for the purposes of this exposition if they apply only to the community of speakers in "our" civilization.

> this was no part of Paul's intention. Lacking that intention, his action is not deception; yet it is (intentional) manipulation that causes false beliefs.

The conclusion: "Manipulations that cause false beliefs are clearly not *ipso facto* deceptions."[27]

The third, ultimate question is whether the combination of the above two conditions—i.e., causing false beliefs by manipulation that involves concealed intentions—is necessarily deceptive. It is not unreasonable to expect that the addition of relevant parameters could cross a certain threshold and thereby enter the scope of a given concept. However, the following example illustrates, I believe, that this does not hold in our case.

> *Stroke*: Nicole's neighbor, Isaac, suffered a stroke; and although he recuperated quite well, Nicole knows that the minor disability that remains evokes feelings of worthlessness in Isaac. Today Nicole needs a new shelf, and was just about to go out to buy one when she recalls that Isaac used to take much pride in his carpentry skills. She also knows that Isaac has a soft spot for her. So, Nicole forgoes visiting her favorite store and knocks on Isaac's door instead. She tells him: "I really need a new shelf, and I remember you are …" Before she completes her sentence, Isaac interjects, "Let's go down and take measurements!" Despite giving up on the shelves from her favorite store, Nicole is happy she could find a way to strengthen Isaac's sense of self-worth.

Nicole solicited Isaac's help in a manipulative manner: she caused him to act by (i) stroking his vanity regarding his artistry, (ii) exploiting his liking for her, and supposedly even (iii) exploiting his manly tendency to want to feel like "the rescuer of a woman in need." And while Nicole expected that Isaac would assume falsely (hence form the false belief) that Nicole's intention was to seek help, Nicole's intention was in fact *to* help. So Nicole intentionally formed a false belief (about her intention) in Isaac. But (as most agree) not every case of causing false beliefs in others amounts to deception, and, in particular, I believe it is far-fetched to claim that Nicole deceived Isaac: she needed a new shelf, and that is what she communicated to him. Her communication was truthful. As we have seen, it could be interpreted as deceptive only if there were "a clear expectation, promise, and/or professional obligation" that information about Nicole's (benevolent) intention be announced.[28] But as anyone who has had neighbors knows, such high expectations of transparency are not

27 Cohen, "Manipulation and Deception," 484.

28 Carson, *Lying and Deception*, 56.

normally part of that type of relationship. More importantly yet, expecting such transparency—vis-à-vis neighbors or whomever—would severely diminish our ability to help people in need while preserving their sense of self-respect; for this reason, as well as others, humanistic societies reject such a norm of transparency. If I am right that it is unreasonable—and, I would add, even dangerous—to call Nicole's communication deception, then Stroke is a case of *nontransparent manipulation that is nonetheless nondeceptive.*

Manipulation and deception have various similarities; it is therefore easy to transfer our intuitions from one to the other. Nicole (benevolently) manipulated Isaac, and given that Isaac acquired false beliefs in the process, it is easy uncritically to assimilate manipulation into deception, and judge erroneously that Nicole deceived Isaac. But manipulation and deception are different creatures, and while Stroke is a case of manipulation, it is not a case of deception.

1.3. No Alternative Arguments

My discussion responded to the argument that all manipulation involves nontransparency regarding intentions, that all such lack of transparency amounts to deception, and that therefore all manipulation is deceptive. I showed that both premises do not withstand scrutiny, and therefore that the conclusion is (doubly) not vindicated. It could be argued that this leaves the possibility of some alternative argument—i.e., one not based on lack of transparency of intentions as a mediating term in a transitive argument—that could vindicate the all-manipulations-are-deceptive claim. Building on what has been already said, I will now argue that such alternative paths are blocked.

Our discussion has shown that it is not true that all nontransparency of intentions in communication amount to deception (and simultaneously that it is not true that these two are extensionally equivalent). This conclusion naturally lends support to the complementary (opposite) possibility: that all deceptions involve nontransparency of intentions. This view is indeed quite intuitive, as it is entailed by the (common) view that all deception is intentional, and the insight that one cannot possibly declare the intention to deceive and still proceed with *deception*.[29] Now if all deception involves necessarily the nontransparency of intentions, then, if we want to hold the all-manipulations-are-deceptions view, we must conclude that all manipulations involve nontransparency of intentions. This, however, has already been shown above to be false. The upshot of this argument is that it is not true that all manipulations are deceptions.

29 Mahon, "The Definition of Lying and Deception."

One small lacuna remains in this argument. It involves the possibility that there exist deceptions that do not involve concealment of intentions—which is only sensible to the extent that those deceptions are nonintentional. But even if we grant the possibility of nonintentional deception, this cannot save the all-manipulations-are-deceptions argument. The reason is that *manipulations* are necessarily (at least to a certain level) intentional, and that which is intentional cannot be completely contained within that which is not.[30] With this realization, the argument against the all-manipulations-are-deceptive view is now complete.

2. ARE ALL DECEPTIONS MANIPULATIONS?

2.1. A Prima Facie *Reasonable View*

While the idea that all deceptions are instances of manipulation has rarely been the subject of elaborate or even explicit articulation, it seems to follow from a straightforward reading of one of the most influential accounts of manipulation in the literature—that of Robert Noggle. Noggle writes: "There are certain norms or ideals that govern beliefs, desires, and emotions. I am suggesting that manipulative action is the attempt to get someone's beliefs, desires, or emotions to violate these norms, to fall short of these ideals."[31] Getting someone to acquire false beliefs, which is what deception does, is a paradigm of making someone's beliefs fall short of the ideals relevant for them; hence, we may conclude, all deception is manipulation. Noggle's reasoning is quite compelling, and it is noteworthy that no one, to my knowledge, has ever attempted to refute it directly. Recently, Noggle wrote explicitly that what he calls "the trickery account" tends to treat manipulation "as a broader category of which deception is a special case."[32] Vladimir Krstić and Chantelle Saville, based on an analysis of some interesting cases, concluded similarly that "while manipulation is not

30 Nobody I know of ventured to claim the opposite, i.e., that manipulation can be strictly unintentional. Marcia Baron writes of reckless manipulation, but she sees recklessness as at most an aspect of *intentional* influence ("The Mens Rea and Moral Status of Manipulation"). Kate Manne eloquently describes a subconscious passive-aggressive manipulative attempt to cause guilt in others, but this too is not unintentional ("Non-Machiavellian Manipulation and the Opacity of Motive"). Rather, Manne's case shows that even if there is no self-aware intention "to *manipulate*" (i.e., that is described to oneself in such terms), there is nonetheless a clear *intention to influence*—and this, in conjunction with other manipulation-constituting attributes of the behavior, is all that is needed to diagnose intentional manipulation.

31 Noggle, "Manipulative Actions," 44.

32 Noggle, "Pressure, Trickery, and a Unified Account of Manipulation," 243.

a species of deception, deception is a species of manipulation." They also suggested a precise identification of the subgroup of manipulations that comprise deception: "purposeful covert manipulations constitute deception . . . whilst those that are not covert constitute manipulations *simpliciter*."[33]

Despite the *prima facie* reasonableness of the view that all deceptions are instances of manipulation, I believe it does not withstand scrutiny. I present below three counterexamples.

2.1.1. Nonintentional Deception

The simplest argument against subsuming all deception under "manipulation" is available to those who hold that deception can be unintentional. Jonathan Adler, for instance, argues: "Deception generally, of course, need not be intentional or voluntary."[34] More radically yet, Gary Fuller refers to the distinction between intentional and unintentional deception as "unimportant."[35] Chisholm and Feehan's classic paper on deception presents a similar view.[36] If, as indeed seems the case, manipulation *must* be intentional, then the conclusion immediately follows that not all deception is manipulation.[37]

In the remainder, I set aside the (minority) view that deception can be unintentional, and present two independent arguments for the claim that some intentional deceptions are not manipulations.[38]

2.1.2. Deception without Intention to Influence

Consider the following case.

> *Liar*: Larry is sometimes described as a pathological liar, since he seems to lie compulsively just about anything, irrespective of any benefit he might get from producing the corresponding false beliefs or their effects. People who know him describe him rather as an "aesthete of deception"—they say he simply relishes making up beautiful false stories in response to questions directed to him, without caring the least about the impact of his fanciful stories on others or about how others would react to those stories.

33 Krstić and Saville, "Deception (under Uncertainty) as a Kind of Manipulation," 835.

34 Adler, "Lying, Deceiving, or Falsely Implicating," 435. Adler opines that lying only must be intentional.

35 Fuller, "Other-Deception," 21.

36 Chisholm and Feehan, "The Intent to Deceive."

37 See note 30, above.

38 Lack of intentionality will indeed feature in the next argument, but it will refer to a particular aspect, as I explain presently; hence, deception will not be unintentional *simpliciter*.

Larry intentionally tells what he knows to be falsehoods to unsuspecting listeners, under circumstances where, we assume, truth is (taken to be) warranted.[39] This is *sufficient* to identify Larry's behavior as deception.[40] Larry follows Oscar Wilde in thinking that the most awesome kind of lying "is Lying for its own sake, and the highest development of this is … Lying in Art."[41] Now in point of fact, Wilde's "lying" is not real deceptive lying, as we do not expect works of art to be factually true; but Larry does tell his (believable) stories in circumstances where (he knows) truth is warranted, and this does make him a deceiver.

Typically, behavior such as Larry's exhibits a complementary characteristic—namely, an intention to change another's mind by implanting false beliefs in the other. However, this linkage is not *necessary*, and in particular, it is not the case in Liar, where self-centered Larry simply relishes making up his imaginative stories "without caring the least about the impact of his fanciful stories on others." His intention is wholly focused on exercising his wild artistic imagination; the audience is, as it were, but a trigger—and not a necessary one at that.[42] (While in most kinds of scenarios stating entails objectively "an invitation to believe," and therefore stating falsehoods qualifies as deception, the fabricator need not *subjectively* intend this invitation. Accordingly, he need not intend to influence.) Since manipulation *necessarily* involves an intention to make some impact on, i.e. to influence, the other, the combination of factors that Liar manifests makes it an example of deception without manipulation. This, if true, shows that it is not the case that deceptions (at least as understood here) are a subtype of manipulation.[43]

In Liar, the intentional telling of falsehoods is separate from the intention to create false beliefs. (Although the two intentions are typically related, they are

39 Hence, we assume that listeners are not aware that Larry is a repeat liar. Let us also assume that the stories Larry tells are believable, and that he realizes that much.

40 For one of the clearest expressions of such a view see Saul, *Lying, Misleading, and What Is Said*, 3.

41 Wilde, "The Decay of Lying," 34.

42 Interestingly, Larry's phenomenon exhibits precisely the opposite characteristics from Harry Frankfurt's "bullshitter": while the bullshitter cares only about the impact of his words on others, and not about their truth, Larry cares only about the (un)truth (i.e., fictional character) of his stories, and not about their impact (Frankfurt, *On Bullshit*).

43 I should add that "aesthetic" lying is not a bizarre, far-fetched phenomenon, as some might initially suspect. As against the view that "lying for the fun of it is a form of craziness" (Burge, "Content Preservation," 474), there exists the idea that "lying is lovely if we choose it, and is an important component of our freedom" (Arendt, *Rahel Varnhagen*, 11). The aesthetic motivation for lying was perhaps never as pithily phrased as by Samuel Butler: "Any fool can tell the truth, but it requires a man of some sense to know how to lie well" (*The Note-Books of Samuel Butler*, 300).

clearly distinct.) The—normally unnoticeable—gap between the two intentions opens a space for deceptions that are not manipulations.

As an addendum, I should mention that an even stronger argument can be made here, though space does not allow me to develop it. A skeptic might claim: when someone intentionally tells falsehoods to others in contexts where truth is typically warranted, there must be at least some indirect sense—be it as derivative and remote as possible—in which an intention to influence others (to cause false beliefs) *can* be attributed to him. In response, even if, for the sake of argument, we accepted this view, this would not change our conclusion. The reason is that, being socially constituted creatures, there is virtually *always* some sense in which our self-directed actions can be simultaneously interpreted as referring indirectly to others. This, therefore, cannot helpfully point to a reasonably circumscribed domain of potential "manipulations." Hence, even if it were insisted that Larry must have some indirect intention to influence others, such a trivial sense of "intention" would be insufficient to constitute manipulation.

2.1.3. Deception without Phenomenological Features of Manipulation

Next, I want to argue that lying as such may not be enough to constitute manipulation, even if there is a direct intention to influence (to cause false beliefs). This argument is independent from any specific understanding of deception. Consider the following example.

> *Grumpy*: Smith woke up in a very grumpy mood this morning, and has no patience to have even the most minimal conversation with anybody. As he is standing in the street corner, waiting for the light to turn green, a passerby asks him, "Excuse me, do you know if this street leads to the market?" Smith knows the answer, but anticipating that a truthful answer might lead to a follow-up question, he just spits out "No clue!" and the passerby continues on his way.

I contend that Grumpy is not a story of manipulation. While Smith lied to the passerby, Smith did not manipulate him.[44] What is the ground for this assertion? In the absence of an authoritative definition of manipulation, we can nonetheless get a reliable assessment of Grumpy in light of paradigmatic characteristics of the phenomenon of manipulation. If, as I suspect, "manipulation" is likely *not* evoked in people's minds upon hearing Grumpy, if it is annexed to lying merely due to some (explicit or implicit) theory that "all deceptions are

44 Smith's grumpy reaction can be meant by Smith and interpreted by the passerby as merely communicating "Get lost!" In such a case, it would not be a deceptive lie. But obviously it *can* be interpreted as a deceptive lie, and this default interpretation is the sense I here intend.

manipulations," then we should return to the phenomenology of manipulation to check whether the generalization imposed by theory does justice to the phenomena. Accordingly, I describe below salient features of the phenomenology of manipulation—features that have not received serious attention—and assess Grumpy in their light.

Definitions of manipulation have all, in one form or another, focused on a problematic attitude toward rationality.[45] While this is important, it is not the only salient feature in the phenomenology of manipulation. Manipulative action is routinely described as "pulling levers" or "pushing buttons," and the metaphor of puppet and puppeteer recurs frequently and seems to embody something distinct and important about the character and "feel" of manipulation.[46] What is conveyed by these is, arguably, a deep sentiment that the manipulator *plays with* his target. Specifically, "playing" refers to some sense of penetrating the mental or psychic machinery of the target, which allows steering the target.[47] Another important, related feature of manipulative action is that in its attempt to obtain control of the target's behavior, it involves at least some minimal *focused attention* on its target, and, I should add, this attention is geared toward *harnessing* the victim to play a role in the manipulator's scheme. Now Grumpy, instructively, does not exhibit these salient features of manipulation.

Let us look first into the parameter of metaphorically "playing with" the other's psyche. Joel Rudinow insightfully describes the manipulator's behavior as "predicated on some privileged insight into the personality of his intended manipulee."[48] In stark contrast, the pure and simple lie of answering no instead of yes does its deceptive job straightforwardly, without any need to "penetrate into the mental machinery" of its victim. Smith's behavior does not express

45 For the best/most influential definitions available, see Faden and Beauchamp, *A History and Theory of Informed Consent*; Noggle, "Manipulative Actions"; and Gorin, "Towards a Theory of Interpersonal Manipulation."

46 For an expression of the thought that the puppet-puppeteer metaphor is important for understanding the concept of manipulation, see Sunstein, "Fifty Shades of Manipulation," 216.

47 "Steering" as such is clearly not enough to capture *manipulative* influence specifically. Steering could be done physically, as with a cattle prod, but this is clearly not "manipulation" in the relevant sense. Even "communicative steering" is not precise, as this could also refer to rational persuasion, which often—though not invariably (Gorin, "Towards a Theory of Interpersonal Manipulation")—constitutes the *antithesis* of manipulation. "Non-rational-persuasion communicative steering" too is not accurate: this could refer, for example, to endless gestures that we all employ in communicative influence (e.g., smiling), and nobody intends to brand that entire dimension of human interaction as "manipulation." Hence, we need a very sensitive analysis to zero in on the relevant parameters constituting the idea of *manipulative* steering.

48 Rudinow, "Manipulation," 346.

any attempt to "operate the passerby from within," as it were; he rather merely utters a falsehood, whose misleading impact is an automatic function of language, requiring neither intention nor understanding of how to operate human beings "from within." Just as when telling the *truth*, the truth "speaks for itself," i.e., the impact of what is said is the direct function of the linguistic message, rendering interpersonal dynamics of influence—and hence manipulation—superfluous, so is the case with the crude simple lie: its inherently misleading nature does not require the interpersonal dynamics and phenomenology of manipulation. Hence, for each case of lying, we need to check whether the dynamics and phenomenology of manipulation are exhibited or not. The distinction alluded to here aligns well with my distinction between manipulation and deception: while manipulation interferes with the workings (the "form") of judgment—and this requires having a grip on the other's psyche—deception as such merely provides false input to judgment, i.e. in contrast to manipulation, it interferes with the *content* of judgment.[49] This latter, I stress, need not exhibit the kind of "managing" of the other so pathognomonic of manipulation (see more below). In a related vein, Todd Long has emphasized the difference between influencing others by providing false information (only) and influencing by gaining control of their psychological mechanisms.[50] While Long's focus is on the question of influence that preserves moral responsibility (the former does, the latter does not), his view that deception as such does not undermine moral responsibility demonstrates a rather similar intuition to the one expounded here: deception presents misleading information (i.e., content), and this as such is distinct from gaining control of the other's inner psychological mechanisms (i.e., judgment)—which is what the manipulator typically does when "pulling his victim's strings."[51]

Manipulation requires at least the minimal interpersonal sophistication needed to understand how to harness the other's psyche to perform as the manipulator wishes. In contrast, when a liar (e.g., grumpy Smith) says no, although the true answer is yes, such misleading requires devoting zero *attention* to the liar's victim or to how circumstances affect the victim's information processing; it requires zero understanding of the other's psyche, and therefore also zero planning of how to maneuver the other. There is nothing in this most minimal act of deflecting another that exhibits the phenomenology of "playing with" the other. The phenomenological difference between manipulation and Smith's lie can also

49 Cohen, "Manipulation and Deception," 486.

50 Long, "Information Manipulation and Moral Responsibility."

51 Long does use the term "information manipulation" for deception, but this is primarily because he writes in the context of the free-will literature, where "manipulation" is used generally for the act of influencing others' decisions and actions.

be presented from a complementary angle. The crude simple lie (as in uttering no instead of yes) exhibits a "mechanical" character of sorts: it can be viewed instructively as the verbal analogue of the physical act of forcing the target's head in the opposite direction, so as to prevent her from seeing reality. This analogy to physical steering suggests a point overlooked in the literature: crude lying can be much closer in its phenomenology to *coercion* than to manipulation!

A typical feature of manipulation involves the manipulator harnessing his victim to become a pawn in his scheme; but Smith is not interested in the passerby playing any role in any scheme of his. Smith just cannot be bothered with giving an iota of consideration to the passerby—and indeed he does not. The extreme lack of attention to the other exhibited by Smith is the very *opposite* of the mindset characteristic of *manipulating* the other. While, unlike Liar, Grumpy does contain the element of attempting to influence the other, the effect sought by Smith is merely to brush off the passerby; and *merely* brushing someone off is, at least sometimes, the wrong sort of influence to constitute manipulativeness. This argument, I should stress, refers to the *process* of manipulating, clearly not to its goal (which can be anything, including brushing someone off). As a process, manipulative influence *engages* the other ("playing with" is a form of engaging). Smith's lie does precisely the opposite: it holds off the other; it is a form of *dis*engaging.

The claim that deceptive lying is invariably manipulative ought to be supported by the phenomenology of manipulation. The novel phenomenological analysis presented here strongly suggests that Grumpy is a case of deception without manipulation. (In the lack of a reasonably comprehensive theory of manipulation, it is virtually impossible to offer a precise delineation of the necessary conditions for manipulation. Hence our phenomenological analysis, while strongly suggestive, cannot be shown to constitute a decisive proof.)

Notice that while my discussion of Liar attempted to demonstrate that there can be deception without intention to create false beliefs, my discussion of Grumpy attempted to show that there can be intention to create false beliefs, which does not qualify as manipulation. In both of these ways, then, there can be deception without manipulation.

To sum up this section, I presented three arguments against the view that all deceptions are manipulations: the first referred to the idea of unintentional deception, the second to the idea of deception as the intentional telling of falsehoods in situations where truth is warranted, yet without intending to influence (by creating false beliefs), and the third referred to deceptive lying that intends to cause false beliefs, but that lacks central phenomenological characteristics constitutive of manipulation. It is worth emphasizing that the three counterexamples are independent of each other, so that any *one* of them is enough to

undermine the thesis that all deceptions are manipulations.[52] While I do not contend that this constitutes a knockout argument (which would require much more extensive treatment than possible here), it does, I believe, offer a new position for further debate, and even shifts the burden of proof. In philosophy, these typically constitute a real step forward.

2.2. Implications for Understanding Manipulation and Deception

If the analysis above is right, it shows that not all deceptions are manipulations. Beyond this, however, it has other interesting implications for our reflection on both deception and manipulation. I will here briefly mention two thoughts.

That lying may not be manipulative is interesting, I believe, because lying has been taken to be a—if not *the*—paradigm of manipulation. For example, in Robert Noggle's influential account (mentioned above), manipulation "leads astray" by making others fall short of ideals for belief, emotion, and desire. However, what precisely are to be taken as ideals for emotions and desires may be difficult to determine objectively. Hence, the clearest case of making others fall short of ideals refers to beliefs; and within this category, the clearest case of making beliefs fall short of the ideals pertaining to them is to induce false beliefs. Lying, the most straightforward way of inducing false beliefs, thus becomes paradigmatic of manipulation. In addition, Noggle's idea on how to characterize formally the ideals with which manipulation interferes is based on the constraint of preserving "a conceptual parallel with lying."[53] Claudia Mills, as another example, sees a deep analogy between manipulative action, as providing bad reasons, and lying, as providing false information—so much so that Mills finds in lying the key to deciphering the moral nature of manipulation, and consequently declares: "If lying is wrong, so is manipulation."[54] Realizing that lying is less paradigmatic of manipulation than it has been taken to be can open the way to novel and perhaps subtler explorations of manipulation.

Reflecting on why lying in Grumpy is not manipulative can advance our understanding also of the theory and ethics of deception. An interesting debate in the ethics of deception concerns the question of whether the *form* of deception (notably, lying versus falsely implicating) has moral significance, and if so, how? The dominant position seems to be that lying is morally worse.[55] However, Clea Rees has argued that falsely implicating (in her terms: "merely

52 Again, while the first and second arguments rely on particular views of deception, the third is not similarly restricted.

53 Noggle, "Manipulative Actions," 47.

54 Mills, "Politics and Manipulation," 103.

55 See, e.g., Webber "Liar!"; and Shiffrin, *Speech Matters*.

deliberately misleading") is worse.[56] The reason, according to Rees, is that while lying breaches trust in assertions only, falsely implicating breaches wider linguistic trust, encompassing conversational implicatures as well as assertions. Our reflection on Grumpy suggests a different reason for why falsely implicating may be worse than lying (when and to the extent that it is): it adds the wrongness of manipulation to that of deception. The liar typically need not bother assessing how his message would be processed by her victim, since what she tells is straightforward (in this, again, false statements are basically similar to true ones). Things are very different with the nonlying deceiver who uses false conversational implicatures or nonlinguistic deception, however. *That* deceiver must give much more consideration to the psyche of her victim—to calculate how to be subtly suggestive in the right way and measure so as to influence her victim into making the wrong inference, and thus falling into the trap. While in lying, the falsehood itself does the deceptive job, in nonlying deception, the misleading is mediated via the victim's misinterpretation of the meaning of the message (which is not a falsehood in the strict sense but only pragmatically) in the given context. The typical nonlying deceiver must therefore *plan* (even if only furtively and subconsciously) how to maneuver her victim's interpretative mechanisms so they draw the misleading conclusion. This kind of tampering with the mental machinery of the other so as to steer it into operating in the way the agent wants them to operate is the typical work of the manipulator. (Think for instance of double bluffing as a clear illustration of such distinctively *manipulative* deceit.) Above I argued that lying is not an apt paradigm for manipulation; our last considerations suggest that *nonlying* deception may provide a more instructive model.

Deception, I conclude, is not necessarily manipulative; in addition, the paradigm for deception that *is* manipulative is probably different from what it has been taken to be.

3. CONCLUSION, AND ETHICAL UPSHOT

3.1. A Partial Overlap

We have seen that it is neither the case that deception is a subtype of manipulation nor that manipulation is a subtype of deception. (Our arguments simultaneously ruled out the possibility that they are coextensive.) This leaves two logical options: either deception and manipulation are completely discrete entities, or they partially overlap. It is patently obvious, however, and denied by no one, that many cases are simultaneously of deception and of manipulation; it is

56 Rees, "Better Lie!"

hence incorrect to think of deception and manipulation as completely discrete. These considerations generate the conclusion that the relation between "deception" and "manipulation" is one of *partial overlap*: while some manipulations are not deceptions and some deceptions are not manipulations, some cases qualify as both deception and manipulation. This conclusion runs against some powerful prevailing intuitions, and it has never been systematically argued for before.

3.2. Moral Conclusions

If manipulations are not essentially deceptive and deceptions are not essentially manipulative, then *moral judgments* regarding the one cannot automatically be transferred wholesale to the other, based on the intrinsic relations between the concepts.

This conclusion is perhaps especially significant with respect to (the rejection of) the view that all manipulations are deceptive. Since deception is usually taken to be *pro tanto* morally wrong, that view implies that manipulations too, being a subset of deceptions, are *pro tanto* wrong. Rejecting that view means that that shortcut to a general moral characterization of manipulation is not available. The debate as to whether manipulation is or is not *pro tanto* wrong therefore remains open.[57]

Similarly, rejecting the view that all deceptions are manipulations means that we cannot, strictly on the basis of the relation between the concepts, transfer wholesale our complex moral judgements regarding manipulations to deceptions. This too is instructive and may prove significant for moral judgment. For instance, in cases where deception, but not manipulation, would maintain the target's moral responsibility (as in Long's view mentioned above), and where that is a salient moral consideration, deception might be all-things-considered morally permissible, so long as it is not manipulative too.[58]

I conclude that while the moral analyses of deception and of manipulation should surely inform each other, they must, ultimately, be approached independently.[59]

Ben-Gurion University of the Negev
shlomoe@bgu.ac.il

57 Most thinkers assume that manipulation *is pro tanto* wrong, but there are plausible dissenting opinions; the latter include: Baron, "The Mens Rea and Moral Status of Manipulation"; Blumenthal-Barby, "A Framework for Assessing the Moral Status of 'Manipulation'"; and Cohen, "Manipulation and Deception."

58 Long, "Information Manipulation and Moral Responsibility."

59 I would like to thank Ron Aboodi for useful comments on a previous draft. This research was supported by the Israel Science Foundation (grant no. 1077/18).

REFERENCES

Adler, Jonathan. "Lying, Deceiving, or Falsely Implicating." *Journal of Philosophy* 94, no. 9 (September 1997): 435–52.

Arendt, Hannah. *Rahel Varnhagen: The Life of a Jewish Woman*. Translated by Richard Winston and Clara Winston. New York: Harvest Books, 1974.

Aristotle, *Nicomachean Ethics*, translated by W. D. Ross. New York: Oxford University Press, 2009.

Augustine. "Against Lying." In *Moral Treatises of Saint Augustine*, translated by C. L. Cornish. New York: Aeterna Press, 2014.

Barnhill, Anne. "What Is Manipulation?" In Coons and Weber, *Manipulation*, 51–72.

Baron, Marcia. "The Mens Rea and Moral Status of Manipulation." In Coons and Weber, *Manipulation*, 98–120.

Beauchamp, Tom L. "Manipulative Advertising." In *Ethical Theory and Business*, edited by Tom L. Beauchamp and Norman E. Bowie. Englewood Cliffs, NJ: Prentice Hall, 1988.

Bělohrad, Radim. "The Nature and Moral Status of Manipulation." *Acta Analytica* 34, no. 4 (December 2019): 447–62.

Blumenthal-Barby, Jennifer S. "A Framework for Assessing the Moral Status of 'Manipulation.'" In Coons and Weber, *Manipulation*, 121–34.

Bruderman, Eli. "The Nature of Aesthetic Manipulation in Consumer Culture." *South African Journal of Philosophy* 35, no. 2 (2016): 210–23.

Burge, Tyler. "Content Preservation." *Philosophical Review* 102, no. 4 (October 1993): 457–88.

Butler, Samuel. *The Note-Books of Samuel Butler*. New York: M. Kennerley, 1913.

Carson, Thomas. *Lying and Deception: Theory and Practice*. New York: Oxford University Press, 2010.

Chisholm, Roderick M., and Thomas D. Feehan. "The Intent to Deceive." *Journal of Philosophy* 74, no. 3 (March 1977): 143–59.

Cohen, Shlomo "Manipulation and Deception." *Australasian Journal of Philosophy* 96, no. 3 (2018): 483–97.

———. "The Moral Gradation of Media of Deception." *Theoria* 84, no. 1 (February 2018): 60–82.

Coons, Christian, and Michael Weber, eds. *Manipulation: Theory and Practice*. New York: Oxford University Press, 2014.

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

Faulkner, Paul. "Lying and Deceit." In *International Encyclopedia of Ethics*, edited by Hugh Lafollette, 3101–9. Hoboken, NJ: Wiley-Blackwell, 2013.

Frankfurt, Harry G. *On Bullshit*. Princeton, NJ: Princeton University Press, 2005.

Fuller, Gary. "Other-Deception." *Southwestern Journal of Philosophy* 7, no. 1 (Winter 1976): 21–31.

Goodin, Robert E. *Manipulatory Politics*. New Haven: Yale University Press, 1980.

Gorin, Moti. "Towards a Theory of Interpersonal Manipulation." In Coons and Weber, *Manipulation*, 73–97.

Hanna, Jason. "Libertarian Paternalism, Manipulation, and the Shaping of Prefersences." *Social Theory and Practice* 41, no. 4 (October 2015): 618–43.

Kaptchuk, Ted J. "Open-Label Placebo: Reflections on a Research Agenda." *Perspectives in Biology and Medicine* 61, no. 3 (Summer 2018): 311–34.

Kaptchuk, Ted J., and Franklin G. Miller. "Open Label Placebo: Can Honestly Prescribed Placebos Evoke Meaningful Therapeutic Benefits?" *British Medical Journal* 363 (October 2018): k3889.

Klemp, Nathaniel. "When Rhetoric Turns Manipulative: Disentangling Persuasion and Manipulation." In *Manipulating Democracy*, edited by Wayne Le Cheminant and John M. Parrish, 77–104. New York: Routledge, 2010.

Krstić, Vladimir, and Chantelle Saville. "Deception (under Uncertainty) as a Kind of Manipulation." *Australasian Journal of Philosophy* 97, no. 4 (2019): 830–35.

Long, Todd. "Information Manipulation and Moral Responsibility." In Coons and Weber, *Manipulation*, 151–75.

Mahon, James Edwin. "A Definition of Deceiving." *International Journal of Applied Philosophy* 21, no. 2 (Fall 2007): 181–94.

———. "The Definition of Lying and Deception." *Stanford Encyclopedia of Philosophy* (Winter 2016). https://plato.stanford.edu/entries/lying-definition/.

———. "Kant and Maria von Herbert: Reticence vs. Deception." *Philosophy* 81, no. 317 (July 2006): 417–44.

Manne, Kate. "Non-Machiavellian Manipulation and the Opacity of Motive." In Coons and Weber, *Manipulation*, 221–45.

Mills, Claudia. "Politics and Manipulation." *Social Theory and Practice* 21, no. 1 (Spring 1995): 97–112.

Noggle, Robert. "The Ethics of Manipulation." *Stanford Encyclopedia of Philosophy* (Summer 2020). https://plato.stanford.edu/entries/ethics-manipulation/.

———. "Manipulative Actions: A Conceptual and Moral Analysis." *American Philosophical Quarterly* 33, no. 1 (January 1996): 43–55.

———. "Pressure, Trickery, and a Unified Account of Manipulation." *American Philosophical Quarterly* 57, no. 3 (July 2020): 241–52.

O'Neil, Collin. "Lying, Trust, and Gratitude." *Philosophy and Public Affairs* 40, no. 4 (Fall 2012): 301–33.
Rees, Clea F. "Better Lie!" *Analysis* 74, no. 1 (January 2014): 59–64.
Rudinow, Joel. "Manipulation." *Ethics* 88, no. 4 (July 1978): 338–47.
Saul, Jennifer. *Lying, Misleading, and What Is Said*. Oxford: Oxford University Press, 2012.
Schaefer, Michael, Tamay Sahin, and Benjamin Berstecher. "Why Do Open-Label Placebos Work? A Randomized Controlled Trial of an Open-Label Placebo Induction with and without Extended Information about the Placebo Effect in Allergic Rhinitis." *PLOS ONE* 13, no. 3 (2018): e0192758.
Sher, Shlomo. "A Framework for Assessing Immorally Manipulative Marketing Tactics." *Journal of Business Ethics* 102, no. 1 (August 2011): 97–118.
Shiffrin, Seana Valentine. *Speech Matters: On Lying, Morality, and the Law*. Princeton, NJ: Princeton University Press, 2014.
Strudler, Alan. "The Distinctive Wrong in Lying." *Ethical Theory and Moral Practice* 13, no. 2 (April 2010): 171–79.
Sunstein, Cass. "Fifty Shades of Manipulation." *Journal of Marketing Behavior* 1, nos. 3–4 (2016): 213–44.
Webber, Jonathan. "Liar!" *Analysis* 73, no. 4 (October 2013): 651–59.
Wilde, Oscar. "The Decay of Lying." 1891. In *The Decay of Lying and Other Essays*, 1–38. London: Penguin, 2010.
Williams, Bernard. *Truth and Truthfulness*. Princeton, NJ: Princeton University Press, 2002.

Journal of Ethics and Social Philosophy
VOL. 25, NO. 2 · AUGUST 2023
https://doi.org/10.26556/jesp.v25i2.1843

FAMINE, AFFLUENCE, AND AQUINAS

Marshall Bierson and Tucker Sigourney

THOMAS AQUINAS famously held that (1) theft is always wrong, and also that (2) it is permissible for a starving man to take the bread he needs from another. He reconciled these two positions by claiming that (3) in cases of great need, it is not theft to take someone else's property when she does not need it herself. As he puts it, "Properly speaking, in a case of extreme need, to take and make use of another's things does not have the character of theft. This is because what someone takes for the purpose of sustaining his own life is made his own in virtue of his need."[1]

On its face, 3 looks like a theoretically costly concession that Aquinas is forced to make in order to reconcile 1 and 2. Surely, the objection goes, the more plausible explanation is that the need *justifies* the theft, rather than somehow transforms the act so that it is not theft at all.

Our principal aim in this paper is to show that claim 3 is not actually a costly concession—that, in fact, there are good independent reasons for adopting it. In sections 1–3, we argue that, given certain plausible intuitions about a range of cases we present, the only reasonable course is to adopt 3—to acknowledge that some of the cases in question are not merely cases of permissible theft, but rather not cases of theft at all. Then, in sections 4 and 5, we note that only some accounts of property are equipped to explain claim 3, and we consider why that might be. Finally, we observe that when we attend to the structure that accounts of this sort generally share, we tend to find that they have radical implications for the duties of the wealthy to give to those in need.

1. THREE CASES

Our argument begins with three cases:

> *Case 1*: Your child is terribly sick and likely to die. There is a medicine that could save her life, but the only dose is currently sitting on a shelf

1 Aquinas, *Summa Theologiae* II-II 66.7 ad 2, henceforth "*ST*." All translations are by Tucker Sigourney.

> in Grushenka's house. Grushenka has no use for the medicine herself—she just likes looking at it. You cannot convince her to sell or give it to you, but you could take it rather easily while she is away.

We will assume that you, the reader, agree with us that it is permissible for you to take the medicine.

> *Case 2*: You and your neighbor, Nikolay, each have a child who is terribly sick and likely to die. There is a medicine that could save their lives, and you were lucky enough to acquire it about a month ago. Unfortunately, there is only one dose, and no less will work.

We will assume that you agree with us that it is permissible for you to give the medicine to your child.

> *Case 3*: You and your neighbor, Nikolay, each have a child who is terribly sick and likely to die. There is a medicine that could save their lives, but there is only one dose, and this time Nikolay is the one who owns it. You could not possibly convince him to give it to you, but you could take it from him rather easily while he is away.

Here, we will assume you agree that it is impermissible for you to take the medicine from Nikolay.

Together these cases raise a puzzle. Case 1 seems to suggest that it is permissible to violate another's property rights when your child's life is on the line. Case 2 seems to suggest that when only one of two children can be saved, it is permissible to prefer your own child. Then why should it not be permissible both to prefer your own child and to violate another's property rights when only one of two children can be saved? What makes Case 3 relevantly different?

2. TWO UNSUCCESSFUL SOLUTIONS

Let us consider two preliminary attempts to resolve this conflict. As a first thought, we might try accounting for the difference as a simple matter of the collected weight of reasons. We might call this the *Arithmetic Hypothesis*: my agent-relative reasons to save my child may outweigh either my agent-relative reasons not to steal or my agent-relative reasons to preserve another child's life, but they do not outweigh both at once.

But the Arithmetic Hypothesis is implausible. If the only relevant difference between Cases 2 and 3 were the fact that there is an act of theft involved, then it would not matter whom you are taking from. But it does matter. Consider another case:

> *Case 4*: You and your neighbor, Nikolay, each have a child who is terribly sick and likely to die. You have a single dose of a medicine that could save their lives. Unfortunately, your child is severely allergic to the medicine, so you cannot administer it without a special antihistamine. The only person who has this antihistamine is Grushenka, who has no real use for it herself except that she enjoys collecting it. Unfortunately, you cannot convince her to give it to you—but you could easily take it from her while she is away.

The choice here is between giving Nikolay the medicine and taking the antihistamine so that you can give the medicine to your own child. Like Case 3, Case 4 pits one property claim and one child's life against your own child's life. If the Arithmetic Hypothesis were true, the sum would come out the same: taking the antihistamine and giving your child the medicine would be impermissible. But it is not.[2] It will be helpful to represent this result in a table (table 1).

Table 1

		Arithmetic Hypothesis			
Action	Intuitive Evaluation	Pro: Your Child Lives	Con: Another Child Dies	Con: Violates Someone's Property	Evaluation
1. Take the medicine from Grushenka	Permissible	✓		✓	Permissible
2. Give your child the medicine	Permissible	✓	✓		Permissible
3. Take the medicine from Nikolay	Wrong	✓	✓	✓	Wrong
4. Take the antihistamine and save your child	Permissible	✓	✓	✓	Wrong*

Note: Asterisks denote instances in which the evaluation generated by the hypothesis does not match the intuitive evaluation.

The Arithmetic Hypothesis does not handle Case 4 correctly, and this suggests that it does not capture the relevant difference between Case 3 and Cases 1 and 2.

What is it, then, that distinguishes Case 3 from Cases 1 and 2? We have tried a solution on which the same *kinds* of acts are implicated in all three cases,

2 We could also establish this conclusion by another route. Suppose we change Case 2 so that Nikolay is the father of *two* sick children, but *both* children could be cured with only half a dose of medicine. Even here, it seems permissible for you to use your dose of medicine to save your own child rather than give it to Nikolay. If the difference between this case and Case 3 were merely a matter of weighing up bad actions, then the act of theft in Case 3 would have to be a weightier consideration than the life of Nikolay's second child. But clearly it is not. If you disagree with our intuition about this case, you can simply ignore this footnote. It is not essential to our argument.

and all that distinguishes them is the *number* of these acts. Perhaps we should jettison that assumption, and instead distinguish the cases by the kinds of acts they involve. The obvious move is to distinguish what you do to Nikolay's child in the two cases. Let us call this the *Causal Hypothesis*: the difference between Cases 2 and 3 is that, if you took the medicine from Nikolay's child in Case 3, in an important sense you would be killing the child—and that is impermissible.

What exactly distinguishes killing from letting die is a matter of controversy, but for our purposes, we can leave the controversy aside. However we are to understand it, some such distinction is plausibly at work here. In Case 2, if you do not act at all, Nikolay's child will still die. In Case 3, that is not so.

But the Causal Hypothesis will not work either—at least, not if this difference between killing and letting die is independent of considerations about property rights. Consider two more cases:

> *Case 5*: You and your neighbor, Nikolay, each have a child who is terribly sick and likely to die. Your neighbor, Katerina, has the only dose of a medicine that could cure the sickness, and she intends to give it to Nikolay. You cannot convince her to give it to you instead, but you could take it rather easily while she is away.

> *Case 6*: You and your neighbor, Nikolay, each have a child who is terribly sick and likely to die. Katerina has the only dose of a medicine that could cure the sickness, and she intends to give it to Nikolay. But Katerina is a good friend of yours. You know that if she learned that your child was sick as well, she would give the medicine to you instead. You meet her in the street on your way home, and she asks you how you have been lately.

In both of these cases, you can do something such that your child will receive the medication and Nikolay's child will not. Both actions would result in a child's death. Yet it seems that your action would be wrong in Case 5, but not in Case 6.

Case 5 is like Case 3. The only difference is that the one who owns the medicine is no longer the sick child's father. But it was not Nikolay's status as the child's father that made the difference in Case 3. (Nikolay could just as well have been trying to save a sick orphan—taking the medicine would still have been wrong.) Case 6, on the other hand, is like Case 2 (where the medicine is yours).[3] Here, although you know speaking up will have dire consequences for Nikolay's child, you are not obligated to hold back or to lie.

3 What makes the difference is not that Katerina is your friend, nor that you are only talking to her. She could just as easily have been selling the medication, with the intention to give it to Nikolay only if nobody bought it first. You would not have been obliged not to buy it.

The Causal Hypothesis fails to distinguish correctly between Cases 5 and 6, which suggests that it, too, fails to capture the relevant difference between Case 3 and Cases 1 and 2. We can add this point to our chart (table 2).

Table 2

		Arithmetic Hypothesis				Causal Hypothesis	
Action	Int. Eval.	Your Child Lives	Another Child Dies	Violates Property	Eval.	Causes a Child's Death	Eval.
1. Take the medicine from Grushenka	Perm.	✓		✓	Perm.		Perm.
2. Give your child the medicine	Perm.	✓	✓		Perm.		Perm.
3. Take the medicine from Nikolay	Wrong	✓	✓	✓	Wrong	✓	Wrong
4. Take the antihistamine and save your child	Perm.	✓	✓	✓	Wrong*		Perm.
5. Take the medicine from Katerina	Wrong	✓	✓	✓	Wrong	✓	Wrong
6. Get Katerian to give you the medicine	Perm.	✓	✓		Perm.	✓	Wrong*

Note: Asterisks denote instances in which the evaluation generated by the hypothesis does not match the intuitive evaluation.

One could try to salvage the causal account by building normative content into *killing* and *letting die*. Thus, in "Killing, Letting Die, and the Trolley Problem," Judith Jarvis Thomson suggests that (in cases of this sort) intervening counts as killing only if the person who dies has a certain *claim* which is violated.[4] And this is compatible with our own solution, as outlined below. In Cases 3 and 5, you are responsible not only for theft, but also for the death of Nikolay's child. So we are happy to agree with Thomson that you are killing Nikolay's child in Cases 3 and 5 and not in the other cases, where "killing" implies responsibility in the way she describes.

At this point, however, we require an explanation for *why* these claims exist in Cases 3 and 5 but not in the other cases. The distinction between killing and letting die does not give us that explanation. (The "wrong" cases involve killing because they also involve a wrongful taking—it is not as though there is a wrongful taking because there is a killing.) So, although our modified causal account can *accommodate* the difference between the cases, it cannot *explain* it.

If we cannot account for the difference between Cases 5 and 6 solely in terms of how you will affect Nikolay's child, then it seems the difference must be in your actions themselves. And that is just what the cases suggest *prima*

4 Thomson, "Killing, Letting Die, and the Trolley Problem," 209–11.

facie. In Case 5, you violate Katerina's right to use her medicine to save Nikolay's child; in Case 6, you do not. A successful solution to our problem, then, must make some appeal to property rights.

3. ONE SUCCESSFUL SOLUTION

Aquinas's view gives us just this sort of solution. Aquinas accepts claim 3; in his view, it is not stealing to take something from someone when you need it desperately and she does not need it at all. And in all of our cases, you have just this sort of desperate need to save your child's life. In Cases 1 and 4, you need something that Grushenka (who currently owns it) does not need. So, in these cases, it is permissible for you to take it from her. But in Aquinas's view, there are constraints on what you can do even in such great need. One such constraint is against theft—and it *is* stealing to take something from someone when she needs it as well.[5] In Cases 3 and 5, Nikolay and Katerina need the medicine for the same reason you do, so it would be wrong for you to take it. (We say Nikolay and Katerina need the medicine because, as we use it here, the term "need" has application to a person who has an important use for something, though that use may be for someone else's sake. Nikolay and Katerina need the medicine to save someone's life. Grushenka, who is using the medicine merely for decoration, has no such need.)[6] In Case 2, there is no question of taking

5 Perhaps you would rather not restrict the word "theft" in this way—that is, to the taking of things you do not need or that someone else needs. But whatever you think about that English word itself, we ought to acknowledge a distinctive act (the one Aquinas picks out with his word "*furtum*") that differs from other ways of taking what is not yours in that it gives rise to a moral constraint.

6 We are leaving "need" here unanalyzed. For logical purposes, its place in our argument is as a primitive. Partly, this move is just pragmatic: we cannot follow every trail of explanation as far as it would take us. But we also want to note that need does seem to play a similarly ineliminable and foundational role in other norms of justice. You may break a promise when you need to, for example, but not when breaking it would merely have better results. And we seem to have duties to rescue those in great need that are not just stronger cases of general duties to benefit people.

To be sure, this leaves unresolved many questions about when a person really needs something. As one anonymous reviewer asks, what if the disease in our cases had been far less serious, but chronic and uncurable except with the medicine? Or what if the disease had been harmless except in rare but fatal cases?

We confess some discomfort with the vagueness of *need* here. For what it's worth, Aquinas acknowledges this vagueness himself in *ST* II-II 32.6, where he distinguishes necessity of two sorts, and he explains that a person may gain or lose much without falling into either excess or deficiency in necessities of the second sort. There is more to be said on that point, and much more work to be done in general on the sorts of difficult questions we just mentioned. Unfortunately, we will have to leave them aside here. In the meantime,

another's property to begin with. It seems, then, that the Thomistic solution gives us the right result in each of our cases.

Table 3

Action	Intital Evaluation	Arithmetic Hypothesis Evaluation	Causal Hypothesis Evaluation	Thomistic Hypothesis It Is Theft	Thomistic Hypothesis Evaluation
1. Take the medicine from Grushenka	Permissible	Permissible	Permissible		Permissible
2. Give your child the medicine	Permissible	Permissible	Permissible		Permissible
3. Take the medicine from Nikolay	Wrong	Wrong	Wrong	✓	Wrong
4. Take the antihistamine and save your child	Permissible	Wrong*	Permissible		Permissible
5. Take the medicine from Katerina	Wrong	Wrong	Wrong	✓	Wrong
6. Get Katerian to give you the medicine	Permissible	Permissible	Wrong*		Permissible

Note: Asterisks denote instances in which the evaluation generated by the hypothesis does not match the intuitive evaluation.

Aquinas also accepts claim 1: he thinks the constraint against theft is absolute. But our argument is neutral on that question. Perhaps you think it would be permissible to commit genuine theft if the stakes were higher—if, for example, you could have saved a hundred children by taking Nikolay's medicine in Case 3.[7] Perhaps you would be right. All our argument requires is that theft is subject to a constraint, absolute or not.

Note also that we are not arguing at this point for Aquinas's general view of property rights. All we have tried to show is that any adequate theory of property must agree with Aquinas that, in cases of need, you do not violate any property rights if you take from others' overabundance, but you *do* violate property rights if you take from those who are also in need. Need thus plays an indispensable role, not just in determining when a property right can be permissibly violated, but in defining the scope of property rights themselves.

---

we believe that the concept of need remains useful for purposes such as ours even before it has been fully elucidated.

7 In fact, our argument is even neutral on whether there are other conditions under which taking what you need from someone else is not an act of theft. For example, you might think it is not theft to take from one person what she needs in order to save a greater number of people. We do not see any reason to adopt such a view, but nothing in our argument rules it out.

4. AN IMPLICATION FOR ACCOUNTS OF PROPERTY

If this is all correct, then we have accomplished our primary goal for this paper. We have shown that there is good reason to accept claim 3: it is not theft for those in need to take from those with plenty. We have provided an argument for a general right of necessity.[8]

Claim 3 has important implications for a range of questions in political philosophy. For example, it defuses one intuitive objection to the idea that the constraint against theft is absolute.[9] But in the remainder of this paper, we want to transition from our argument's implications regarding theft to its further and more general implications for accounts of property. In this section, we show that our argument constrains which accounts of property we should accept because only accounts of property of a certain sort are equipped to explain 3. Then, in the next and final section, we argue that many of the accounts of property that are equipped to explain 3 also have rather radical implications regarding the duties of the wealthy to give to the poor.

Consider, then, what our argument implies for accounts of property. It shows that there is an intimate connection between what you need and what you have rights to make use of. Whatever account of property one adopts, it should have the resources to explain this connection—this right of necessity. An account of property that cannot explain the right of necessity has at least our argument to be counted against it.

Take, for example, the Lockean view articulated by Robert Nozick. Nozick seems to allow for something like the right of necessity in his explanation of the "Lockean Proviso," which he articulates first in the case of initial acquisition, and then by extension to ownership. The principle is this: ownership is unjust (and so illegitimate) if someone else's situation is worsened on balance because of it.[10] Note that, in place of need, this principle appeals to a comparison between

8 We mean the sort of right of necessity discussed in Mancilla, "What the Old Right of Necessity Can Do for the Contemporary Global Poor," 607–20.

9 For another example, it implies that ownership must essentially be a relation to persons rather than merely to property. After all, it is not theft for a starving man to "steal" bread, nor for someone else to "steal" bread on a starving man's behalf, but it *is* theft for a well-fed man to steal bread for himself. If ownership were only a relation to property, it would not switch on and off like this between one person and another. So, then, our argument also lends support for Kant's claim that "speaking strictly and literally, there is also no (direct) right to a thing. What is called a right to a thing is only that right someone has against a person" (Kant, *The Metaphysics of Morals*, 6:261).

10 Nozick, *Anarchy, State, and Utopia*, 174–82. Nozick's principle also allows for acquiring something even if others suffer because of it so long as one makes fair restitution. Our argument here works irrespective of that qualification.

the state a person finds herself in *prior to* someone else's ownership of some piece of (otherwise common) property and her state *posterior*.

Now, how we apply this Lockean principle depends somewhat on Nozick's notions of better and worse, and what advantages are legitimately attributed to a thing's being held in common. But no matter what Nozick's view has to say on those questions, it will yield one of the following two results. Either Nikolay's claim to the medicine will be *illegitimate* in Case 3, since by owning it himself he makes you (or your child) worse off; or else, if we are *not* to understand Nikolay's claim in Case 3 as making you worse off (and therefore illegitimate), then the same will be true of Grushenka's claim to the antihistamine in Case 4. But that is the wrong result. In Case 3, Nikolay has a legitimate right to the medicine which Grushenka does not have in Case 4. Without appealing to need, Nozick's account will not be able to distinguish in the right way between Nikolay's claim to the medicine (or Katerina's, or yours) and Grushenka's claim.

The right of necessity, correctly understood, is a right of those in need against those with plenty (as suggested by Cases 1 and 4). But importantly (as suggested by Cases 3 and 5), it is not equally a right against others who are also in need. So an account of property must explain not only why those in need can take from those with plenty, but also why those in need *cannot* take from others who are also in need. This would rule out, for example, accounts like that of Thomas Hobbes. Hobbes defends the right of necessity by arguing that those in great need no longer stand to benefit from the laws of justice, and so, for them, the authority of justice is dissolved.[11] This does indeed explain why those in need would be permitted to take from those with plenty, but it would equally justify those in need taking from others also in need. So Hobbes's account also fails to provide an adequate explanation of the right of necessity.

What, then, does it take for an account of property to appropriately explain the right of necessity? Let us consider two examples, beginning with Aquinas's account.

11 Hobbes argues thus:

> If a man by the terrour of present death, be compelled to doe a fact against the Law, he is totally Excused; because no Law can oblige a man to abandon his own preservation. And supposing such a Law were obligatory; yet a man would reason thus, "If I doe it not, I die presently; if I doe it, I die afterwards; therefore by doing it, there is time of life gained;" Nature therefore compells him to the fact.
>
> When a man is destitute of food, or other thing necessary for his life, and cannot preserve himselfe any other way, but by some fact against the Law; as if in a great famine he take the food by force, or stealth, which he cannot obtaine for mony nor charity; or in defence of his life, snatch away another mans Sword, he is totally Excused, for the reason next before alledged. (Hobbes, *Leviathan*, ch. 27)

According to Aquinas, under natural law, everything a person needs in order to live is hers, given to her by God. All necessities belong collectively to those for whom they are necessities. Any further claims to possession are grounded in conventional human law. This law is perfectly legitimate, but only so long as it does not conflict with natural law.[12] If, for example, I possess a loaf of bread that I need, I have rights to the bread that are grounded in natural law (and perhaps also in human law). If I possess a loaf of bread that I do not need, I may have a right to the bread grounded in human law, but only so long as someone else does not have a right to it grounded in natural law—for a human law in conflict with the natural law is unjust, and thus forfeits its authority. Now, if there are several people who need the loaf of bread (which is mine under human law), then I may choose whom to give it to, since, as Aquinas says, "to each of us is committed the stewardship of his own things."[13] That is, my bread is due to those who hunger, but since it cannot feed them all, and since in the meantime it is mine to distribute as I think best, no one may take it from me any more than I may keep it for myself (unless, of course, I refuse to distribute it as I am obligated to). Here, I am in the position of Katerina in Case 5.

Another account which would reconcile our six cases is given by Hugo Grotius. Grotius distinguishes two schemata of property, one of which is tied closely to need, the other of which is not. Of the first, the "primitive schema," Grotius says "God conferred on humankind in general a right to the things of this inferior natural order" such that "whatever someone had taken to himself, another could not take from him except in wronging him."[14] And because this world is given in common to all for satisfying human need, the scope of the primitive schema is limited by need. So, on the primitive schema, we have a certain right to "use the things in the common, and to use them up inasmuch as nature demands it."[15] That is, our property rights are absolute, but only so long as our property is necessary to us. The second, the "private schema," is created by our collective agreement to further divide up natural holdings (to allow for industry, innovation, and so on). But this secondary system of property rights might not refer at all to necessity in determining who owns what. Indeed, that it allows us to possess luxuries is part of the point. So, in this schema, need plays no essential role.

For Grotius, the right of necessity is explained by the fact that the private schema is instituted with the implicit intention that it not deviate from the primitive schema. It therefore has built into it an implicit exception "in a case

12 Aquinas, *ST* II-II 66.7.

13 Aquinas, *ST* II-II 66.7.

14 Grotius, *De jure belli ac pacis*, II.II.I.1. All translations are by Tucker Sigourney.

15 Grotius, *De jure belli ac pacis*, I.II.I.5

of dire need."[16] And this is why he concludes that it is not theft for a starving man to take bread from those with plenty, but it is theft to take from those who are themselves in need.

There is an important similarity between Aquinas's and Grotius's accounts. Both explain the right of necessity—how it arises, and also the fact that it does not apply against others who are also in need—by distinguishing two systems. Within the first, or fundamental, system (natural law for Aquinas, the primitive schema for Grotius), need plays a central role in the establishment of property rights. Within the second, or posterior, system (human law or the private schema), there is no such central role for need, but since the second system is truly *second*, in cases of need, it is constrained or superseded by the first system. In this way, we get an explanation for our judgments about the permissibility of taking what you need from those who do not need it. But—and here is the point—the explanation depends on the thought that this act of taking is permissible because it *is not stealing*: it is a taking of something to which you have a certain right, and this right is prior to the right of the one from whom you are taking. In this way, necessities *belong* to those who need them.[17]

5. A Further Implication for the Duties of the Wealthy

So far, we have said that an adequate account of property must explain the right of necessity, and that accounts that explain it successfully will invoke a distinction between property rights: one set of rights does not depend on need (of the sort Grushenka has to her medicine in Case 1), while a deeper or more fundamental set of rights does depend on need (of the sort you have to Grushenka's medicine). Thus, these accounts admit a sense in which necessities belong to those who need them, and they explain why this should be so. (And, of course, they also acknowledge a sense in which necessities belong to those now in possession of them, needing them or not. Otherwise, there could be no duty to give them.)

These theories have a certain elegance to them, insofar as they make what you are permitted to take and use simply a matter of what belongs to you, in

16 Grotius, *De jure belli ac pacis*, II.II.VI.2.

17 Our argument here is an argument to the best explanation. Of course, Aquinas's and Grotius's are not the only accounts of this sort, much less are theirs the only possible theories of property that can answer to our argument. Our claim is only that this general structure that they share, on which the right of necessity is explained by a foundational principle by which rights to own and use things are given first to those who need them, is the most natural and elegant way to accommodate our six cases. In fact, it is the only plausible way to accommodate them that either of us knows of.

cases of great need just as in ordinary cases. And while it is too strong to say this *proves* anything about the duties of the wealthy—for matters of this sort always give rise to complications—it does strongly suggest that, just as those in great need are entitled to take from the superabundance of the wealthy, equally the wealthy are obligated to give from their superabundance to those in need.

Consider Aquinas's view. For Aquinas, the natural world is given to us all by God for the purpose of meeting our needs. This is a first principle of the natural law. Aquinas then distinguishes two ways in which you might possess something. You possess something *simply* so long as you have any kind of standing to use or dispose of it; you possess something *for yourself* when you have standing to use it for your own ends. The manager of a blind trust, for example, possesses the trust funds *simply* but not *for herself*. She has a certain standing to use them, but not for her own private ends. Similarly (for Aquinas), a wealthy glutton possesses his extra bread *simply* but not *for himself* because, under natural law, the bread is his to use only for the purpose of succoring the poor.[18] And this obligation of the glutton's arises straightforwardly from the natural law, by which the bread—a part of this natural world just like anything else—is given first to those who need it. (This, of course, is the very same principle that would ground a starving man's right to take the bread.) For these reasons, Aquinas holds that if something of ours is a luxury for us but a necessity for someone else, and if we knowingly keep it from her, we are thereby committing theft. He writes:

> For this reason, those things which someone has in superabundance are due by the natural law to the sustenance of the poor. Hence Ambrose says (and it is found in the Decretals, dist. 47), "it is the bread of the hungry which you hold back; it is the clothing of the naked which you store away; the ransom and absolution of those in distress is the very same money which you bury in the earth."[19]

So, for Aquinas, the very principle that explains the right of necessity also grounds a duty for the wealthy to give from their abundance to the poor.

The same is true of Grotius. In cases of need, property rights are given by the primitive schema—and within that schema, I have rights to objects only "as far as nature requires," i.e., as far as I have need. So Grotius explains a starving

18 This is only the mouth of a much deeper cave, of course, and it is more than we can do here to explicate a full account of property. But it is worth noting that these same distinctions are drawn by others. Kant, for example, distinguishes between "possession"—merely dispositive control—and "use"—standing to use something for one's own purposes—in his account of parental right (6:281–82). He argues that parents *possess* their children: they have the right to manage and direct them, but only for the children's own good, not for the parents' (6:281–82).

19 Aquinas, *ST* II-II 66.7.

man's right to take bread by appealing to a special normative context within which (in some sense) he, and not the wealthy man, has a right to the bread. Here, too, the same principle that justifies the starving man in taking the bread equally undermines the wealthy man's right to keep it for himself.

On Thomas Hobbes's account, by contrast, the right of necessity arises only because *all* duties of justice are dissolved for those in great need. In this sense, necessity returns the desperate to a state of war. Hobbes's account does not entail any duty on the part of the wealthy to give to those in need, but as we saw, for the very same reasons, it also fails to imply that those in need cannot take from others in need.

In summary, then, it seems that accounts which successfully explain a right of necessity, such as those of Aquinas and Grotius, are successful precisely because they invoke some principle on which things belong first to those for whom they are necessities—and this principle grounds a right of the needy to take from the wealthy as well as an obligation of the wealthy to give to the needy. Accounts that invoke no such principle need not imply any such obligation for the wealthy, but neither do they successfully explain the right of necessity. At a minimum, then, we should acknowledge a sense in which, in Case 1, the medicine sitting on Grushenka's shelf is *yours to use* (in a deeper sense than that in which it is Grushenka's), whereas in Case 3, the medicine on Nikolay's shelf is not.

Of course, one might still resist this implication. It may be that the correct account of property captures the contours of the right of necessity without also entailing duties of giving for the wealthy. Still, we have suggested that accounts of property that adequately explain claim 3 seem to have a symmetry to them—arising from a principle on which necessities belong first to the needy—that equally grounds a right for the needy to take what they need and an obligation for the wealthy to give it. We have not ruled out the possibility of another account that could adequately explain 3 without invoking the Thomistic (or Grotian) symmetry.[20] But we have, at least, restricted the range of adequate theories of property such that the symmetric accounts represent a larger share of the remaining options. And, dialectically speaking, if Aquinas's somewhat radical conclusion is to be rejected, we have also shown the need for an account of this sort, which explains 3 without the Thomistic symmetry. Without any such account at the ready, resistance to the Thomistic conclusion

20 This, of course, raises the question as to what would be required to establish (rather than merely suggest) this duty for the wealthy to give to those in need. What this would require (so it seems to us) is just to establish the correct account of the normativity of property. If we could establish the correct explanation of the fact that property is governed by the right of necessity, we could then simply look and see whether that explanation implies a strong duty of charitable aid.

seems unmotivated. (Or perhaps all too motivated.) We are left with a picture that, beside the Thomistic one, appears *ad hoc* and incomplete.[21]

As our title suggests, these considerations leave us with a position reminiscent of Peter Singer's in "Famine, Affluence, and Morality."[22] In fact, Singer himself cites Aquinas as a philosopher who shares his "radical" conclusion, referring to the same passage we quoted above.[23]

Although Singer and Aquinas do argue for similar obligations to dispense wealth, there are two important differences between their positions. First, Aquinas distinguishes two ways in which we can give to the poor. On the one hand, we can give from what we have in surplus to alleviate need. We might donate unneeded income to an effective aid organization, for example. For Aquinas, this is a matter of justice, subject to what he calls a "precept." On the other hand, we could give from what we do need—or we could go further by seeking out new ways to help the poor. We might forsake philosophy to pursue data science, for example, in order to double the amount of income we can donate. And while there is good reason in Aquinas's view to do this sort of thing, it is reason of a different kind. It is a matter of charity and is subject not to a precept but to a counsel. So, for Aquinas, a failure to give in the first way is theft, and therefore an injustice, whereas a failure to give in the second way might be a failure of charity. Singer does not acknowledge this distinction in ways of giving. For him, spending too much on a watch is the same kind of failure as taking up a more enjoyable but less lucrative career. Both are failures to sacrifice something of lesser moral value in order to prevent great suffering.[24]

Second, Aquinas acknowledges a principled difference between need and other kinds of lack, whereas Singer does not. Suppose all serious poverty were suddenly eliminated from the world, but significant income disparity remained.

21 Of course, it is true of any argument that the reader may find it less plausible to accept the conclusion than to deny a premise. Perhaps, for example, a reader will be inclined to deny that Grushenka acts wrongly if she keeps her medicine on her shelf instead of using it to save someone's life. But, in the opinion of the authors, this sort of thing should be done only with great caution, given the general human tendency to rationalize immoral behavior. We, like most of our readers, recognize that we are wealthy by the standard of need. We therefore have more reason to doubt our intuitions that we are permitted to keep our luxuries than to doubt our intuitions that the poor are permitted to take what they need from us.

22 Singer, "Famine, Affluence, and Morality," 229–43.

23 Singer, "Famine, Affluence, and Morality," 239.

24 What grounds this difference is the fact that the Thomistic view is exclusively a view about *property*. This has other intuitive implications that separate it from Singer's view. For example, just as the Thomistic view would not require Katerina to give her labor, neither does it require her to give her kidneys, since neither her actions nor her organs are *property* in the relevant sense. (Aquinas makes this point in *ST* II-II 66.3.)

In Aquinas's view, a wealthy person—call him "Bill"—would still have reasons to give to the less fortunate, but again these would be reasons of a different kind: charity or liberality, perhaps, rather than justice. Maybe Bill should use some of his wealth to supply nicer homes for families crammed into apartments, for example, but he would not be stealing from those families if he refused. For Singer, on the other hand, although death and destitution are especially bad things, they are of the same stock as any other bad thing. As long as it does not require sacrificing something nearly as important, Bill has the same kind of obligation to supply better houses for the citizens of his wealthier world (supposing that is an effective use of his money) as we have now to supply basic medical care to the citizens of rural Madagascar.

So Aquinas endorses a difference in kind between, on the one hand, giving to the poor from our surplus, and, on the other, devoting our lives to the poor or giving to those not in extreme need. Though Aquinas's articulation of that difference is tied up with the specifics of his view—distinctions between precept and counsel, justice and charity, and so on—we can also put the difference in general terms. If it is wrong, say, for me to study philosophy instead of data science, it is wrong because I am using my own life poorly. And if it is wrong for Bill to upgrade his mansion rather than upgrade a family's crowded apartment, it is wrong because he is using his own money poorly. But if I keep my surplus income when others are in desperate need, then I am not just using *my own* resources poorly. Rather, I am misusing resources that properly belong to *someone else*. This is a further wrong—an injustice—beyond selfishness or thoughtlessness.

We take it to be an advantage of Aquinas's account that it accommodates this distinction. For one thing, it takes seriously our intuitions that, although we might sometimes have obligations to give up a career in philosophy to better serve the poor, that obligation would be different (and generally weaker) than an obligation of justice. But, on the other side of that same coin, it also takes seriously just how deep a failure it is to withhold from the poor what they need. If Aquinas is right, then a great deal of what we keep and spend is not just ill-used, but the very ransom and absolution of those who are in distress.[25]

Florida State University
marshall.bierson@gmail.com
tssigourney@gmail.com

25 We would like to thank Bob Bishop, Paul Rezkalla, John Schwenkler, and Spencer Smith for their valuable contributions to our thinking on the matter of this paper, as well as a group of high school students at the Victory Briefs Institute, who were the first to engage with the argument.

REFERENCES

Aquinas, Thomas. *Summa Theologiae*. Rome: Typographia Polyglotta, 1888–1906.

Grotius, Hugo. *De jure belli ac pacis*. Paris: Nicolas Buon, 1625. Published online by the Bibliothèque nationale de France. https://gallica.bnf.fr/ark:/12148/btv1b86069579/.

Hobbes, Thomas. *Leviathan*. London: Andrew Crooke, 1651. Published online by Project Gutenburg, 2002. https://www.gutenberg.org/files/3207/3207-h/3207-h.htm.

Kant, Immanuel. "The Metaphysics of Morals." In *Practical Philosophy*. Translated and edited by Mary J. Gregor. Cambridge: Cambridge University Press, 1996.

Mancilla, Alejandra. "What the Old Right of Necessity Can Do for the Contemporary Global Poor." *Journal of Applied Philosophy* 34, no. 5 (November 2017): 607–20.

Nozick, Robert. *Anarchy, State, and Utopia*. Oxford: Blackwell, 1974.

Singer, Peter. "Famine, Affluence, and Morality." *Philosophy and Public Affairs* 1, no. 3 (Spring 1972): 229–43.

Thomson, Judith Jarvis. "Killing, Letting Die, and the Trolley Problem." *The Monist* 59, no. 2 (April 1976): 204–17.

Journal of Ethics and Social Philosophy
VOL. 25, NO. 2 · AUGUST 2023
https://doi.org/10.26556/jesp.v25i2.1726

PERSONAL REACTIVE ATTITUDES AND PARTIAL RESPONSES TO OTHERS

A PARTIALITY-BASED APPROACH TO STRAWSON'S REACTIVE ATTITUDES

Rosalind Chaplin

STRAWSON'S discussion of the reactive attitudes in "Freedom and Resentment" distinguishes between three types of reactive attitudes: the *personal*, the *impersonal*, and the *self-reactive* attitudes.[1] According to Strawson, personal attitudes paradigmatically reflect our concern that we ourselves are treated with good will and regard; impersonal attitudes paradigmatically reflect our concern that others receive the same good will and regard that we demand for ourselves; and self-reactive attitudes paradigmatically reflect the demand we make on ourselves to treat others with good will and regard.[2] Thus, when someone insults *me*, I react with the personal attitude of resentment; when someone insults a person sitting next to me on the bus, I react with an impersonal attitude, such as moral indignation or disapprobation; and when *I* am the insulter and later judge my own behavior to be unacceptable, I react with a self-reactive attitude, such as guilt, shame, or remorse.

1 For ease of reference, all citations of Strawson's "Freedom and Resentment" refer to page numbers in Pamela Hieronymi's recent reprint in *Freedom, Resentment, and the Metaphysics of Morals*.

2 Note that these distinctions do not track Strawson's distinction between the *participant* attitude (or stance) and the *objective* attitude (or stance). Rather, the personal, impersonal, and self-reactive attitudes are all species of participant attitudes for Strawson. Thus, the objective attitudes are properly characterized not as *impersonal* but rather as *non*reactive in Strawson's view. In this paper, I do not take a stand on exactly how the reactive attitudes as a general class are to be characterized, though I do adopt Strawson's claim that they are responses to the quality of will displayed in our and others' actions (remaining agnostic as to exactly how quality of will is to be understood). And in line with Strawson, I also adopt a relatively permissive approach to which attitudes can count as reactive, assuming (against someone like Wallace in *Responsibility and the Moral Sentiments*) that reactive attitudes can pertain to attributions of responsibility in both the *attributability* and *accountability* senses articulated in Watson's "Responsibility and the Limits of Evil."

A number of philosophers working in the Strawsonian tradition have taken this tripartite taxonomy to reflect a more general commitment to distinguishing between reactive attitude types according to *who* is the target of good or ill will and *who* is the subject displaying good or ill will. Hieronymi's recent characterization typifies this approach, though similar claims can be found in the works of various other prominent moral responsibility theorists.[3] Hieronymi writes:

> In general, then, a reactive attitude is *x*'s reaction to *x*'s perception of or beliefs about the quality of *y*'s will toward *z*. In the impersonal reactive attitudes, *x*, *y*, and *z* are different persons. In the case of the personal reactive attitudes, the same person stands in for *x* and *z*. In the case of self-directed attitudes, the same person stands in for *x* and *y*.[4]

Notice that if we accept this characterization without emendation, then a personal attitude, such as resentment, always occurs as a reaction to one's own treatment; an impersonal attitude, such as moral indignation, always occurs as a reaction to the treatment of another; and a self-reactive attitude, such as shame, always occurs as a reaction to one's own treatment of others.[5] One may even be tempted to say that what *makes* resentment different from moral indignation is that the former expresses a kind of self-concern, whereas the latter expresses our concern for others (and so on for other attitude pairs).

However, close readers of Strawson will immediately notice that this way of interpreting Strawson's taxonomy cannot be completely correct, for Strawson himself says that "one can feel indignation on one's own account."[6] If one can feel indignation on one's own account (i.e., for a wrong done to oneself), then the difference between indignation and resentment cannot be that indignation concerns a wrong done to another, whereas resentment concerns a wrong done to oneself. Likewise, it cannot be the case that indignation *essentially* expresses concern for another, whereas resentment *essentially* expresses concern for the self. Strawson's own remarks thus provide some reason to reassess

3 See Helm, *Communities of Respect*, ch. 3; McKenna, *Conversation and Responsibility*, 66; Wallace, *Responsibility and the Moral Sentiments*, ch. 2; and Watson, "Responsibility and the Limits of Evil," 223n4.

4 Hieronymi, *Freedom, Resentment, and the Metaphysics of Morals*, 8.

5 In this paper, I focus on shame rather than guilt because of what I take to be the special plausibility of claiming that shame can concern the behavior and character of others as well as oneself. In contrast, I believe guilt is linked to making amends in a way that does, in fact, limit it to our own wrongdoing. This said, I am open to the possibility that guilt has a wider scope than I have acknowledged, and I thank an anonymous referee for convincing me that one could reasonably disagree on this point. See note 25 below for further discussion of how the intuitions I have just registered concerning guilt are consistent with my overall proposal.

6 Strawson, "Freedom and Resentment," 121.

the standard interpretation of the distinction between personal, impersonal, and self-reactive attitudes.

In this paper, I propose an improved way of understanding Strawson's distinction between fundamental reactive attitude types, and I argue that this new alternative better captures the core insight animating Strawson's discussion in "Freedom and Resentment." What matters most in Strawson's original framework is not whether an attitude arises as a response to the treatment of the *self* or *another*, or whether an attitude is directed at the *self* or *another*. Instead, what matters most is whether an attitude expresses *partial* or *impartial* concern. Resentment is an attitude expressing partial concern, and this is what distinguishes it from attitudes such as moral indignation, which Strawson calls "impersonal." In contrast, moral indignation arises from our impartial concern that all people are treated with good will and regard, and this impartial character (rather than its other-concerning nature) is what distinguishes it from resentment. In principle, we can have impartial concern both for others and for ourselves, and this insight is at the heart of Strawson's claim that one can have indignation on one's own account.

I also argue that once we distinguish between reactive attitudes according to their partiality or impartiality (rather than according to whether the subject and target of the attitude is the self or another), we are better able to accommodate an important fact about our moral lives; namely, many reactive attitudes have a wider scope than is often acknowledged, and the attitudes that express partial concern play an especially important role in the maintenance of our close personal relationships. For although attitudes like resentment and gratitude (Strawson's "personal" attitudes) *can* reflect the special concern we have for our own treatment, they can also reflect the special concern we have for our close ties. Similarly, attitudes such as shame (Strawson's "self-reactive" attitudes) *can* reflect our interest in our own behavior and character, but they can also reflect the special interest we have in the behavior and character of those with whom we stand in close relationships (and can similarly have third-party manifestations). The traditional characterization of reactive attitude types (adopted by Hieronymi and others) obscures these facts, whereas a bipartite distinction between attitudes that express partial and impartial concern sheds light on them.

The plan for the paper is as follows. In section 1, I briefly review the elements of Strawson's discussion that have motivated the traditional way of characterizing his distinction between basic reactive attitude types, and I highlight some further parts of "Freedom and Resentment" that suggest he may have embraced a different picture. In section 2, I present four cases that motivate the conclusion that the attitudes Strawson calls "personal" and "self-reactive" are, in fact, unified by a common characteristic: they reflect the *partial* concern we have for the treatment and behavior of certain agents (including but not limited to ourselves).

If my assessment of these cases is correct, then attitudes like resentment, gratitude, pride, and shame have a wider scope than the standard characterization of Strawson's taxonomy allows for. In section 3, I consider whether my proposed expansion in scope can be undermined by the claim that a particular kind of concern for the *self* always grounds our partial concern for others. If this is the case, then the attitudes Strawson calls "personal" may always concern our *own* treatment after all, and the attitudes he calls "self-reactive" may always concern our *own* behavior (or character traits, depending on the case). I respond to this objection by showing that the cases discussed in section 2 involve concern for the treatment and behavior of others *for their own sakes* (i.e., independently of any effects on us). Given this, they cannot plausibly be reduced to cases of self-concern. Finally, in section 4, I close with some reflections on how a bipartite, partiality-based taxonomy of fundamental reactive attitude types relates to Strawson's claim that only the "impersonal" attitudes deserve "the qualification 'moral.'"[7] In saying this, I take Strawson to be expressing his commitment to the idea that morality demands impartiality, for he thinks the *impersonal* attitudes are uniquely "moral" because they express *impartial* demands. While my discussion of the role of partial reactive attitudes in helping us fulfill our relationship-based obligations may be in tension with this part of Strawson's view, it clearly fits his understanding of the nature of the basic reactive attitude types—namely, as expressing either partial or impartial concern.[8]

1. The Standard Tripartite Taxonomy of Strawson's Reactive Attitudes

As Hieronymi's discussion (quoted above) makes clear, it is tempting to think that Strawson appeals to a distinction between *self* and *other* to generate the tripartite taxonomy of reactive attitudes that has become the standard reading of his basic classificatory scheme. And indeed, Strawson's own characterizations sometimes suggest that he intends for the distinction between basic reactive attitude types to be understood in this way. For instance, he writes that the

7 Strawson, "Freedom and Resentment," 121.

8 Though I argue below that the partial reactive attitudes help us fulfill the demands of our close relationships, I hope to remain ecumenical as to the nature of reactive attitudes more generally. For instance, I think reactive attitudes often play communicative roles and often make demands, but they need not always do so. I also remain neutral on whether they rest on or include judgments and on whether they must be expressed or sometimes can be privately held. For a variety of different views on these issues, see Helm, *Communities of Respect*; Hieronymi, "Articulating an Uncompromising Forgiveness"; Macnamara, "Reactive Attitudes as Communicative Entities"; McKenna, *Conversation and Responsibility*; Wallace, *Responsibility and the Moral Sentiments*; and Watson, "Responsibility and the Limits of Evil."

personal attitudes are "essentially those of offended parties or beneficiaries" and that they are "essentially reactions to the quality of others' wills towards us, as manifested in their behaviour."[9] When introducing the impersonal attitudes, he describes them as "reactions to the quality of others' wills, not towards ourselves, but towards others," and he initially glosses moral indignation as "resentment on behalf of another, where one's own interest and dignity are not involved."[10] Finally, turning to the self-reactive attitudes, Strawson says: "Just as there are personal and vicarious reactive attitudes associated with demands on others for oneself and demands on others for others, so there are self-reactive attitudes associated with demands on oneself for others."[11]

Thus, there is good reason to take seriously the idea that Strawson bases his taxonomy of reactive attitude types on the distinction between self and other. His discussion does, at times, seem to suggest that personal reactive attitudes reflect our interest in how others treat us; impersonal reactive attitudes reflect our interest in how others treat others; and self-reactive attitudes reflect our interest in how we treat others. Summarized in a table, this standard picture looks as follows:

Table 1

Personal reactive attitudes	Self-reactive attitudes	Impersonal reactive attitudes
Examples: · resentment · gratitude · hurt feelings · reciprocal love · forgiveness	*Examples:* · shame · guilt · remorse · feeling obligated · pride	*Examples:* · disapprobation · indignation · disapproval · admiration · approbation · approval
Reactions to the demands we make on others concerning our own treatment	Reactions to the demands we make on ourselves concerning others' treatment	Reactions to the demands we make on others concerning others' treatment

Notably, Strawson himself does not mention all the particular attitudes included in this table. For instance, he does not explicitly mention pride, admiration, or approbation. But Strawson qualifies his discussion by remarking that the reactive attitudes belong to a "field of phenomena" "too complex" to be neatly characterized, and as many scholars have noted, it therefore seems in keeping with the spirit of "Freedom and Resentment" to enrich his list, per the table above.[12]

9 Strawson, "Freedom and Resentment," 121.

10 Strawson, "Freedom and Resentment," 121.

11 Strawson, "Freedom and Resentment," 122.

12 Strawson, "Freedom and Resentment," 111. Scholars who advocate adding attitudes like pride, admiration, and approbation to the list of reactive attitudes include Clarke, McKenna,

However, as noted above, Strawson also indicates that his initial characterization of the basic reactive attitude types is, in some respects, "misleading."[13] As he explains, "one can feel indignation on one's own account," and so it cannot be the case that indignation is resentment on behalf of another, as he had earlier claimed.[14] Strawson thus corrects himself by saying that the impersonal reactive attitudes should be understood not as "essentially vicarious" but rather as "essentially capable of being vicarious."[15] What is important is that they express a kind of "disinterested or generalized" concern—that is, they express "the demand for the manifestation of a reasonable degree of goodwill or regard on the part of others, not simply towards oneself, but towards all those on whose behalf moral indignation may be felt."[16] With this in mind, Strawson remarks that the impersonal reactive attitudes are therefore distinctively deserving of "the qualification 'moral.'"[17] Since they reflect our demand that all people as members of the moral community be treated with good will and regard, they express a distinctively moral demand (at least in Strawson's eyes).

If Strawson's initial characterization of a tripartite taxonomy is misleading in these respects, what are we to make of his basic conceptual framework? That is, how should we understand the difference between a "personal" attitude, like resentment, and an "impersonal" attitude, like indignation, and is a better characterization of Strawson's basic taxonomy of reactive attitude types available to us? I turn now to an articulation and defense of a *bipartite* approach. As I argue, Strawson's framework is best understood as distinguishing between reactive attitudes according to whether they express *partial* or *impartial* concern (rather than whether they reflect concern for the *self* or concern for *others*). Thus, the attitudes

---

and Smith, *The Nature of Moral Responsibility*; Helm, *Communities of Respect*; and Watson, "Responsibility and the Limits of Evil." Against this, Wallace has argued in *Responsibility and the Moral Sentiments* that we should restrict the reactive attitudes to resentment, indignation, and guilt, since only these three attitudes are plausibly part of our practices of holding one another morally responsible (per Wallace). However, as Wallace himself acknowledges, this means we must abandon the Strawsonian claim that involvement in interpersonal relationships is inseparable from susceptibility to the reactive attitudes, since resentment, indignation, and guilt do not seem required for interpersonal relationships *as such* (30). Note that although I do not endorse Wallace's restriction of the reactive attitudes, my arguments below still bear on his account insofar as he adopts the traditional approach to distinguishing between personal, impersonal, and self-reactive attitudes.

13 Strawson, "Freedom and Resentment," 121.

14 Strawson, "Freedom and Resentment," 121.

15 Strawson, "Freedom and Resentment," 121.

16 Strawson, "Freedom and Resentment," 121–22.

17 Strawson, "Freedom and Resentment," 121.

Strawson calls "personal" and "self-reactive" express two different kinds of *partial* concern, while the attitudes he calls "impersonal" express *impartial* concern.

2. AN IMPROVED BIPARTITE TAXONOMY OF REACTIVE ATTITUDES

We have seen that if the standard approach is correct, "personal" attitudes such as resentment and gratitude express our concern with our *own* treatment, and "self-reactive" attitudes such as pride and shame express our concern with how we *ourselves* treat others. Only "impersonal" attitudes can arise as reactions to how *others* treat *others.* However, consider the following four cases, which suggest that both the personal and the self-reactive attitudes are capable of being appropriate responses to the treatment and behavior of certain others—namely, those with whom we stand in close personal relationships:

> *Case 1: Resentment*. Your lifelong best friend discovers that his partner has been having an affair. You are outraged, resent your friend's partner, and realize that it will be very difficult for you to forgive him for what he did to your friend.
>
> *Case 2: Gratitude.* Your partner's colleague nominates her for a community service award. Your partner feels that her hard work often goes overlooked, and you know how much the nomination means to her. You are grateful to your partner's colleague and decide to express your appreciation when you next see her.
>
> *Case 3: Pride*. Your brother has a bad habit of snapping at others and making mean or embarrassing remarks when irritated. He regrets this habit and decides it is time to make a meaningful change. After months of incremental progress, you realize that his interactions with others have become noticeably more sensitive and respectful. You know how much effort it took for your brother to change, and you are proud of him for improving.[18]
>
> *Case 4: Shame*. You spend your semester abroad in college with your best friend, Sam. One night, Sam drinks too much and vandalizes a temple.

18 See Philippa Foot's *Virtues and Vices and Other Essays in Moral Responsibility* for one expression of the traditional view that one must take the object of pride to be one's own achievement (76). But as cases like this suggest, pride can be an appropriate response to the achievements of one's loved ones and can play an important role in signaling or expressing a special commitment to one's loved ones.

> You are ashamed of her and worry that you will not be able to have the same kind of friendship anymore.[19]

In each of these cases, agents involved in close personal relationships respond to their close ties' treatment and behavior with attitudes that the standard approach limits to situations involving our *own* treatment (in the case of the personal attitudes of resentment and gratitude) and our *own* behavior (in the case of the self-reactive attitudes of shame and pride).[20] Recall Hieronymi's claim: "a reactive attitude is *x*'s reaction to *x*'s perception of or beliefs about the quality of *y*'s will toward *z*. . . . In the case of the personal reactive attitudes, the same person stands in for *x* and *z*. In the case of self-directed attitudes, the same person stands in for *x* and *y*."[21] If we accept this characterization of personal and self-reactive attitudes, then we cannot account for the four cases just described. And yet in each of these four cases, the agents' responses appear to be both intelligible and appropriate. We might even go as far as to say that their responses are exemplary of *good* relationships of the relevant type. In case 1, your inclination to resent your friend's partner is grounded in the strength of your friendship; as a close friend, you *should* be prepared to stand up for him, and your resenting his wrongdoer promotes this end. In fact, if you responded to your friend's bad treatment in the same manner in which you would respond to the bad treatment of a perfect stranger, we might worry that you have revealed a problematic kind of disinterestedness in your friend. Close friends should not react to one another's injuries in the way in which they would react to the injuries of perfect strangers; rather, they should react with the same kinds of attitudes they manifest in cases involving their *own* bad treatment. Similarly, in case 2, given the nature of your relationship with your partner and the special concern you have for her, it is fitting for you to

19 One might object here that you must feel only disappointment in (or perhaps contempt for) your friend unless you see her behavior as somehow reflecting on you. However, given a close enough relationship between the two of you, I do not find this objection plausible. Consider a case where Sam is your sister. Here, there is not a plausible case to be made that her behavior reflects badly on you since your association with her is not voluntary. Still, shame on your part seems like an appropriate response to her conduct, given that you care about her in the right kind of way. Thus, although I think our friends sometimes reflect badly on us (in which cases we may experience self-directed shame), there are other cases in which our shame responses are fully independent of self-evaluation and, instead, reflect our special concern for our close ties.

20 Here I do not mean to deny that shame often (or even paradigmatically) concerns a person's character. But we often take behavior to reveal character, and so, in practice, a particular instance of behavior is often what prompts shame. I also want to remain open to the possibility that we can be ashamed of behavior directly, i.e., independently of what it reveals about character.

21 Hieronymi, *Freedom, Resentment, and the Metaphysics of Morals*, 8.

respond to her benefactor with gratitude rather than with the more disinterested kind of approval that would be appropriate in contexts involving strangers. Cases 3 and 4 should elicit similar judgments. While it would be strange for you to feel ashamed or proud of a stranger, it is not inappropriate for you to feel pride and shame for your siblings, your children, and your close friends, and these reactions can even reveal the depth of your concern for them.[22]

Assuming these observations are correct, what conclusions can we draw? We can now articulate the following explanation of *why* attitudes like resentment, gratitude, pride, and shame are appropriate in cases such as the ones described above (and why they have a wider scope than the traditional approach allows).[23] In particular, all four of these attitudes evince *partial* forms of concern, and it is their *partiality* that explains why they extend beyond circumstances involving ourselves to situations involving our close ties. Just as we resent our own wrongdoers because of our special concern for our *own* well-being and treatment, so too we resent our loved ones' wrongdoers because of our special investment in *their* well-being. Just as I am grateful to my own benefactors because I have a special reason to care about *myself*, so too I am grateful to my loved ones' benefactors because I have a special reason to care about *them*.[24] Turning to the so-called self-reactive attitudes of pride and shame, we can say

22 Further considerations can bolster the judgment that apt attitudes in situations involving our close ties are different in kind from the attitudes we have as third-party observers in situations involving strangers. For one, the attitudes we have toward our loved ones' wrongdoers come with different ranges of dispositions to action than do the attitudes we have toward strangers' wrongdoers, even when the seriousness of the wrongs are the same (e.g., we may be disposed to intervene in situations involving our loved ones in ways that we would not in cases involving strangers). Second, phenomenological differences in the attitudes arguably are different in kind, rather than merely in degree; i.e., apt responses in impersonal cases involving strangers are not simply less intense manifestations of the same attitudes we have in cases involving our loved ones. When a loved one is wronged, our reactive attitudes do not simply feel like more intense versions of the very same attitudes we have toward strangers' wrongdoers, and this reflects the fact that our loved ones matter *differently*, and not merely *more*, to us than do strangers.

23 What should we say to someone with the impartialist judgment that, intuitions aside, we ought to react in the same way to *everyone* (even when it comes to our reactive attitudes)? Although full discussion of this exceeds the scope of this paper, one point to make is that this would imply a quite radical revision of the Strawsonian approach since it would imply that we really ought to restrict ourselves to the "impersonal" attitudes alone, even when it comes to our own treatment. I thank Jackson Bittick for raising this issue.

24 Here it is important to distinguish between partial concern for another that has an egoistic basis and partial concern for another that does not have an egoistic basis. If I take a special interest in my family member's welfare because of how that person's flourishing stands to benefit me, my partial concern reduces to partial egoistic concern. But I might also have partial concern for my family member *for her own sake*, and I intend for the cases I have

something similar. I have a special interest in my *own* character and quality of will, which explains why I feel pride or shame when I behave especially well or poorly; similarly, I have a special interest in my loved ones' character and quality of will, which explains why I feel pride or shame when *they* behave in ways that are admirable or shameful, respectively.[25] More generally, then, what distinguishes attitudes like resentment and gratitude from attitudes like indignation and approval is that the former, but not the latter, are manifestations of partial concern for the well-being of the wronged or benefited party. Similarly, what distinguishes attitudes like pride and shame from attitudes like admiration and disapproval is that the former, but not the latter, are manifestations of partial concern for the agency or character of the attitude's target.[26]

Notice also that we have a ready explanation as to why partiality is appropriate in each of the four cases: in each case, the agent's personal relationship grounds the appropriateness of a partial response. In case 1, your relationship with your friend makes it appropriate for you to resent his cheating partner, for friends *should* be especially invested in one another's good treatment. In case 2, your relationship with your partner makes it appropriate for you to react with gratitude to her good treatment since intimate relationships entitle us to have special concern for our partners' well-being. Turning to cases 3 and 4, we can similarly say that our relationships with our close ties call on us to respond partially to circumstances that reflect *their* good or bad quality of will, even when our own treatment is not at issue. Indeed, being in a close relationship with someone often requires special concern not only for their well-being but also for the kind of agent they are (the character traits they display in their actions, the quality of will they reveal toward others, and so on).[27] Thus, our close relationships entitle

---

just discussed to be understood in this way; partial concern for our relationship partners for their own sakes grounds the appropriateness of third-party resentment and gratitude.

25 As noted in note 5 above, I do not discuss the attitude of *guilt* because I take it to be tied to making amends in a way that (typically) renders it fitting only as a response to one's own wrongdoing. (We typically cannot make amends for others' wrongs.) Is this a problem for my view, given that guilt is the paradigmatic self-reactive attitude for moral responsibility theorists following in the tradition of Strawson? I think it is not, since in my view guilt remains a fundamentally partial response in the sense that it expresses the special concern we all have for our own conduct and moral agency. This said, I am also open to the possibility that one can, in principle, feel third-party guilt if one is in a position to make amends as a third party.

26 In distinguishing here between agency and character, I mean to leave room for the possibility that we sometimes care about our close ties' behavior because of what their behavior reveals about their character, while at other times we care about their behavior because we care about their quality of will (independently of our conception of their character).

27 Helm invokes a related distinction between partial concern for a person's *well-being* and partial concern for her *identity* ("Love, Identification, and the Emotions"). According to

us to have a special interest both in the well-being and in the agency of our close ties. In the four cases above, these two forms of partial concern are on display.[28]

If these remarks are correct, then we should understand Strawson's fundamental taxonomy of reactive attitude types differently than the standard approach. Instead of classifying attitudes according to whether they concern ourselves or others (i.e., according to their objects), we should first classify attitudes according to whether they express partial or impartial concern. An appropriately formulated basic taxonomy, therefore, looks like this:

Table 2

Partial reactive attitudes (reactions of partial concern)	Impartial reactive attitudes (reactions of impartial concern)
Examples:	*Examples:*
· resentment	· disapprobation
· gratitude	· indignation
· shame	· disapproval
· pride	· admiration
· hurt feelings	· approbation
· reciprocal love[29]	· approval
· guilt	
· remorse	

There are a variety of advantages to adopting this bipartite taxonomy. First, as already noted, it allows us to acknowledge that attitudes such as resentment, gratitude, pride, and shame have a wider scope than proponents of the traditional characterization allow. In many cases, our close relationships call on us to respond with these attitudes even when our *own* treatment or behavior

Helm, in loving a person we care not just about her well-being, but also about her identity, just as we take a special interest in our own well-being and our own identity.

28 Could it be appropriate to have shame and pride responses concerning, say, one's country or other group association? I want to remain neutral on this and commit only to saying that if it is appropriate to feel shame or pride in one's country or other association, then it is appropriate because partial attitudes concerning these groups can be appropriate for their members.

29 I include the qualifier "reciprocal" here to account for the fact that Strawson recognizes a species of disinterested love that does not belong to the participant stance at all. Strawson writes:

> The objective attitude may be emotionally toned in many ways, but not in all ways: it may include repulsion or fear, it may include pity or even love, though not all kinds of love. But it cannot include the range of reactive feelings and attitudes which belong to involvement or participation with others in inter-personal human relationships; it cannot include resentment, gratitude, forgiveness, anger, or the sort of love which two adults can sometimes be said to feel reciprocally, for each other. (Strawson, "Freedom and Resentment," 116)

is not at issue. Insofar as this is the case, it is important that our taxonomy of reactive attitude types makes room for this. Indeed, allowing for this wider scope should be especially appealing to theorists working in the Strawsonian tradition as it allows us to appreciate the role of partial responses like resentment, gratitude, shame, and pride in our close personal relationships, the arena where Strawson noted that our grip on the importance of the reactive attitudes to our ordinary lives is most secure.[30]

Second, a bipartite, partiality-based taxonomy can also explain Strawson's claim that "impersonal" attitudes do not, in fact, always concern wrongs done to others. As we have seen, Strawson writes that "one can feel indignation on one's own account." This means that the difference between indignation (an impersonal attitude) and resentment (a personal attitude) cannot be that one is a reaction to a wrong done to someone else, while the other is a reaction to a wrong done to the self.[31] A partiality-based approach can explain the difference as follows. When I take a disinterested look at my own injury, abstracting from the partial concern for myself that I usually feel, I am able to experience the indignation that I normally feel on behalf of injured parties with whom I have no special ties. That is, though my reactions to my own injuries normally manifest my special interest in my own well-being and dignity (giving rise to resentment), I can, in principle, react to my own treatment in a manner that recognizes the agent-neutral fact that a wrong done to me is morally on a par with the same wrong done to any other person.[32] This is compatible with the idea that an agent who never resented wrongs done to herself would arguably lack an important kind of concern for herself; for both the impartial and the partial attitudes are important insofar as they help us to secure different values in our moral lives.[33]

Finally, a third advantage of a bipartite, partiality-based taxonomy is that it can be flexible as to exactly *which* attitudes have the capacity to be both self- and other-concerning. Above, I have suggested that the partial attitudes of

30 Strawson, "Freedom and Resentment," 111.

31 Strawson, "Freedom and Resentment," 121.

32 In fact, Strawson himself suggests this in a reply to criticisms from Jonathan Bennett: "I freely reaffirm the central importance of that sense of sympathy, and of a *common* humanity, which underlies not only my indignation on another's behalf but also my own indignation on my own" (Strawson, "P. F. Strawson Replies," 266).

33 This is similar in some respects to Wolf's claim in "Morality and Partiality" that considerations of partiality and impartiality often reflect different (and sometimes competing) values. But whereas Wolf stresses that considerations of partiality sometimes reflect our *nonmoral* values, I highlight cases in which it is at least *prima facie* plausible to think that our reasons to react partially *are* moral (being tied to the special obligations we have to both ourselves and others).

resentment, gratitude, pride, and shame are flexible in this regard for many agents (and so do have a wider scope than is often acknowledged). But nothing in the fundamental characterization of attitudes as either partial or impartial demands that *all* agents manifest the reactive attitudes as both self- and other-concerning. Plausibly, there are some partial attitudes (such as guilt and remorse) that most agents experience only as self-directed.[34]

3. Antireductionism about Partial Concern for Others

The arguments above have attempted to show that the "personal" attitudes are, in the first instance, reactions of *partiality* rather than reactions of *self*-concern (allowing that we are often partially concerned with ourselves and our own circumstances). Likewise, the so-called self-reactive attitudes are also reactions of partial concern rather than of *self*-appraisal and can (at least in some cases) arise as responses to the behavior of our close ties. Thus, rather than classifying reactive attitudes in terms of *whose* treatment and behavior they concern, we should instead classify them in terms of whether they are expressions of partiality or impartiality.

However, one way of resisting this argument might go as follows. Whereas I began by pointing to cases where attitudes like resentment, gratitude, pride, and shame are appropriate as responses to the treatment and behavior of our close ties, one might argue that partial concern for others in these cases in fact reduces to a special kind of self-concern. If this is correct, then the cases I have offered as evidence for a partiality-based approach can be assimilated into the traditional tripartite model after all.[35]

We can begin to see how this objection might be articulated by considering what a proponent of the standard approach might say about each of the four cases discussed in section 2 above. In case 1, perhaps you resent your friend's partner only because you see your friend as such an important part of your life that you regard a wrong done to him as a wrong done to you. In case 2, perhaps you feel gratitude toward your partner's coworker because the good of your partner is a part of your good such that any benefit conferred on her is also a benefit conferred on you. In case 3, perhaps you are proud of your brother because you identify with him such that you regard his goals as your goals and his accomplishments as your accomplishments. And in case 4, perhaps you are

34 See notes 5 and 25 above for further remarks relevant to this.

35 Arguments like this can be found in the literature on forgiveness, where some scholars defend the view that putative cases of third-party forgiveness are always hidden cases of victim forgiveness. See Murphy and Hampton, *Forgiveness and Mercy*; Walker, "Third Parties and the Scaffolding of Forgiveness"; and Zaragoza, "Forgiveness and Standing."

ashamed of your friend's behavior because of the way in which her identity is tied up with your own; her behavior reflects badly on you, and its significance for your appraisal is what explains the appropriateness of your shame response. Analyses of cases such as these suggest that the personal and self-reactive attitudes always involve self-concern or self-assessment after all. As the objection goes, it may look like third-party resentment, gratitude, pride, and shame are possible, but a proper understanding of the cases reveals that this is not so; our reactive responses to our close ties' circumstances are grounded in the relationship those circumstances have to our *own* well-being or appraisal. In other words, they are covert cases of *self*-concern and *self*-appraisal.

To assess the plausibility of an objection of this sort, we should first distinguish more carefully between two ways in which it might be interpreted. How should we understand the idea that our concern for the well-being and appraisal of our close ties reduces to our concern for our *own* well-being and appraisal? First, perhaps the objector means to suggest that we care about the well-being and appraisal of our close ties because of the *instrumental* relationship it has to our own good. The proposal concerning case 4 makes especially plausible an account like this, for in that case, it is plausible to think that the friend's behavior has downstream significance for the agent having the shame response. Plausibly, the agent thinks her friend's shameful behavior will cause others to appraise her negatively, in which case her response to her friend's bad behavior would be grounded in the negative instrumental value she thinks it has for *her*. Alternatively, the objector might instead argue that our concern for our close ties is always grounded in the fact that their good partially *constitutes* our own good. The proposals for cases 1-3 above are especially susceptible to an analysis like this; perhaps we care about the harms, benefits, and accomplishments of our loved ones only because those harms, benefits, and accomplishments contribute to our own overall good as constitutive parts.

Although I do not wish to deny that there are cases in which our reactive responses to situations involving our close ties reduce to self-concern or self-appraisal, I do not think it is plausible to argue that *all* cases of special concern for our close ties are susceptible to this kind of reduction. First, notice that in either version of the reductivist proposal (i.e., the instrumental or the constitutive one), the strength of your reactive responses should be proportional to the good or bad done to *you*. That is, if the reductivist objection is correct, then in the case involving your friend's cheating partner, the strength of your resentment should be proportional to the harm done to *you*, given that (according to the reductivist's proposal) the harm done to *you* is what justifies your resentment. But this is an implausible result. Instead, the strength of your resentment should track the seriousness of the wrong done to your friend;

since the wrong done to your friend is serious, it is appropriate for you to have strong resentment for your friend's cheating partner. This is true even though it is not plausible to suggest the harm done to you was very serious.[36]

Connected to this is a second point, which is that both the instrumentalist and the constitutivist reductivist objections suggest an implausible story about the *focus*, or *object*, of our concern in cases involving our close ties. Consider first the instrumentalist alternative. Although I want to allow that we *sometimes* regard our close ties as sources of instrumental value for ourselves, the norms governing good close relationships are incompatible with our having only instrumental concern for our close ties. Insofar as we are good friends, good romantic partners, and good family members, our relationships call on us to care about the well-being and appraisal of our close ties, notwithstanding the instrumental utility or disutility it has for us. If this is correct, then it is implausible to suggest that when we respond reactively to circumstances involving our loved ones, our responses are justified solely by the instrumental importance of situations involving our loved ones for *us*. Instead, we should regard the well-being of our close ties not only as instrumentally good but also as intrinsically good, and the story we tell about the justification of our reactive attitudes should reflect this.[37]

A similar point can be made concerning a constitutive parts version of the reductivist objection. Although a constitutive parts view can accommodate the intuition that we should regard our close ties as intrinsically rather than as merely instrumentally valuable, it too struggles to tell a plausible story about how we should conceive of the special concern we have for our close ties in personal relationships (and about the focus, or object, of our reactive responses in situations involving our close ties). To see why this is so, consider the gratitude you feel toward your partner's coworker in case 2. Although we can certainly imagine circumstances in which this gratitude is an expression of the concern you have for your partner's good *qua* constitutive part of your

36 Note that it is not just the instrumentalist version of the objection that cannot tell a compelling story about the appropriateness of the strength of our reactive responses in cases involving our close ties. Even if the good of a close tie is a constitutive part of my own good, it still is not plausible to suggest that a very serious wrong done to friend, which has a significant impact on his overall good, has an equally significant impact on my overall good. And yet it seems that the strength of my reactive responses should be sensitive to the impact on the friend's overall good rather than to the impact on my overall good.

37 This also helps to explain why we do not usually think of *ourselves* as needing apologies or recompense when our loved ones are wronged. Because our resentment is not justified by a wrong done to *us*, an apology to us has no bearing on our decision to forswear resentment or refuse to forswear resentment for wrongs done to our loved ones. For further discussion of how this affects debates about the nature of forgiveness, see Chaplin, "Taking It Personally."

good, your role as a good relationship partner calls on you to feel gratitude toward your partner's benefactor independently of the import of the situation for your own good. Indeed, an especially loving partner might even feel gratitude in a case where the overall upshot for her is bad (a circumstance which is certainly possible).[38] Similarly, in the cases involving pride and shame, even if the good of an agent's close tie turns out to be partially constitutive of the agent's own good, the norms of close relationships suggest that the agent's reactive responses should be able to get a hold notwithstanding that agent's concern for her own appraisal. My pride in my brother should be compatible with the possibility that his accomplishments have no impact on anyone else's evaluation of me, and likewise, my shame in my friend should be compatible with my thinking that her behavior reflects only on her own character and not at all on mine.[39] In short, even if we admit that our reactive responses to our loved ones' treatment and behavior sometimes stem from self-concern and an interest in our own appraisal (whether in an instrumental or constitutive guise), they need not always do so, and the norms of our relationships suggest that they should not always do so. Indeed, in many paradigmatic cases involving our close ties, it is implausible to suggest that reactive attitudes apparently manifesting our concern for others in fact manifest self-concern. In light of this, we should conclude that an objection appealing to reductivism about partial concern for others fails.[40]

38 It is not difficult to imagine cases of this sort. Perhaps her recognition leads to a promotion that means her partner must take on many more of the household duties and chores, making her partner's daily life much less enjoyable.

39 An especially radical version of the reductivist's objection might be based on the view that *all* reasons to be altruistic are grounded in egoistic reasons (e.g., see Brink, "Self-Love and Altruism" and "Impartiality and Associative Duties"). In a view like this, when my special concern for my friend motivates me to resent his wrongdoer, my special concern for my friend is intelligible only in light of the fact that promoting his good treatment contributes to my overall good. But notice that in a view like this, we can never have a reason to promote the well-being of our loved ones at the expense of our own overall good, even in principle. For readers that take this to be an implausible result, this constitutes a further reason to reject a radical version of an egoistic reduction such as the one just described.

40 Another kind of objection, which I do not take up at length here, suggests that attitudes like resentment, gratitude, pride, and shame need not even express *partial* concern. Consider, for instance, the possibility of feeling gratitude to a great philanthropist for all they have done to fight disease. Or consider the possibility of feeling ashamed by what human beings have done to the planet. One might object that these seem to be cases in which attitudes like resentment and shame express other-regarding and yet *impartial* concern (since they apparently rest on concern for humanity as a whole). However, I think cases like these are most plausibly interpreted in a way that confirms my main claim that attitudes like gratitude and shame are essentially partial. For when I feel shame for what humanity as a whole has done to the planet, my shame is fitting only insofar as *I* am a member of the

Finally, note that the claims just made are fully compatible with the idea that in close relationships there is a sense in which we take on the good or flourishing of our relationship partners *as our own ends*. For to say that I consider it one of *my* aims for my partner to flourish is not necessarily to say that I think of my partner's flourishing as identical to my own, or even as a constitutive part of my own flourishing. Rather, to say that I take my partner's flourishing as my end may just be to say that I have made my partner's ends my special concern (viz., into something I am especially committed to promoting for its own sake). That is, my concern for my partner's flourishing is *partial*, just like my interest in my own good is my special concern, but this need not involve my ceasing to distinguish between my good and my appraisal on the one hand, and my partner's good and my partner's appraisal on the other. Thus, even if we allow that the reasons we have to care about our close ties are in some respects similar to the reasons we have to care about ourselves (insofar as they are sources of partial concern), it does not follow that special concern for our close ties reduces to a kind of self-concern. Rather, we can grant that caring for our close ties involves "making their ends our own" (on some understanding of this locution) and hold that this is simply to be understood as a gloss on what it is to have partial concern for someone else.[41]

4. Final Reflections

My discussion above has aimed to show that reactive attitudes such as resentment, gratitude, pride, and shame are best understood as attitudes of partial concern (either for ourselves or for others), and so they should not be understood as belonging to classes of attitudes defined by the notions of *self* and *other*. In line with this, I have also aimed to show that the appropriateness of these attitudes in cases involving our close ties cannot be explained away by an appeal to the way in which our own well-being and appraisal depend on the well-being and appraisal of others. As I have argued, although it is true that our

group who has damaged the planet. That is to say, my shame is in fact partial, for if *I* were not a member of the group whose behavior is shameful, I would not feel shame. Similarly, if my attitude toward the philanthropist does not express any form of partial concern at all, then I think we should say that gratitude is not fitting, properly speaking (though attitudes like approval and admiration would be).

41 Though I do not wish to endorse his account of love in particular, Helm's discussion of "person-focused emotions" is helpful in articulating how we may want to understand the partiality involved in partial reactive attitudes. According to Helm, the person-focused emotions that play a central role in love evince a kind of concern that is the same as the kind of concern we have for ourselves, but self-concern is *not* conceptually prior (Helm, "Love, Identification, and the Emotions," 42).

close relationships often involve taking on our close ties' interests as our own, it is nonetheless implausible to suggest that self-concern is the sole justificatory or explanatory basis of our partial responses concerning our loved ones.

In closing, I now want to make one final set of remarks about Strawson's claim that only the "impersonal" attitudes qualify as "moral." Recall that Strawson says the "impersonal" reactive attitudes uniquely deserve the label "moral" because they are "disinterested" and demand "a reasonable degree of goodwill or regard on the part of others, not simply towards oneself, but towards all those on whose behalf moral indignation may be felt."[42] I take this to be an expression of Strawson's commitment to the view that morality demands impartiality. As Strawson sees it, although partial attitudes such as resentment and gratitude should be our starting point for theorizing about *moral responsibility* (since they give us our first grip on what it is to regard one another as morally responsible), there is a different sense in which we use the label "moral" to mark out only the *impartial* demands that we make on (and on behalf of) all people. Indeed, for Strawson, the claim that the "impersonal" attitudes are uniquely *moral* is fully compatible with the claim that our theorizing about moral *responsibility* should start with observations about, as Strawson writes, "what it is actually like to be involved in ordinary personal relationships," where our commitment to impartiality may not always be manifest.[43]

But this observation leads to a second point deserving of emphasis. Although in this paper I have been arguing that a bipartite, partiality-based taxonomy of reactive attitudes captures Strawson's fundamental concerns in "Freedom and Resentment," there is one respect in which I may be advocating a departure from Strawson; namely, in arguing that our close personal relationships sometimes require us to respond with the partial reactive attitudes, I have suggested that our moral obligations and concerns are not thoroughly impartial. As I see it, attitudes such as resentment, gratitude, pride, and shame help us fulfill the obligations of our close relationships by supporting us in our efforts to care for and attend to one another. Indeed, I have even argued that it would be problematic for an agent's reactive attitudes to register no difference between, say, the wrong done to her partner and the wrong done to a perfect stranger.[44] While I have not argued for the claim that the obligations of our

42 Strawson, "Freedom and Resentment," 121–22.

43 Strawson, "Freedom and Resentment," 113.

44 Perhaps we do not have obligations to have *particular* attitudes in particular instances (since we cannot directly control our attitudes), but if we frequently fail to show any kind of partial concern *at all* for our close ties, something is morally amiss. Plausibly, failure to have any partial attitudes whatsoever within the context of close relationships indicates a failure of uptake with respect to a person's significance to you.

close relationships are *moral*, I take it to be a plausible one, and this aspect of my view may be at odds with Strawson's suggestion that moral demands are fundamentally impartial.[45]

However, even if I have departed from Strawson in this way, I have not departed from his core understanding of the *nature* of the reactive attitudes, for, as I have argued, Strawson's discussion does suggest that the core distinction our framework of reactive attitude types should capture is the distinction between attitudes that express partial and impartial concern.[46] Moreover, allowing for a role for partiality in morality may in fact give some readers further reason to embrace the entire account developed in this paper. For instance, proponents of relationship-based obligations should be especially eager to embrace a partiality-based taxonomy of basic reactive attitude types, for if relationships of love, family, and friendship generate special moral obligations and entitlements to prioritize caring for some people over others (as proponents of relationship-based obligations hold), then our reactive attitudes ought to register this fact.[47] That is, while we should expect to see some other-concerning responses

45 However, Strawson's "Social Morality and Individual Ideal" is friendly toward the idea of some role for partiality in morality. There, Strawson argues that morality need not be a system of universal principles, but it does need to be a system whose participants recognize at least some reciprocal claims. Strawson writes: "What is universally demanded of the members of a moral community is something like the abstract virtue of justice: a man should not insist on a particular claim while refusing to acknowledge any reciprocal claim" (11). Elsewhere in the article, Strawson indicates a willingness to embrace the notion of an "internal morality of an intimate personal relationship" (7), and, more generally, he is content with role-based obligations and the partial requirements they sometimes make on us. So this paper's claim about relationship-based obligations and the supporting role of partial reactive attitudes may not be at odds with Strawson's considered understanding of morality after all.

46 Is my proposal also compatible with the Strawsonian idea that some attitudes are "generalizations" of others? Are the impartial attitudes "generalized analogues" of the partial ones? While spelling out the precise relationship between partial and impartial attitudes requires a separate paper, I think impartial attitudes may be articulable as generalizations of partial ones and that this may provide fruitful material for developing a broader theory of partial and impartial concern. I thank John Fan for suggesting this to me.

47 For some arguments for relationship-based obligations, see Bazargan-Forward, "The Identity-Enactment Account of Associative Duties"; Brink, "Impartiality and Associative Duties"; Darwall, "Responsibility within Relations"; Hardimon, "Role Obligations"; Kolodny, "Which Relationships Justify Partiality?"; Scheffler, "Morality and Reasonable Partiality"; Stroud, "Permissible Partiality, Projects, and Plural Agency"; Svirsky, "Responsibility and the Problem of So-Called Marginal Agents"; Wallace, "Duties of Love"; and Wolf, "Morality and Partiality." (Wallace, however, explicitly denies that relationship-based obligations count as *moral*, so I assume he would resist at least this aspect of my proposal. Similarly, Wolf argues that some relationship-based obligations are founded in impartial moral demands, but other considerations of love and friendship

that manifest impartiality (responses involving attitudes like indignation, disapprobation, approval), we should also expect to see other-concerning responses that reflect our special investment in the well-being and behavior of our close ties (responses involving attitudes like resentment, gratitude, pride, and shame).[48] Indeed, given that the reactive attitudes can play a variety of communicative, defensive, and even sanctioning roles, and given how much we care about what reactive attitudes others have concerning us, it is no surprise that the partial reactive attitudes play an especially important role in the obligations and entitlements associated with our close personal relationships.[49]

All this said, let me end by stressing that Strawson's claim that the impersonal attitudes are uniquely "moral" does get at an important point—namely, that there is an important moral difference between attitudes that reflect our partial concern for ourselves and others and attitudes that reflect our impartial concern for ourselves and others. Both kinds of attitudes play important roles in our moral lives, but they express our different commitments to general principles of impartiality, on the one hand, and principles that allow for (and sometimes demand) differential concern for ourselves and our close ties, on the other. As I have argued, the attitudes Strawson calls "personal" and "self-reactive" play especially important roles in the aspects of our moral lives that allow for, and sometimes demand, *partiality*. And although these attitudes *can* express the special concern we have for ourselves, they can also express the special concern we have for our loved ones. When we are alive to this feature of the so-called personal and self-reactive attitudes, we see that more fundamental than the distinction between self- and other-concerning attitudes is the

---

give us nonmoral reasons to be partial, which can compete with morality.) Broad's "Self and Others" also makes a compelling case that relationship-based obligations are a part of commonsense morality (in the form of what he calls "self-referential altruism"), but he does not explicitly commit to the thesis that self-referential altruism is true.

48 In some relationships, partial concern may even be *constitutive* (or at least partially constitutive) of the relationships. For instance, perhaps I cannot *be* a friend to someone unless I have differential concern for her welfare, her projects, her character, and so on. Whiting's "Friends and Future Selves" makes this especially plausible (though there are egoistic elements in Whiting's account of concern for friends that I do not wish to endorse). Additionally, though she does not focus on partiality, Svirsky shows how regarding norms and expectations as partly constitutive of close relationships can help to explain the responsibility of so-called marginal agents.

49 Consider, for example, how my resenting my friend's wrongdoer might help me defend him and protect his interests by motivating me to hold his wrongdoer to account. Or consider how my pride in my brother might signal my support for him and my commitment to helping him achieve his important personal ends. We can tell similar stories for other attitudes (like shame and gratitude) insofar as they have motivational and social significance for their targets.

distinction between partial and impartial ones. As I have argued, a bipartite characterization of Strawson's fundamental taxonomy of reactive attitudes best captures this important point.[50]

University of North Carolina at Chapel Hill
rchaplin@unc.edu

REFERENCES

Bazargan-Forward, Saba. "The Identity-Enactment Account of Associative Duties." *Philosophical Studies* 176, no. 9 (September 2019): 2351–70.

Brink, David O. "Impartiality and Associative Duties." *Utilitas* 13, no. 2 (July 2001): 152–72.

———. "Self-Love and Altruism." *Social Philosophy and Policy* 14, no. 1 (January 1997): 122–57.

Broad, C. D. "Self and Others." In *Broad's Critical Essays in Moral Philosophy*, edited by David Cheney, 262–82. Abingdon: Routledge, 1971.

Chaplin, Rosalind. "Taking It Personally: Third-Party Forgiveness, Close Relationships, and the Standing to Forgive." In *Oxford Studies in Normative Ethics*, vol. 9, edited by Mark Timmons, 73–94. Oxford: Oxford University Press, 2019.

Clarke, Randolph, Michael McKenna, and Angela M. Smith, eds. *The Nature of Moral Responsibility: New Essays*. Oxford: Oxford University Press, 2015.

Darwall, Stephen. "Responsibility within Relations." In Feltham and Cottingham, *Partiality and Impartiality*, 150–68.

Feltham, Brian, and John Cottingham, eds. *Partiality and Impartiality: Morality, Special Relationships, and the Wider World*. Oxford: Oxford University Press, 2010.

Foot, Philippa. *Virtues and Vices: and Other Essays in Moral Philosophy*, Oxford: Oxford University Press, 2002.

Hardimon, Michael. "Role Obligations." *Journal of Philosophy* 91, no. 7 (July 1994): 333–63.

Helm, Bennett W. *Communities of Respect: Grounding Responsibility, Authority, and Dignity*. Oxford: Oxford University Press, 2017.

———. "Love, Identification, and the Emotions." *American Philosophical*

50 For invaluable feedback on previous drafts of this paper, I am grateful to Lucy Allais, Jackson Bittick, David Brink, Cory Davia, John Fan, Dana Nelkin, Robert Wallace, Shawn Wang, Monique Wonderly, audience members at the 2020 Central APA, and an anonymous referee for *JESP*.

Quarterly 46, no. 1 (January 2009): 39–59.
Hieronymi, Pamela. "Articulating an Uncompromising Forgiveness." *Philosophy and Phenomenological Research* 62, no. 3 (May 2001): 529–55.
———. *Freedom, Resentment, and the Metaphysics of Morals*. Princeton: Princeton University Press, 2020.
Kolodny, Niko. "Which Relationships Justify Partiality? The Case of Parents and Children." *Philosophy and Public Affairs* 38, no. 1 (Winter 2010): 37–75.
Macnamara, Coleen. "Reactive Attitudes as Communicative Entities." *Philosophy and Phenomenological Research* 90, no. 3 (May 2015): 546–69.
McKenna, Michael. *Conversation and Responsibility*. Oxford: Oxford University Press, 2012.
Murphy, Jeffrie G., and Jean Hampton. *Forgiveness and Mercy*. Cambridge: Cambridge University Press, 1988.
Scheffler, Samuel. "Morality and Reasonable Partiality." In Feltham and Cottingham, *Partiality and Impartiality*, 98–130.
Smith, Angela M. "Responsibility as Answerability." *Inquiry: An Interdisciplinary Journal of Philosophy* 58, no. 2 (January 2015): 99–126.
Strawson, P. F. "Freedom and Resentment." In Hieronymi, *Freedom, Resentment, and the Metaphysics of Morals*, 107–34.
———. "P. F. Strawson Replies." In *Philosophical Subjects: Essays Presented to P. F. Strawson*, edited by Zak van Straaten, 260–96. Oxford: Clarendon Press, 1980.
———. "Social Morality and Individual Ideal." *Philosophy* 36, no. 136 (January 1961): 1–17.
Stroud, Sarah. "Permissible Partiality, Projects, and Plural Agency." In Feltham and Cottingham, *Partiality and Impartiality*, 131–49.
Svirsky, Larisa. "Responsibility and the Problem of So-Called Marginal Agents." *Journal of the American Philosophical Association* 6, no. 2 (Summer 2020): 246–63.
Walker, Margaret Urban. "Third Parties and the Social Scaffolding of Forgiveness." *Journal of Religious Ethics* 41, no. 3 (September 2013): 495–512.
Wallace, R. Jay. "Duties of Love." *Aristotelian Society Supplementary Volume* 86, no. 1 (June 2012): 175–98.
———. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press, 1994.
Watson, Gary. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." In *Agency and Answerability: Selected Essays*, 219–59. Oxford: Oxford University Press, 2004.
Whiting, Jennifer. "Friends and Future Selves." *Philosophical Review* 95, no. 4 (October 1986): 547–80.

Wolf, Susan. "Morality and Partiality." *Philosophical Perspectives* 6 (1992): 243–59.

Zaragoza, Kevin. "Forgiveness and Standing." *Philosophy and Phenomenological Research* 84, no. 3 (May 2012): 604–21.

Journal of Ethics and Social Philosophy
VOL. 25, NO. 2 · AUGUST 2023
https://doi.org/10.26556/jesp.v25i2.1783

SEPARATING THE WRONG OF SETTLEMENT FROM THE RIGHT TO EXCLUDE

TERRITORY AND SOCIOCULTURAL STABILITY

Daniel Guillery

An important part of the history of modern colonialism has been a history of *settlement*. One major form that colonial subjugation has taken has been that of settler colonialism, in which a group of settlers moves and together establishes a home in a new, already-inhabited geographical location, aiming in some sense to replace its existing inhabitants and create an outpost of the society from which they came.[1] (The settlers aimed to *replace* existing inhabitants often in a literal sense, through extermination or enforced displacement, though not always; sometimes the project of replacement was rather one of shaping the territory according to the *settlers'* practices, goals, and ideals.) The colonization of the Americas and the South Pacific are core instances, but other histories of imperialism have exhibited the traits of settler colonialism to varying degrees. Needless to say, the settler colonial record is morally hideous. It involved widespread murder, rape, exploitation, enslavement, forced displacement, political subjugation, and cultural imposition and domination. But at the core of *settler* colonialism is the act of settlement, permanent relocation to a new geographical home, which might on its face seem a morally innocuous one. We might wonder, then, whether what marks settler colonialism out as a distinct form of imperial relationship (namely, settlement) is, from a *moral* point of view, merely an incidental feature of a project that was wrong for other reasons. Or is its distinguishing element a morally significant one: wrongful settlement?

Settler colonialism is a complex historical phenomenon that emerged at a particular time and place (or places). Its various manifestations are characterized by a bundle of motivations, ideas, and practices, grouped, most plausibly,

1 Note that the term "colonialism" is sometimes used to refer *exclusively* to a phenomenon of this sort, centrally involving settlement, distinguished from "imperialism," the exercise of power of some sort by one state, nation, or people over another (see Kohn and Reddy, "Colonialism"; and Moore, "Justice and Colonialism," 447–48). On settler colonialism, see for instance Bell, *Reordering the World*, ch. 2; and Veracini, "Introducing."

by a family resemblance. Various elements of this bundle are straightforwardly wrongful, often egregiously so: from racist devaluation or dehumanization of indigenous people and concerted campaigns of extermination to forced assimilation and cultural imposition. The historical phenomenon is distinguished by the particular way in which these elements came together. But the question that concerns me here is whether settler colonialism is a distinctive *normative* phenomenon, as well as a historical one. Not all of the distinguishing historical features are normative ones, and many of the wrongs involved do not, on their own, distinguish settler colonialism from other forms of imperialism. The natural place to look for one that does is the act of settlement itself. My question, then, is whether *settlement* (of a certain kind) can itself be wrongful. There is some intuitive temptation to think so. As Margaret Moore has recently argued, it is natural to resist fully assimilating the moral story we tell about settler colonialism to a wider group of colonial or imperial projects.[2] Additionally, as she points out, members of groups subjected to settler colonialism often describe the wrongs perpetrated against them or their ancestors as bound up with *territory* or *land*, and complaints against settlement do seem capable of persisting even when political subjection and straightforward violence are not so prominent.[3] Consider, for instance, Haunani-Kay Trask's claim that "in less than a hundred years after Cook's arrival [in Hawai'i in 1778], my people had been dispossessed of our religion, our moral order, our form of chiefly government, many of our cultural practices, and our lands and waters."[4] This describes a century that saw significant influxes of visitors (traders, missionaries, and so on) and settlers (who by 1890 made up 55 percent of the population), but no formal political subjection (which followed in 1898) and relatively little direct violence.[5] Of course, the Hawaiian case, like all historical cases, is complex and messy; it does not pinpoint the question exactly, and independent wrongs were certainly committed. But at the core of what happened during the period described by Trask with a sense of moral outrage were settlement, trade, and evangelism.

Yet if we are attracted by the sort of cosmopolitan view that rejects the idea of exclusionary rights over territories, we might seem to be led straightforwardly to deny the possibility of settlement that is itself morally wrongful.[6] If "settlement" is just another term for immigration (perhaps in a context of imperial

2 Moore, "The Taking of Territory," 88, and "Justice and Colonialism," 448, 455.

3 Moore, "The Taking of Territory," 88.

4 Trask, *From a Native Daughter*, 10.

5 Kauanui, *Paradoxes of Hawaiian Sovereignty*, 87.

6 Advocacy of open borders is gaining wider currency; see, for instance, Carens, "Aliens and Citizens," and *The Ethics of Immigration*; Cole, *Philosophies of Exclusion*; Oberman, "Immigration as a Human Right"; Huemer, "Is There a Right to Immigrate?"

domination of some kind), then it might seem natural to conclude that we cannot both see settler colonialism as distinguished by a *wrong* and deny the existence of rights to exclude from territory. If that is correct, we can either reject this sort of cosmopolitan position or deny the distinctive wrongfulness of settler colonialism.[7] The intuitive cost to the latter option leads Moore to take the first horn of the apparent dilemma and to posit exclusionary territorial rights. It is my aim in this paper, however, to dismantle the dilemma. We can accept that settlement can be wrong, and so settler colonialism is not distinguished merely by the morally incidental form that imperial subjugation happened to take, but without granting exclusionary rights over territory to anybody. Or so I will argue.

We should be clear, though, that it is an option very much open to the cosmopolitan denier of exclusionary territorial rights simply to deny that there *is* anything distinctively wrong with settler colonialism. This would not force any obviously wrong judgments about historical cases: their wrongfulness can easily be located elsewhere. There is some disagreement in recent philosophical discussion about what, if any, is the essential or distinctive wrong of *colonialism* more generally (understood broadly to encompass settler colonialism as well as a variety of other imperial relations).[8] While some take the essential wrong of colonialism to be the violation of exclusive property-like territorial rights, others take it to be, or involve, political subjugation or domination of a certain kind.[9] It could be, then, that what was wrong with historical instances of settler colonialism was (most centrally) just what it shared with other forms of colonialism, and if we take the view that this was some form of political domination, no territorial rights are needed. Others hold that there is *no* essential wrong of colonialism, and what made historical instances of it grievously wrong was just the litany of other wrongs with which it was contingently connected.[10] Settler colonialism has been accompanied by a diverse bundle of such wrongs, including the deceit, force, and violence through which it was achieved, and

7 The phrase "distinctive wrongfulness of settler colonialism" could be read in two ways: here, I do not mean it to imply that there is a unique wrong associated with settler colonialism, but rather that the distinguishing feature of settler colonialism (the *settlement*) is a wrong.

8 These are two different questions, though the existing literature does not seem always to notice this. It may be that certain essential or necessary features of colonialism are wrong, though not distinctively so: the wrong-making features might not be *sufficient* for something to count as colonialism, and so be shared with other phenomena.

9 On the former, see Ferguson and Veneziani, "Territorial Rights and Colonial Wrongs." On the latter, see Ypi, "What's Wrong with Colonialism?"; and Stilz, "Decolonisation and Self-Determination." On the debate generally, see Moore, "Justice and Colonialism."

10 Valentini, "On the Distinctive Procedural Wrong of Colonialism." Cf. van Wietmarschen, "The Colonized and the Wrong of Colonialism."

murder, rape, exploitation, and enslavement that went alongside. So, we will have no difficulty finding wrongful actions in the history of settler colonialism without turning to the act of settlement. But, as I have suggested, such a story will not satisfy everybody and might seem to miss something. It is at least worth considering, then, whether complaints against *settlement* can be taken seriously.

My aim in this paper, then, is to offer an alternative account of a possible wrong of settlement that does *not* require us to posit any exclusionary rights over territory. The wrong I describe is certainly not the whole story about historical (and current) cases of settler colonialism. A recurring feature of these is the prevalence of various forms of disrespectful treatment of colonized people: the devaluation of their practices, beliefs, and identities, social marginalization, discrimination, and so on. These wrongs (as well as the others mentioned above) will form an important part of the moral story about the history of settler colonialism. Nevertheless, I do think the account I will provide below gives another crucial *part* of that story.

The account I will put forward posits an interest in sociocultural stability, in constancy of the background social conditions on the basis of which we orient ourselves in the world, and which shape and frame the options available to us. Our well-being and agency, I will suggest, depend on some degree of sociocultural stability of this kind. In some cases, these background cultural practices can involve patterns of land use in particular geographical areas that would be disrupted by certain patterns of settlement by new inhabitants. The interest I describe will only ground a weak, *pro tanto* right, but it is of sufficient importance, I think, that in particular circumstances it would be wrong to settle in an area in which others already have interests of this kind (if you have no correspondingly strong interest in using that particular area of land). Importantly, the sociocultural stability rights I posit are grounded in an interest in what I will call "orientation," not in the sort of interests in making and pursuing plans that Moore (and Anna Stilz) appeal to, and for this reason they are rights to *stability*, not *control*. Thus, they allow us to account for a possible wrong of settlement, but not because the existing inhabitants have any sort of *exclusionary rights* over the territory. The wrong I will describe is not an essential or necessary feature of settlement, nor is it a wrong that can only be committed through settlement, but I will claim that the act of settlement itself *can* (and sometimes does) constitute a wrong of this kind (and so settler colonialism can be understood in moralized terms, as distinguished by *wrongful* settlement).[11]

11 Since the wrong I identify is not essential to settlement, the normative category of wrongful settlement I identify may not map perfectly onto the historical category of settler colonialism.

The paper thus has two main aims: first, to defend skepticism about exclusionary territorial rights from the concern that it prevents us from accounting for a distinctive wrong involved in the historical phenomenon of settler colonialism (as there is some intuitive temptation to do), and second, to identify a significant feature of the moral universe, relevant not only to evaluation of the past, but also potentially to action and policy here and now. I begin by describing the plan-based accounts of occupancy rights given by Moore and Stilz, and distinguishing their function from that of the account I will give. I then set out my account of the interest in sociocultural stability and the rights grounded in it. Next, I describe how these rights can be violated, most obviously by physical displacement, but also by settlement. Finally, I argue that they do not support exclusionary rights, or property-like territorial rights of any kind.

1. BACKGROUND

It is reasonably straightforward to account for a distinctive wrong involved in colonial settlement if we attribute exclusionary, property-like rights over territory to groups of inhabitants. Uninvited settlement of the territory then becomes a simple violation of its inhabitants' collective right to a certain range of control over the territory, or to determine for themselves the conditions of access to it (an asymmetrical right that outsiders lack). Both Margaret Moore and Anna Stilz, two leading theorists of territorial rights who have offered explanations of the wrong involved in settlement, pursue this route.[12] The accounts they offer differ substantially, and they differ notably in the extent of the justification for exclusion that they are willing to grant to territorial-right holders. Nevertheless, both arrive at territorial rights that are exclusionary in the sense important for my purposes. These are rights to a substantial degree of *control* over the territory in question, rights such that it would make sense to say of their object that it, in some restricted sense, *belongs* to the right holder; they have a certain kind of asymmetrical *authority* over it; it is in a sense *theirs*.

In neither case are the control rights envisaged absolute or unlimited; they are rights to control the territory in certain respects only and within certain limits. Both also acknowledge that in some cases exercising control to prevent access to a territory would be unjust even where the controlling agent possesses genuine legitimate authority over the matter. Moore countenances a fairly extensive justification for discretionary exclusion from legitimately held territories.[13]

12 Moore, "The Taking of Territory"; and Stilz, "Settlement, Expulsion, and Return." Also Moore, *A Political Theory of Territory*; and Stilz, *Territorial Sovereignty*.

13 Moore, *A Political Theory of Territory*, ch. 9.

Stilz is much more restrictive, arguing that relatively stringent conditions have to be met for exclusion to be justified.[14] Still, though, as for Moore, the territorial rights that Stilz defends *are* exclusionary and involve a *right to exclude* in the sense I have in mind. To explain this, let me distinguish two kinds of question in political philosophy. First, we may ask what justice requires, or what a justifiable policy would look like. This is the sort of question we ask when deciding, for instance, what policy to vote for. But second, we may also ask what procedures, or which people, have the legitimate authority to make a particular decision, and to impose it on others. Stilz carefully distinguishes these two questions. The account she gives of the justifiability of exclusion (and the limits to discretion she imposes) is an answer to the first question; it is an account of the substance of a just immigration policy, not its legitimacy.[15] Although an immigration policy that excluded harmless immigrants would be unjust, Stilz thinks that a self-determining people (with the kind of occupancy rights and jurisdictional rights she defends) has the *right*, or *legitimate authority*, to set its own immigration policy. If such a people were to make the wrong decision, outsiders would still be obliged to respect it (at least up to a certain point). For the purposes of this paper, I want to reserve the terms "right to exclude" and "exclusionary territorial rights" for an answer to the *legitimacy* question: for a state or people to have the right to exclude in this sense is for it to have legitimate authority over the matter of exclusion from a particular territory. If a state is, or would be, *justified* in excluding, I will say that it possesses an "exclusion justification." Thus, in these terms, Stilz holds that legitimate states *do* have the right to exclude, though they have only quite a limited (and certainly not a *discretionary*) exclusion justification.

What, then, is the basis for the kind of exclusionary control right that writers like Moore and Stilz posit? Moral considerations called "occupancy rights" play an essential role in both Moore's and Stilz's accounts. These are quite limited, "primitive" rights (in Stilz's phrase) over space or land that do not depend on the existence of an entity capable of governing a territory, but serve as stepping stones in justifying the full-blown "territorial rights" that both defend.[16] The occupancy rights that the two writers defend are quite different (notably, for Stilz they are held by individuals, while for Moore they are group rights, though she also posits partially derivative individual "residency rights"), but for both these are property-like rights (in the sense that they are rights to a certain extent of *control* over a space or object, only in this case the extent of control

14 Stilz, *Territorial Sovereignty*, ch. 7, and "Settlement, Expulsion and Return," 363.

15 Stilz, *Territorial Sovereignty*, 188.

16 For the phrase "primitive rights," see Stilz, "Property Rights" and *Territorial Sovereignty*, ch. 3.

is somewhat less than that involved in a full liberal property right).[17] In both cases, these rights are also pre-institutional in that they do not depend on any established institution (or artificial human convention) that grants occupancy rights in a particular place to particular individuals or groups. It is natural facts about people's (individual or collective) connections to places that give rise to these rights and their correlative obligations.

For both writers, these occupancy rights play a necessary and central role in justifying the kind of control rights they think territorial-right holders have, and, as a result of this, in explaining the wrong of settler colonialism.[18] For both, they seem to be, in Stilz's terminology, the "foundational title" on which territorial rights are built (and that attaches states or peoples to particular spaces and provides the necessary link between the valuable functions served by territorial control and rights over a particular space). (Although, unlike Stilz, Moore does not seem to take occupancy rights to be sufficient on their own to account for the wrong of settlement, the territorial rights that allow her to do so depend necessarily on the former for their justification.)

Interests in some sort of collective self-determination play an equally important role in both accounts of exclusionary territorial right. But group self-determination rights, as both writers seem to acknowledge, are not, and do not on their own include, rights to control any particular physical objects or spaces in the external world (and it is for this reason that occupancy rights are needed). To see this basic point, it is sufficient to notice that there are groups that seem plausibly to have a right to be collectively self-determining to a substantial degree, but where the right to self-determination has no territorial (or external-object-involving) dimension at all. Consider a voluntary association like a book club, for instance, or a religious community. We might plausibly think it matters that some such groups be free to determine to a reasonable extent their own internal affairs according to their shared goals, preferences, or ideals. Yet achieving this clearly does not require book clubs, rugby teams, churches, or mosques to have control over an area of land. Such an association may, of course, own property, and perhaps their right to self-determination entitles them to make their own collective decisions about how to use property they legitimately own under an existing legitimate property regime, but it does not seem necessary in order for a book club, say, to be self-determining in the relevant respects that it owns property. A right to self-determination (individual or collective) cannot be a right to do whatever one chooses, and so there

17 Stilz, *Territorial Sovereignty*, 33–36.; Moore, *A Political Theory of Territory*, 43–45.

18 Stilz, "Settlement, Expulsion, and Return" and *Territorial Sovereignty*, 26–27; Moore, *A Political Theory of Territory*, 36–37, and "The Taking of Territory."

is no reason to *assume* that what a self-determining group must be free to do includes the exercise of control over land.

Now, of course, the kind of groups that are held to have territorial rights are importantly different from groups like religious associations and voluntary shared-interest associations. One might, then, combine the self-determination idea with the thought that groups of a particular kind (most likely, groups with some sort of shared *political* project, as well as the capacity to deliver the goods that such a project can provide) require control over an area of land in order to achieve the valuable function that self-determination for such a group can fulfill. Moore and Stilz do not take this route (though there are certainly elements of such a story in their accounts). To summarize very briefly, the problem is that this kind of story cannot explain what binds others to respect the unilateral claims over *particular* areas of land that a group happens to make (in the absence of any overarching institutions or conventions that could legitimate such claims).[19] It does not explain what would be wrong with an outsider group *B* turning up and using land *T* in which group *A* is currently exercising political self-determination if there are other, equally good places where *A* could perform the same valuable functions instead, or if *B* could equally well perform the same functions in *T*. (The *mere* fact of first arrival does not seem to be a morally significant one; at least, we need some explanation of what is morally significant about first arrival. I will return to this point below.)

For these reasons, Moore and Stilz need the "occupancy rights" they defend to connect the interest in collective self-determination with control over particular geographical areas.[20] The move is from limited "primitive" control rights over an area to more substantial territorial rights, supported by the ways in which these more substantial rights enable groups already holding basic control rights in a place to serve their interest in being collectively self-determining. In defending the foundational "occupancy rights" they need, both writers appeal in turn to interests in the ability to plan and bring plans to fruition or to pursue stable projects and commitments over time.[21] It is this appeal that permits the

19 For similar arguments, see Moore, *A Political Theory of Territory*, ch. 5; and Stilz, *Territorial Sovereignty*, ch. 4.

20 See Stilz, *Territorial Sovereignty*, 26–27.

21 For Stilz, this is quite straightforwardly explicit (*Territorial Sovereignty*, 11, 40ff.). For Moore, occupancy rights are group rights, making things more complicated, but these group rights seem nevertheless to be grounded in interests in developing *shared* plans and projects over time. The importance of collective *identities* (and the relationship of these to particular places) plays an important part in her justification of occupancy rights (*A Political Theory of Territory*, 40), but the structure of the argument seems to be that because of the importance of collective identities to group members, shared projects of groups that possess such an identity matter analogously to the way individual projects matter. (On this picture, I think,

move to *control* that, as we have seen, both writers want to make. It will not be possible, or so the thought goes, to make and pursue a stable plan over time, without some *control* over the necessary background conditions on which the plan relies. The interest we have in developing and pursuing plans over time can only adequately be served by making use of external objects in the world, and, in particular, physical space. The successful pursuit of plans depends on the ability to rely on continued access to (and ability to use in planned ways) elements or parts of the world involved in the plans you have made. Because of the importance of this human interest, the involvement of an object or space in a person's (or group's) plans can give rise to obligations in others to refrain from using it in ways that conflict with those plans. This entails an extent of moral *control* over the object or space on the part of the initial planner.

I think it is plausible that we have these planning interests and that they are morally significant. I think this may be part of the explanation of why it is a good thing to have a system of reliable property rules that allocates rights to access, use, and exercise control over external objects. I am skeptical, though, that these interests are sufficient to ground obligations to respect others' unilateral appropriations (whether as individuals or groups) independent of any legitimate human institutions allocating asymmetrical rights over particular things to individuals or groups. In other words, I am skeptical that they ground *natural* rights over things or places. There is not space here to give any sort of full argument against that idea. It is enough, though, for now, to point out that there are good reasons to be doubtful. The fact of an object's involvement in a person's plans is certainly a morally significant one, but, on its own, will not resolve any conflicts: a single object or space may often be involved in the plans of multiple people, and these

---

group projects matter in a way not reducible to the importance of their individual subprojects to individual group members, but the importance of the group projects is nevertheless derived from individual interests.) The argument, then, is analogous to Stilz's individualist one, only here the focus is on *collective* plans and projects as well as individual ones: it is still an interest in developing plans and projects over time that does the work. (Moore sometimes talks about the disruption of *identities* themselves by geographical displacement, but I do not think this should be taken literally. Displacement does not really disrupt an *identity*: a place can figure in a group identity without the group's being physically there. One may identify, for instance, as a member of a group displaced from territory *T*. What might be disrupted by displacement are plans, projects, or relationships whose importance is explained by their significance for a shared identity.) One other reading of Moore's argument here would see it as much closer to the argument I will make below (suggested by her talk of "attachment" and "feeling at home in the world" [*A Political Theory of Territory*, 43–44]). I will argue, though, that the interests I appeal to support rights to a certain kind of stability, but *not* rights to control. Nothing Moore says (*other* than the plan-based argument) justifies the move from interests in things like "attachment" or "feeling at home in the world" to *control* rights.

plans may conflict. The defender of natural, plan-based, property-like rights must distinguish *mere* involvement of an object in plans from actual incorporation in use. It is not very clear, though, how exactly this distinction is to be made, or what its moral significance is. We might attempt to make it in terms of some sort of physical contact. It is hard, though, to see what is so morally significant about physical contact (and why, say, someone who has had cursory physical contact with a plot of land has a stronger moral claim to it than someone who has made extensive plans concerning it at a distance). Is it simply first arrival that does the trick? Again, though, it is hard to see what is morally significant about first arrival. A plan that does not start out involving a particular object can later come to essentially depend on it. Why should the fact that I got to this object first and involved it in *a* plan of mine before you did have overriding moral significance, especially considering that it might have later come to be much more centrally involved in higher-level life projects of yours to which you are deeply committed? It is not clear, further, that there are any universally acceptable criteria for comparing depth of commitment or centrality of a plan, and it is not obvious that you will be morally bound to respect my deeper or more central plans in the absence of such criteria. If there is an established *convention* in place that grants rights according to a "first-come, first-served" rule (or some other rule), then, of course, things are different. But the fact that we would be better off with such a stable framework is a reason to establish property conventions of some sort, not a reason to respect the unilateral claims of others in the absence of these conventions.

None of this is conclusive, but it is worth bolstering this *prima facie* case with an appeal to authority: the view that there *are* natural, property-like rights is, I think, a minority one.[22] Occupancy rights, of course, are rights to a more limited range of control than typically argued for by defenders of natural property rights, but the reasons for doubt seem similar. It seems worth exploring, then, how far we can get *without* positing pre-institutional control rights over land or territory, on the part of either individuals or groups. What I will argue below is that we do not need such control rights in order to account for the distinctive wrong of settler colonialism. If we did, that would be one reason to posit their existence. But if what I say below is correct, it is possible to hold on to both skepticism about exclusionary territorial rights and the conviction that there is a distinctive wrong associated with settler colonialism, one that

22 For defenses of this minority viewpoint, see Simmons, *The Lockean Theory of Rights*, 271–77; Stilz, "Property Rights," 247–49; Sanders, "Projects and Property"; and van der Vossen, "Imposing Duties and Original Appropriation," 77–78. Moore herself argues against the idea of natural property rights (*A Political Theory of Territory*, 19–20). It seems to me, though, that her own view depends on an analogous anti-conventionalism about group *territorial* rights that faces similar problems.

has to do with the settlement itself. The right I will defend (violated, in some cases, by settlement) will not support the right to exclude—in the sense of the legitimate authority to make immigration policy—that Moore's and Stilz's occupancy rights are supposed to support.

2. AN INTEREST IN SOCIOCULTURAL STABILITY

The experience of disorientation and dislocation that tends to go along with sudden transplantation to a new and different environment and, especially, sociocultural environment, is probably familiar to many. When we lose the ability to understand what is going on around us, how things in our environment behave and interact, and how they will respond to our choices and actions, it can be debilitating and distressing. Simply finding oneself in a new topographical situation is perhaps the most banal source of disorientation: if you do not know the lay of the land, it is likely to be difficult to get anywhere useful. When we relocate to an unfamiliar *cultural* environment, in particular, it can become challenging to navigate the *social* world. We may become lost, both metaphorically and quite literally. We may struggle with things as mundane as getting around the physical urban environment, or finding things to eat, as we familiarize ourselves with the local practices for doing these things. We may find it difficult to understand the social significance of our actions and how we are perceived by others; we may miss subtle social cues or fail to grasp the options open to us and the expectations held of us. We might, for instance, unintentionally offend, or take offense from a well-meant gesture. We may find it more difficult to make social connections or develop relationships of trust, as we attempt to relearn the norms governing these. A shift of this kind may of course be exciting, for the possibilities it opens up, for the opportunity to learn new modes of social cooperation and new ways of understanding the world. But even where excitement predominates, it tends to go along with disorientation and confusion, which at their worst can be debilitating.

What I think is suggested by these observations is that there is a basic morally significant interest we can have in a certain kind of moderate environmental stability, with importance for our individual agency and well-being. This interest in moderate stability, I will suggest, is derived from a basic interest in what I will call "orientation." To be "oriented," in my sense, is to be able, literally and metaphorically, to find your way around your environment. Orientation is a form of understanding. To understand is, in some sense, to grasp something about the relations between elements of the world.[23] It is a cognitive relation:

23 See Grimm, "Understanding."

achieving understanding requires an *accurate* grasp of relevant features of the world. To achieve orientation is to grasp successfully, i.e., to understand those relations in the world the understanding of which enables a practical orientation to one's environment, to understand how things behave and how they can be located in a way that enables you to predict how the world will respond to your actions and how it can be used to achieve things. Of course, an individual's orientation in the world is a matter of degree, dependent on the extent and usefulness of their practically relevant understanding.

We achieve our understanding of the world, and our relation to it, in large part by making use of stable regularities. We navigate our local area with the aid of stable, familiar points of reference. Similarly, we navigate our lives, and the choices we face, using constant patterns that we observe in the world around us. These familiar regularities allow us to make sense of the various elements of the world that we experience, to predict the behavior of objects and agents we encounter, and to understand the possible ways that we can interact with them. It is clear, then, that a certain kind of stability plays an essential role in establishing this capacity for orientation. It is, quite obviously, not the case that the world needs to be perfectly static for us to be able to make sense of it, or to navigate it. The practically relevant understanding we are trying to achieve is, in large part, an understanding of how the world *changes*. But we make sense of change by reference to stable constants. We predict the future on the assumption that it will, in certain ways, resemble the past.

Of particular importance are the *social* regularities that structure our orientation in our environment. We are social beings, and for this reason, a large, and especially practically significant, part of the world we inhabit is socially constituted. Central, then, to the environmental stability that our orientation depends on is a degree of stability across the *social* patterns and regularities that surround us. We are typically surrounded by, participate in, contribute to, and can be constrained by a wide variety of established social practices. These are ongoing, mutually reinforcing patterns of behavior shaped by shared values, beliefs, structures of meaning, patterns of expectations, conventions, and so on.[24] It is a familiar point (made particularly by defenders of liberal multiculturalism) that the options open to us, the goals, projects, and relationships we can pursue, are culture dependent.[25] These options are both created and given meaning by existing cultural practices. My point is a related but more

24 On the nature of social practices, see Haslanger, "What Is a Social Practice?"; see also Kuper, *Culture*.

25 See, for instance, Dworkin, "Can a Liberal State Support Art?" 228–33; Kymlicka, *Liberalism, Community, and Culture*, ch. 8, and *Multicultural Citizenship*, 82–84; Margalit and Raz, "National Self-Determination," 448–49; and Raz, "Multiculturalism," 176. For discussion,

basic one that highlights not only the options, goals, and relationships available to us, but more generally the way we orient ourselves in the world.[26] The way we understand the world and our place in it is heavily culturally mediated. Established social practices account for a substantial portion of the regularities and fixed points that allow us to make sense of our environment. First, social practices have the special virtue of making possible social cooperation and coordination, and providing the framework within which it takes place.[27] Mutual intelligibility, and hence social interaction, depends on convergence on conventions, or salient regularities in behavior that establish stable expectations about the behavior of others.[28] Understanding these practices is thus essential to our understanding of, and ability to navigate, an especially important element of our world as social animals: our shared life and cooperation with others.

As well as the objects of our understanding being cultural, the social practices that surround us also condition our understanding of the physical world by providing us with the necessary conceptual tools. Existing practices of agriculture, to give one example, provide us both with bases for understanding social cooperation with others, as well as with particular ways of understanding land, its purpose, and our place in it, different to those available in pre-agricultural societies. Similarly, the ability to find one's way around an urban milieu depends on a background of social practices concerning things like roads, their meaning, the way they are used, and so on.

When we lose these practices (or find ourselves surrounded by unfamiliar ones), we risk becoming disoriented. If the complex structure of practices around us forms a major part of the scaffolding we use to find our way around, to understand what we do and the environment in which we do it, when it is removed (or significant parts of it are removed) we are lost. As mentioned above, some degree of disorientation of this kind can be all-things-considered healthy and good. By encountering unfamiliar cultural practices, we may learn new ways of understanding the world and open up new possibilities. And over time, we generally adapt to new social environments. But where the loss of familiar practices is too extensive and sudden, its effect can be drastic.

---

see also Patten, *Equal Recognition*, ch. 3; and compare Lenard, "Culture, Free Movement, and Open Borders."

26 The constitution of options is *one* way in which social practices form a basis for orientation in the world, but not the only one.

27 Haslanger, "Culture and Critique," 154–57, "What Is a Social Practice?" 7–8, and "Cognition as a Social Skill"; Lewis, *Convention*; and Bourdieu, *Esquisse d'une Théorie de la Pratique*, 166–68.

28 See Lewis, *Convention*, 76.

It is also not the case that all social change is *disorienting*. It is a normal part of the course of social life that shared practices change and evolve constantly.[29] They change for all sorts of reasons: they change as we learn new things, as social knowledge accumulates, and as we adapt to changing external circumstances. They change also as we have new ideas and as we deliberately reshape our practices. And they change as new people become involved in them, and as different practices influence each other and combine. It would be quixotic, and indeed undesirable, to seek to maintain perfect sociocultural stability, and this sort of ordinary change need not impair our ability to understand our environment at all. As mentioned before, precisely what is involved in orientation is the capacity to predict and make sense of *changes*. Our social practices would not do a good job at orienting us in the world if they were overly rigid and inflexible. (There are also important independent reasons that it is better to have cultural practices that are not *too* stable, that are adaptable and not stagnant. It might be thought that through cultural exchange and the meeting of minds we produce *better* cultural practices: we best address the problems we face by constantly being ready to learn from each other.[30] Cultural exchange and flexibility might also be valuable in that it promotes the ability to understand and empathize with others.[31] And finally, we need our practices to be adaptable in order for it to be possible to question and alter unjust and oppressive practices.[32])

But on the other hand, social practices would not serve an orienting function, and would not really qualify as social practices, if they did not exhibit a certain degree of stability over time. The interest I am describing is thus merely an interest in avoiding *excessive* and *overly rapid* sociocultural change. The line between those changes we have an interest in avoiding and those that are part of the normal course of cultural evolution is not one I intend to draw in any precise way.[33] *Magnitude* of change, *breadth* of change across the full set of practices that individuals or groups draw on, and *speed* of change are all relevant to fixing this line. The interest will only clearly be set back when there are changes significant on all three dimensions (i.e., large, broad, *and* rapid

29 See Scheffler, "Immigration and the Significance of Culture."

30 See Waldron, "Minority Cultures and the Cosmopolitan Alternative"; Kuper, *Culture*, 243; Appiah, *The Lies That Bind*, ch. 6.

31 Cf. Nussbaum, *Cultivating Humanity*.

32 None of these considerations need conflict with the interest in stability so long as they can be achieved through an openness to ordinary gradual evolution rather than sudden dramatic change.

33 It is worth noting that the kind of change an individual has an interest in avoiding is determined objectively by what causes the kind of disorientation I have described, but the degree or kind of change that does this may vary from individual to individual.

changes).[34] Further, breadth of change for an individual or group is in turn a function not only of the number of practices that disappear or are replaced, but also of the *importance* of those practices. An individual's set of practices is more broadly affected in this sense when a practice central to their way of life or orientation is lost than when a more peripheral practice is lost.

2.1. Societal or National Cultures

As noted, somewhat similar arguments to the one I have just made (although usually focused more narrowly on the *options* available to us) were put forward by "liberal multiculturalists" as part of a case for granting group rights to minority cultural groups. These writers generally argued that choice (and typically they appealed to the stronger ideal of *autonomy*) depends not only on social practices (*culture*, uncountable) and their maintenance, but also on access to *a* particular culture (the countable concept, a discrete individuable body of cultural practices unified in some way). In particular, they have argued that a "societal" or "encompassing" culture is necessary. A "societal culture," Kymlicka tells us, is "a culture which provides its members with meaningful ways of life across the full range of human activities, including social, educational, religious, recreational, and economic life, encompassing both public and private spheres." One feature of an "encompassing group" for Margalit and Raz is that its members share a culture across various aspects of life.[35] But this aspect of the liberal multiculturalist view has been convincingly disputed by a number of writers, who argue, in my view correctly, that the liberal multiculturalist falsely reifies (or essentializes) cultures as discrete, delineable wholes.[36] Raz is right to claim that cultural practices come in interlocking webs: individual practices are intertwined with each other and often depend on each other.[37] But these interlocking webs do not (generally) clump together into separable, unified *cultures* shared by delineable, non-overlapping groups of people.[38] Rather, there is a sea of interlocking practices, and the set of practices in

34 Speed is not a fully independent dimension: a set of practices changes rapidly when a broad range of its elements change all at once, or in quick succession.

35 Kymlicka, *Multicultural Citizenship*, 76; Margalit and Raz, "National Self-Determination," 80. Raz also defends this idea ("Multiculturalism"). See also Miller, *On Nationality*, 85–87.

36 See, for instance, Barry, *Culture and Equality*, 11, 258–64; Benhabib, *The Claims of Culture*; Carens, *Culture, Citizenship, and Community*, ch. 3; Phillips, *Multiculturalism without Culture*; Scheffler, "Immigration and the Significance of Culture"; and Waldron, "Minority Cultures." See also Appiah, *The Lies That Bind*, ch. 6; Clifford, "Introduction," 19; Haslanger, "What Is a Social Practice?" 8; and Wedeen, "Conceptualizing Culture."

37 Raz, "Multiculturalism," 177.

38 Cf. Benhabib, *The Claims of Culture*, 60; Waldron, "Minority Cultures," 781–86. Patten (*Equal Recognition*, ch. 2) gives the best account of how it might be possible to make sense

which an individual participates, and that forms the background by which they orient themselves, is likely to differ slightly from the equivalent set for their neighbor. There may be certain groups that have more salience than others in terms of cultural commonalities. But the groups among which social practices are shared are quite heterogeneous: some may exist at a quite local level, others at a supranational, regional, or even global level, while yet others cross-cut national or geographic boundaries.

Whether or not this is right, nothing in the view I have set out above commits me to the thought that anyone has an interest in the stability of *national* or *societal* cultures, or of *any* sort of bounded, delineable *cultures*, or the *survival* of individuable cultures generally.[39] What we need to orient ourselves in the world is for there to be a relatively stable set of social practices on which we can rely (there needs to be stable *culture*, not *a* stable culture). There is no reason to think this requires a unified body of practices shared with a discrete homogeneous group of others.

Perhaps, though, one might still be concerned that the charge of reification could be leveled at my account, even when distinguished from the liberal multiculturalist view. If Benhabib, for instance, is right that *cultures* are essentially contested, and "internally riven by conflicting narratives," maybe the same could be said for the social practices that my account does depend on.[40] If this objection is thought to entail that there *are no such things* as social practices that can be relied on for purposes of orientation and that can be held relatively stable over time, I think it is false. And if not, I do not think it conflicts with the above. Even if it is the case that social practices are constituted through processes of contestation (and so constantly open to challenge and redefinition), that does not entail that there are no practices in existence that could, at least for some time, provide a fixed point for understanding the world. As I have said above, my account does not depend on the assumption that social practices can be insulated against change entirely, nor that they have a fixed essence independent of ongoing processes of creation and contestation. My account also does not depend on any claims about the individuation of practices. It could be that there are no bounded, delineable practices with a single determinate social meaning shared by all and only the participants in the practice. Perhaps

---

of such a clumping, but to the extent that he is successful in offering a way to individuate cultures, I think it will have the result that there are *very many* overlapping and cross-cutting cultures. I do not find plausible his claim that some of these cultures will constitute *societal* cultures in Kymlicka's sense (*Equal Recognition*, 62–64), or at least not that are of significant size.

39 On the latter, see also Taylor, *The Politics of Recognition*.

40 Benhabib, *The Claims of Culture*, ix.

there is nothing but a fluid, undifferentiated mass of patterns of behavior, social meanings, expectations, shared beliefs and values, and so on. My claim is merely that we have a significant interest in a reasonable degree of stability across this web of social patterns on which we rely. The elements of the web drawn on are likely to vary from individual to individual, but each, I claim, has an interest in some degree of stability across that part of the web closest to them. One can of course always bring into question what one has previously taken as a fixed point. But one cannot question everything at once, and a loss of too many of one's fixed points in quick succession can be disorienting in a damaging way.

2.2. Land Use

There is one final observation that can be added to this account of the interest in sociocultural stability. The practices across which we may have an interest in maintaining some degree of stability are often intimately bound up with *land use* in a couple of different ways.[41] That is, keeping these practices stable will often require the people involved to remain in a particular geographical location, and for their ability to use a physical area of land in certain ways to be maintained. First of all, *social* practices are created and maintained *communally*. They thus depend for their existence and stability on the existence and stability of the *communities* whose practices they are. This is not to suggest that these communities need ever have a fixed membership, or be protected against compositional change. Nor need it be to suggest that there are unified "encompassing" communities that share practices across the full range of human activities. But a practice will normally disappear when the community engaged in it disappears or disperses. And these communities are often geographically located. Thus, stability in social practices that are like this will derivatively depend on the continued geographical proximity of their participants.

Second, the cultural practices we have an interest in maintaining may themselves *be* practices of land use. A good range of the cultural practices in which an individual is engaged will be practices that in some way make use of land, and so in which an area of land is essentially involved. Such practices may involve transforming the land itself in a productive way or making use of natural resources, or they may be practices that require a certain amount and/or kind of physical space to be carried out. Some practices require only access to *some* land, and *which* area of land they are carried out in is incidental (in some cases only the *amount* of land will matter, while in others land of a certain kind, with certain generic features, will be necessary). Many agricultural practices are like this,

41 Roughly the same point has been made by Stilz, *Territorial Sovereignty*, 41; and Moore, "The Taking of Territory," 94.

as are many practices of modern urban life. Other practices require access to a *specific* area of land, perhaps because of certain unique characteristics that it has (whether natural characteristics or features with which it has been endowed by human activity), or perhaps because of its symbolic or emotional significance to those engaged in the practice. A number of religious practices involve particular places in this way (and religious practices are often especially central to an individual's orientation in the world). Religious practices involving sites in Jerusalem, Mecca, Amritsar, or Rome, for instance, may be of this kind, while much larger areas of land and natural features play central roles in various indigenous American religions.[42] Certain agricultural or hunting practices are also tied to particular places, such as the fishing practiced in collaboration with dolphins in Laguna, Brazil; Sioux buffalo hunting in the American Plains; or Sámi reindeer herding in northern Scandinavia and Russia.[43]

3. A *PRO TANTO* RIGHT TO STABILITY OF LAND-USE PRACTICES

This interest in sociocultural stability, then, is derived from the importance of a somewhat stable background of social practices for what I am calling *orientation*. The moral significance of that, in turn, may be twofold. First, I suspect that orientation may make a non-derivative contribution to well-being. For cognitive processors like us, it seems possible that there is a distinctive value to the successful exercise of cognitive capacities for *practical* purposes. I do not have a worked-out theory of what such a value would be, and nothing will turn on whether this is correct, but the idea seems to have some intuitive plausibility. Second, and more importantly, orientation is of derivative importance for individual *agency*. A certain degree of orientation is, I think, a necessary precondition for an individual to achieve agency, where "agency" is the status of a being that intentionally *acts* in the world. To see oneself as an agent is to see oneself, and crucially, one's intentional states as, in certain ways, shaping the world, not merely being shaped by it. Action, in the sense we are interested in, involves some sort of interaction between an agent's internal states (or events) and the external world (in the right agent-to-world direction).[44] Agency, in this sense, seems plausibly to be a basic and morally significant feature of those creatures that possess it. This status, in addition, seems to impose moral demands on others. Respect for another with a capacity for agency requires treating their exercise of this capacity with sufficient

42 Deloria, *God Is Red*, 75–81.

43 On fishing, see Tennenhouse, "These Fishermen-Helping Dolphins Have Their Own Culture"; on the Sámi, see Benko, "Sámi." For discussion of the Plains buffalo hunters, see Stilz, "Settlement, Expulsion, and Return," 360.

44 Cf. Schlosser, "Agency," sec. 3.

concern. It seems that achieving this status is not merely a binary matter of successfully acting on some occasion: you can possess agency to a greater or lesser extent as the "domain" or scope of your action (or possible action) varies. The more extensive the domain in which you act (assuming there is some way to quantify this), the greater the extent of your agency. If the range of things that you can do and the range of spatio-temporal locations in which you can act is very limited, it makes sense to say that your agency is stunted or restricted (even if, in a minimal binary sense, you still qualify as an agent). Respect for another as an agent, it seems plausible to think, involves refraining from avoidably stunting their agency in this way. It seems there is a vague threshold of agency below which you cannot consider yourself a genuinely active part of the world, and this threshold, though vague, seems to have moral importance.

Intentional action depends (if not always, then at least nearly always) on *some* understanding of the world. Since, in most cases, the possibility of performing any given intentional action depends on some understanding of the elements of the world you intend to involve in your action (and their interrelationships), the scope of your (possible) agency will tend to expand with your understanding of your surroundings. This is unquestionably true of complex *social* actions. To engage in social interaction requires some understanding of human behavior, an ability to interpret the movements and utterances of others, and, probably, some limited capacity for "mind reading" (inferring mental states from the observable behavior of others). The understanding we need for these purposes is precisely that which I have been referring to as "orientation," an understanding of the regularities and fixed patterns and relationships that structure your environment. Insofar as the moral significance of orientation is derivative in this way, not all practically relevant understanding will be of equal importance. Some elements of environmental understanding are central to our overall orientation, and hence to our ability to act, while others are more peripheral. Stability in those aspects of the environment that play a more crucial role will thus be of more importance than stability in others. Further, while greater practically relevant understanding will generally expand the scope of agency, what will matter most is that you be sufficiently "oriented" to meet the vague threshold for genuine agency mentioned above.

The ideal of agency I appeal to here is different from, and more basic than, the kind of ideal of autonomy or planning that is appealed to in defense of pre-institutional territorial or property rights (discussed above). The latter ideal could be cashed out in various ways, but central to it will need to be some sort of capacity for temporally extended planning and some reasonable ability to count on success in bringing projects developed over time to fruition. Agency is a much more basic prerequisite of such an ideal. To have agency in this sense is simply to be a

being that acts in the world over a sufficient proportion of its life; it is a further achievement to string this together into coherent projects extended over time. The claim I want to make here is that this weaker ideal is enough to account for a possible wrong of settlement (via the idea of orientation). The much weaker idea that we have some natural obligations to respect others' agency, and, derivatively from that, their need for orientation, is sufficient for this purpose and does not lead to any justification of exclusionary rights over territory.

3.1. A Pro Tanto *Right*

There is, then, a morally significant interest in orientation. I think that that interest suffices to support a weak, *pro tanto* right to some degree of environmental, and notably sociocultural, stability. Because this right is grounded in interests in orientation and agency, it is a right to *stability*, not *control*. Absent comparably significant countervailing considerations, the suggestion is, it would be *wrong* to do something that severely disrupts the web of social practices on which someone relies against their will. Just as I have said that we have no interest in perfect sociocultural stability, in protecting our practices generally against change and evolution, there is also no right to perfect sociocultural stability. The *pro tanto* right is merely to a moderate degree of stability across our cultural practices; it is a right against excessive and overly rapid changes to the overall web of practices on which we rely. This right is not a property-like right *over the land*. Rather, it is a right to do certain things—namely, to continue to participate in and rely on a moderately stable range of social practices, including, notably for our purposes, practices of land use. I will elaborate this point further below.

The rights I am describing are individual rights, even if it is not possible to describe them without reference to groups. To accept this picture of a right to sociocultural stability and of a possible wrong of settlement, there is no need to believe in groups or collectives with the kind of ontological standing to be right holders. In many cases, the practices or patterns of land use to which individuals have a right will be irreducibly collective. But the *right* to stability in these practices (along with the corresponding interest) is held by individuals. It is individuals, on my story, that come to depend on particular background patterns of social practices for their orientation in the world. And so, even if these practices are necessarily collective practices, it is individual participants in them that have a right to their maintenance.

3.2. Limitations and Objections

The propensity of change in practices to provoke disorientation does not depend on the practices in question having any sort of value. Even if some

practices treat you oppressively or unjustly, you may still be disoriented by their loss. That disorientation, considered in isolation, is a *pro tanto* bad; it is a respect in which your interests have been set back. If the injustice is significant, though, that bad will clearly be outweighed. There can be no *right*, however, against disruption of social practices that are morally objectionable. The fact that you may be harmed by the disorientation you would experience at the loss of such a practice does not give rise to a right against such a harm when you are anyway morally bound to be rid of the practice. (It is also worth noting at this point that, although the right is held by all, those with greater social and economic power are much less *likely* to be victims of wrongful cultural disruption. Social and economic power tends to bring with it (a) means to control and shape the social practices that surround you, and (b) the ability to develop means and strategies for adapting to and orienting oneself in new social and cultural environments.)

It is also worth clarifying that this does not constitute a general defense of social stability, or a general call for the deceleration of cultural change. This is the case in two respects. First, the argument I have given does not offer any reason to think that traditional ways of doing things are good in virtue of being traditional, or that tradition, as such, is normative.[45] I have stressed that the right is a right against *extreme* cultural change—change that is rapid, substantial, and *broad*, i.e., that extends across a wide range of the cultural practices in the web that an individual draws on. The picture is *not* one according to which stability, in whatever degree, is a good thing but minor instabilities are outweighed. Rather, there is no complaint at all against changes that do not provoke severe disorientation. Thus, the right to sociocultural stability does not give us reason, for any individual practice in isolation, to preserve it from change. It *only* gives us reason to pay attention to the overall web of practices, and to ensure that it is not too radically or rapidly overhauled. Only when there is risk of this does the right give us reason to protect any individual practice from change.

The second respect in which this is not a general defense of sociocultural stability is that the right to sociocultural stability is only one consideration among many relevant to all-things-considered moral judgments. As I have said, the right is only a weak, *pro tanto* one (on which more below). There are many independent values that may outweigh the interest in sociocultural stability and demand change even despite the severe disorientation that it will bring. The lesson that we *should* draw in cases like this is that the disorientation caused by such rapid change *ought to be taken into account*. And where it is ultimately

45 On the normativity of tradition, see Scheffler, "The Normativity of Tradition"; and Jeffers, "The Ethics and Politics of Cultural Preservation."

outweighed, it should not simply be forgotten. It may be incumbent upon us, for instance, to pursue whatever means are available to *limit* or *mitigate* the disorientation caused by otherwise positive change.

4. Settlement and the Violation of Sociocultural Stability Rights

People have weak, *pro tanto* rights to moderate stability in social practices that are often bound up with land use. These rights can thus be violated when people's ability to be in or use land in particular ways is disrupted. The most obvious way in which this might be done is when individuals or whole groups are physically removed from an area in which the practices to which they have rights are located. If you are suddenly forcibly removed from the area in which you live, you will most likely be separated from the communities with whom the practices familiar to you are shared, and the particular area of land on which some of the relevant practices may depend. But I think it should also be apparent that this is not the only way in which sociocultural stability rights might be violated. In particular, the *settlement* of a large or powerful group of newcomers in an area, bringing with them different, incompatible land-use practices, may do the same.[46] Moore and Stilz have convincingly argued that settlement can disrupt the life plans and projects of existing residents in an area.[47] It is no less plausible, I think, that settlement may, in certain cases, severely disrupt a background web of social practices so as to disorient existing inhabitants in a way that violates the right described above.

Of course, it is not the case that settlement *generally*, as a matter of course, does cause disruption of such significance. Settlement can *only* violate the rights described when it involves the importation of land-involving practices that are incompatible with, and so disrupt, those of existing inhabitants. The account offered here could give *no* complaint against settlers who arrive and join or adopt the practices already prevalent in the area. And this right only makes settlement wrongful where the disruption it brings about is significant, broad, and rapid enough to create serious and harmful disorientation. But it does seem that in certain particular kinds of case the settlement of a large group could have such an effect. As Moore has pointed out, different land-use practices may be incompatible with one another, so a settler group's simply settling

46 Settler colonialism frequently involved both the *coercive* imposition of new cultural practices and forms of epistemic injustice involving disrespectful treatment of existing cultural practices of indigenous groups. These things plausibly exacerbate the wrong done by settlement, and are wrong independently of the settlement itself; neither are *necessary* for the wrong I describe.

47 Stilz, "Settlement, Expulsion, and Return," 360; Moore, "The Taking of Territory," 94–98.

in an area where an existing group already has certain ways of using the land, and pursuing their own practices of land use, without attempting to remove the indigenous group from the land, may be enough to make it impossible for the indigenous group to maintain their existing practices. For instance, Moore says, "settled farming in enclosed fields is disruptive of nomadic hunting and gathering or slash-and-burn agriculture."[48] The movement of white settlers across the American Plains that Stilz describes seems like another example.[49] This settlement drove away the buffalo on which hunting practices core to the Plains tribes' mode of existence depended. The Hawaiian case mentioned in the introduction also seems like it might fit this model. Foreign settlers (and missionaries and traders) in Hawai'i brought with them different systems of using and dividing land, and their influence led to the "Māhele," a privatization of land, a radical shift in ways of relating to territory. This seems to have caused significant disorientation among the indigenous population, who had lost a familiar framework for understanding their social and territorial world, a fact settlers exploited to shift land into their hands.[50]

The practices disrupted need not be agricultural or economic practices. An interesting example is that of the indigenous people of North America, for many of whom religious belief was closely tied both to particular places and to particular geographical communities (for many, subsistence *also* depended on the use of large areas of land of a particular kind).[51] Settlement that altered these peoples' access to the relevant places (or that altered features of these places with deep religious significance), then, seems likely to have struck at practices at the core of their members' understanding of their place in the world. Rapid settlement by a large group of newcomers could also change the *social* environment without altering the possibilities for land use directly.

48 Moore, "The Taking of Territory," 96.

49 Stilz, "Settlement, Expulsion, and Return," 360.

50 See Osorio, *Dismembering Lahui*, ch. 2; Silva, *Aloha Betrayed*, 39–43; and Kauanui, *Paradoxes of Hawaiian Sovereignty*, ch. 2. The historical evidence here is complicated. These changes were legal changes made by the Hawaiian king. Most of the evidence we have of the impact of these changes on Hawaiians comes from these ruling classes and the settlers themselves, so any claim about disorientation suffered by ordinary Hawaiians is necessarily speculative. But there is some evidence that there were effects of this kind, for instance, from the many petitions made by ordinary Hawaiians to the government expressing concern about foreign ownership of land and the stability of traditional systems of chiefly rule, as well as the success with which the local ruling classes and foreign settlers were able to exploit ordinary Hawaiians' loss of familiar frameworks for understanding their social and territorial world in order to shift land into their hands.

51 See Deloria, *God Is Red*, 75–81, 200–201. Thanks to Liz Reese for drawing my attention to this.

Changing the cultural practices (linguistic practices, for instance) prevalent in the area could make it suddenly difficult for existing inhabitants to find their way around the *social* world in which they live. Where this is excessively rapid and broad, it could be seriously disorienting.

Thus, I think the right to sociocultural stability can account for at least a *possible* wrong of settlement. In addition, it seems plausible that, although the wrong described is not *conceptually* tied to settlement with colonial ambitions, the kinds of attitudes, ideas, and goals associated with historical projects of settler colonialism make it particularly likely that such a wrong will be done. Where settlement goes along with a conceptualization of in-fact-inhabited land as empty, an idea of existing inhabitants as racially or culturally inferior, and aspirations to recreate the "civilization" of the motherland in a supposedly "uncivilized" territory, there is good reason to expect, at a minimum, callous disregard for the disorientation of prior inhabitants.[52]

Finally, an *individual* settling on their own in an area in which existing inhabitants have weak rights to sociocultural stability is unlikely ever to violate these rights. An individual's settlement on its own will rarely, if ever, cause sufficient disruption. It is only when sizeable groups settle in an area *together* that a wrong might be done. It seems clear that there can be wrongdoing, rights violation, or injustice that only occurs when a group of individuals *all* behave in certain ways, i.e., where no individual's action is wrong in the absence of the actions of a number of other individuals. The wrong of settlement is usually such a case. This might lead us to wonder, though, when exactly (if ever) an *individual* acts wrongly by settling in a new area. This raises tricky questions about the distribution of collective wrongs to individuals.[53] I do not have an answer to these questions, but for what it is worth, it does seem plausible that, at least sometimes, an individual's choosing to settle in the context of a large number of others' doing so, and in full knowledge that they are doing so (and that collectively they will cause serious and wrongful disruption to existing inhabitants), will be an individual wrong.

5. THE RIGHT TO EXCLUDE

So, it seems that the description of the right to sociocultural stability I have given, if plausible, offers one way to account for the thought that there can be something distinctively wrong with *settler* colonialism. The right to moderate

52 See Bell, *Reordering the World*, 38–39.

53 On this, see for instance Kutz, *Complicity*; Smith, "Non-Distributive Blameworthiness"; Kagan, "Do I Make a Difference?"

sociocultural stability I posit is, like the occupancy rights Moore and Stilz posit, a right that people have independent of any institutions or conventions granting these rights to individuals. It is a right that flows more basically from an imperative to respect the agency of others. It is, though, a much weaker and more limited right than the occupancy rights that, for Moore and Stilz, support a right to exclude or legitimate authority over access to a territory. Unlike the stories told by Moore and Stilz, my account does *not* support any sort of property-like *control* rights over territory.

To be wronged by settlement in a territory, all that needs to be the case is that the settlement unnecessarily severely disrupts the scheme of practices on which you rely to orient yourself in the world. You do not have to have any special claim to the territory or legitimate authority over access to it. It does not in any (even minimal) sense have to be *yours*. And you do not have to have any more claim *to the territory* than do the settlers. As noted before, since it is grounded in an interest in avoiding disorientation, not a plan-based interest, the right is a right to *stability*, not control. That you may be wronged in certain cases by others entering a territory does not mean that you have the right to decide who may and may not enter. (As noted above, *mere* entry will never violate the right: to do so, settlers must bring with them incompatible land-use practices.)

We all have interests in and rights to sociocultural stability of equivalent weight. These impose duties on others to do what is necessary to allow you to maintain an appropriate degree of sociocultural stability where possible without setting back interests of comparable significance. Where sociocultural stability for an individual or group involves stability of land use, outsiders will be under a *pro tanto* duty to refrain from disrupting the relevant practices. Current occupation of a space does tend to generate an additional interest in continued use of it that non-occupiers do not have, insofar as orientation in the world tends to depend on a particular place in which one is a resident. But none of this is because existing residents have any claim or authority over the land that outsiders lack. If outsiders *also* have a significant interest in using the same area of land that (for whatever reason) cannot be met without disrupting the practices of existing users, this may suffice to outweigh the right. Their interests or rights are not to be given any less weight on account of their being outsiders.

The right to sociocultural stability is only *pro tanto*. Thus, it will not *always* be wrong (all things considered) to cause severe disorientation; it is wrong just when the disruption is not required for any comparably weighty interests or rights to be met. Because of the disorientation that results from a significant and sudden disruption to a set of cultural practices, the interests of outsiders in using an area of land in a way that would cause such a disruption can only justify

doing so if there is no other feasible and less costly way of meeting the interest. But suppose that a group of outsiders *needs* to settle in territory *T* for some very weighty interest to be met (say, to survive), and if they do not reconstitute some of their existing practices there, *they* will suffer severe disorientation in their new environment. Suppose also that their existing practices will require using the land in *T* in a way incompatible with the practices of *T*'s existing inhabitants. The group of outsiders cannot meet their weighty interest in survival without causing severe disorientation to *either* themselves or *T*'s existing inhabitants. The fact that the latter were there first is of no moral significance on this account. In such a case, there is no obligation on the outsiders to bear the "disorientation cost" of their settling in *T*.[54]

Let us finish with one final question: Does it follow from this that those wronged by settlement have the right to exclude in the sense of the right to *enforce* demands about immigration? The answer, I think, is no. It does not *follow* from the fact that *A*'s action would be wrong that it would be permissible for you to force *A* not to do it. There are a good many moral duties that are not permissibly enforced. It is usually wrong, we tend to think, to break a promise, but we do not usually think that it is permissible to *force* a promisor to keep their promise. So, it does not follow from the conclusion that settlement is sometimes wrong that any inhabitants of a territory have the right to forcibly keep others from *settling* in it. I think it is quite plausible that forcibly resisting wrongful settlement will *sometimes* be justifiable, but this is not an immediate consequence of my account of the wrong. Certainly, it would be justifiable to forcibly resist settlement *accomplished by the use of force* and to resist forcible *removal*. This is, I think, unproblematic. There is, though, no reason to think that the cases in which forcible resistance to settlement is justified will be all those in which settlement would be wrong.

6. Conclusion

I have presented an account of an interest people can have in moderate stability across the social practices that surround them, derived from the necessity of a degree of such stability for an individual's ability to orient themselves in the world, which may matter both independently and as a precondition for agency. This offers an alternative explanation of how individuals can come to have legitimate expectations of continued use of a territory, and so rights that could be violated by settlement, to the usual plan-based story. This allows us to

54 This is where my account diverges substantially in its practical consequences from Stilz's, despite her relative skepticism about the extent of the *exclusion justification* held by possessors of territorial rights.

account for a possible wrong of settlement, and so a wrong in settler colonialism independent of the features it shares with other forms of colonialism and imperialism, without positing any exclusionary territorial rights on the part of those wronged. Not only do we not need to say that inhabitants of a territory are generally *justified* in excluding from that territory, but we also do not need to say that they have the *legitimate authority* to do so.[55]

London School of Economics and Political Science
daniel.guillery@gmail.com

REFERENCES

Appiah, Kwame Anthony. *The Lies That Bind: Rethinking Identity*. New York: Liveright Publishing Corporation, 2018.

Barry, Brian. *Culture and Equality: An Egalitarian Critique of Multiculturalism*. Cambridge, MA: Harvard University Press, 2001.

Bell, Duncan. *Reordering the World: Essays on Liberalism and Empire*. Princeton: Princeton University Press, 2016.

Benhabib, Seyla. *The Claims of Culture: Equality and Diversity in the Global Era*. Princeton: Princeton University Press, 2002.

Benko, Jessica. "Sami: The People Who Walk with Reindeer." *National Geographic*, November 2011. https://www.nationalgeographic.com/magazine/2011/11/sami-reindeer-herders.

Bourdieu, Pierre. *Esquisse d'une Théorie de la Pratique, précédé de Trois Etudes d'Ethnologie Kabyle*. Geneva: Librairie Droz, 1972.

Carens, Joseph H. "Aliens and Citizens: The Case for Open Borders." *Review of Politics* 49, no. 2 (Spring 1987): 251–73.

———. *Culture, Citizenship, and Community: A Contextual Exploration of Justice as Evenhandedness*. Oxford: Oxford University Press, 2000.

———. *The Ethics of Immigration*. Oxford: Oxford University Press, 2013.

Clifford, James. "Introduction." In *Writing Culture: The Poetics and Politics of*

55 This paper owes a lot to discussions with a number of people. Thanks first to an anonymous reviewer for this journal whose comments helped me to substantially revise the paper. The paper also profited greatly from discussions with audiences at the 2020 PPE Society meeting in New Orleans, the 2021 Lisbon Conference on the Philosophy of Migration and Asylum, the Faculty Work-in-Progress seminar at the University of Chicago Law School, the online migration ethics workshops organized by Kieran Oberman, and comments from students in the University of Chicago Law and Philosophy workshop. Special thanks for valuable extended discussion and/or written comments are owed to Joe Carens, Sarah Fine, Erin Miller, Martha Nussbaum, Kieran Oberman, Liz Reese, and Tyler Zimmer.

Ethnography, edited by James Clifford and George E. Marcus, 1–26. Berkeley, CA: University of California Press, 1986.

Cole, Phillip. *Philosophies of Exclusion: Liberal Political Theory and Immigration*. Edinburgh: Edinburgh University Press, 2000.

Deloria, Vine, Jr. *God Is Red*. New York: Grosset and Dunlap, 1973.

Dworkin, Ronald. "Can a Liberal State Support Art?" In *A Matter of Principle*, 221–33. Cambridge, MA: Harvard University Press, 1985.

Ferguson, Benjamin, and Roberto Veneziani. "Territorial Rights and Colonial Wrongs." *European Journal of Philosophy* 29, no. 2 (June 2021): 425–46.

Grimm, Stephen. "Understanding." In *Stanford Encyclopedia of Philosophy* (Summer 2021). https://plato.stanford.edu/archives/sum2021/entries/understanding/.

Haslanger, Sally. "Cognition as a Social Skill." *Australasian Philosophical Review* 3, no. 1 (2019): 5–25.

———. "Culture and Critique." *Aristotelian Society Supplementary Volume* 91, no. 1 (June 2017): 149–73.

———. "What Is a Social Practice?" *Royal Institute of Philosophy Supplement* 82 (July 2018): 231–47.

Huemer, Michael. "Is There a Right to Immigrate?" *Social Theory and Practice* 36, no. 3 (July 2010): 429–61.

Jeffers, Chike. "The Ethics and Politics of Cultural Preservation." *Journal of Value Inquiry* 49, nos. 1–2 (March 2015): 205–20.

Kagan, Shelly. "Do I Make a Difference?" *Philosophy and Public Affairs* 39, no. 2 (Spring 2011): 107–41.

Kauanui, J. Kehaulani. *Paradoxes of Hawaiian Sovereignty: Land, Sex, and the Colonial Politics of State Nationalism*. Durham, NC: Duke University Press, 2018.

Kohn, Margaret, and Kavita Reddy. "Colonialism." In *Stanford Encyclopedia of Philosophy* (Fall 2017). https://plato.stanford.edu/archives/fall2017/entries/colonialism/.

Kuper, Adam. *Culture: The Anthropologist's Account*. Cambridge, MA: Harvard University Press, 1999.

Kutz, Christopher. *Complicity: Ethics and Law for a Collective Age*. Cambridge: Cambridge University Press, 2000.

Kymlicka, Will. *Liberalism, Community, and Culture*. Oxford: Oxford University Press, 1989.

———. *Multicultural Citizenship: A Liberal Theory of Minority Rights*. Oxford: Oxford University Press, 1995.

Lenard, Patti Tamara. "Culture, Free Movement, and Open Borders." *Review of Politics* 72, no. 4 (Fall 2010): 627–52.

Lewis, David. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press, 1969.

Margalit, Avishai, and Joseph Raz. "National Self-Determination." *Journal of Philosophy* 87, no. 9 (September 1990): 439–61.

Miller, David. *On Nationality*. Oxford: Oxford University Press, 1997.

Moore, Margaret. "Justice and Colonialism." *Philosophy Compass* 11, no. 8 (August 2016): 447–61.

———. *A Political Theory of Territory*. New York: Oxford University Press, 2015.

———. "The Taking of Territory and the Wrongs of Colonialism." *Journal of Political Philosophy* 27, no. 1 (March 2019): 87–106.

Nussbaum, Martha. *Cultivating Humanity: A Classical Defense of Reform in Liberal Education*. Cambridge, MA: Harvard University Press, 1997.

Oberman, Kieran. "Immigration as a Human Right." In *Migration in Political Theory: The Ethics of Movement and Membership*, edited by Sarah Fine and Lea Ypi, 32–56. Oxford: Oxford University Press, 2016.

Osorio, Jonathan K. Kamakawiwo'ole. *Dismembering Lahui: A History of the Hawaiian Nation to 1887*. Honolulu: University of Hawaii Press, 2002.

Patten, Alan. *Equal Recognition: The Moral Foundations of Minority Rights*. Princeton: Princeton University Press, 2014.

Phillips, Anne. *Multiculturalism without Culture*. Princeton: Princeton University Press, 2007.

Raz, Joseph. "Multiculturalism: A Liberal Perspective." In *Ethics in the Public Domain: Essays in the Morality of Law and Politics*, 170–91. Oxford: Oxford University Press, 1994.

Sanders, John T. "Projects and Property." In *Robert Nozick*, edited by David Schmidtz, 34–58. Cambridge: Cambridge University Press, 2002.

Scheffler, Samuel. "Immigration and the Significance of Culture." *Philosophy and Public Affairs* 35, no. 2 (Spring 2007): 93–125.

———. "The Normativity of Tradition." In *Equality and Tradition*, 287–311. New York: Oxford University Press, 2010.

Schlosser, Markus. "Agency." In *Stanford Encyclopedia of Philosophy* (Winter 2019). https://plato.stanford.edu/archives/win2019/entries/agency/.

Silva, Noenoe K. *Aloha Betrayed: Native Hawaiian Resistance to American Colonialism*. Durham, NC: Duke University Press, 2004.

Simmons, A. John. *The Lockean Theory of Rights*. Princeton: Princeton University Press, 1992.

Smith, Thomas H. "Non-Distributive Blameworthiness." *Proceedings of the Aristotelian Society* 109, no. 1 (April 2009): 31–60.

Stilz, Anna. "Property Rights: Natural or Conventional?" In *The Routledge Handbook of Libertarianism*, edited by Jason Brennan, Bas van der Vossen,

and David Schmidtz, 244–58. New York: Routledge, 2018.
———. "Settlement, Expulsion, and Return." *Politics, Philosophy, and Economics* 16, no. 4 (September 2017): 351–74
———. *Territorial Sovereignty: A Philosophical Exploration*. Oxford: Oxford University Press, 2019.
Taylor, Charles. "The Politics of Recognition." In *Multiculturalism and "The Politics of Recognition": An Essay by Charles Taylor*, edited by Amy Gutmann, 25–73. Princeton: Princeton University Press, 1992.
Tennenhouse, Erica. "These Fishermen-Helping Dolphins Have Their Own Culture." *National Geographic*, April 9, 2019. https://www.nationalgeographic.com/animals/2019/04/dolphins-fishermen-brazil-culture.
Trask, Haunani-Kay. *From a Native Daughter: Colonialism and Sovereignty in Hawai'i*. 1993. Rev. ed. Honolulu: University of Hawai'i Press, 1999.
Valentini, Laura. "On the Distinctive Procedural Wrong of Colonialism." *Philosophy and Public Affairs* 43, no. 4 (Fall 2015): 312–31.
Van der Vossen, Bas. "Imposing Duties and Original Appropriation." *Journal of Political Philosophy* 23, no. 1 (March 2015): 64–85.
Van Wietmarschen, Han. "The Colonized and the Wrong of Colonialism." *Thought* 7, no. 3 (September 2018): 170–78.
Veracini, Lorenzo. "Introducing." *Settler Colonial Studies* 1, no. 1 (2011): 1–12.
Waldron, Jeremy. "Minority Cultures and the Cosmopolitan Alternative." *University of Michigan Journal of Law Reform* 25, nos. 3–4 (1992): 751–94.
———. "Superseding Historic Injustice." *Ethics* 103, no. 1 (October 1992): 503–19.
Wedeen, Lisa. "Conceptualizing Culture: Possibilities for Political Science." *American Political Science Review* 96, no. 4 (December 2002): 713–28.
Ypi, Lea. "What's Wrong with Colonialism." *Philosophy and Public Affairs* 41, no. 2 (Spring 2013): 158–91.

Journal of Ethics and Social Philosophy
VOL. 25, NO. 2 · AUGUST 2023
https://doi.org/10.26556/jesp.v25i2.1117

ETHICS AND THE QUESTION OF WHAT TO DO

Olle Risberg

THE AIM of this paper is to present and defend an account of a distinctive form of "practical" or "deliberative" question that is central in several debates in ethics, metaethics, and metanormativity more generally. Most writers assume that this question concerns some special normative issue, such as what we ought to do "all things considered."[1] I will argue against this assumption and instead endorse an alternative view, which combines elements of both metaethical cognitivism and noncognitivism. A notable consequence of this view is that even if there are truths about how we (all things considered) ought to act—truths that may even be objective, irreducible, and so on—the "central deliberative question," as it is has sometimes been called, does not concern those truths.[2] Instead, that question does not have a true answer.

One debate that highlights the relevant kind of question is the one about normative uncertainty.[3] Since we are not epistemically flawless beings, it seems that we are often (or at least sometimes) not in a position to know what we ought to do. As many have noted, such situations make it natural to ask questions like: "I don't know what I ought to do—*now* what ought I to do?" For obvious reasons, however, it is unclear how this question should be understood. After all, what the agent ought to do is precisely what she does not know!

Another example concerns choices in the face of conflicting normative requirements.[4] If we must choose between promoting the common good and promoting our own good, for instance, the requirements of morality might

1 Similarly to Mark Schroeder and others, I will generally use the term "normative" to mean, roughly, "having to do with value, oughts, reasons, duties, and the like" (Schroeder, "Realism and Reduction," 3), though see section 9 for a discussion of other things that can be meant by "normative."

2 For this expression, see, e.g., Lord, "What You're Rationally Required to Do and What You Ought to Do (Are the Same Thing!)," 1110; and McPherson, "Explaining Practical Normativity," 621.

3 See, e.g., MacAskill, Bykvist, and Ord, *Moral Uncertainty*; Sepielli, "What to Do When You Don't Know What to Do"; and Weatherson, *Normative Externalism*.

4 See, e.g., Chang, "All Things Considered"; and Baker, "Skepticism about Ought Simpliciter."

clash with those of prudence in such a way that we cannot satisfy both. Such situations invite other questions that are difficult to understand, such as: "Which ought—the moral or the prudential one—ought I *really* to satisfy?" Here too it is unclear how to understand the question that is raised, since after all, it is *really the case* that we ought morally to satisfy the requirements of morality, *and* that we ought prudentially to satisfy the requirements of prudence.

I will argue that the salient question in these and other choice situations does not strictly speaking concern what we ought to do, in any sense of "ought." Nor does it concern any other normative question. The reason, as I will argue, is that this question may well remain unanswered even in choice situations where all the truths, *including all the normative truths*, are known. The best explanation of this fact, I suggest, is that while uncertainty about normative questions amounts to uncertainty about the truth of some normative proposition—concerning, e.g., what one ought to do—the "central deliberative question" is instead the question of what *to* do. I further suggest that we understand the question of what to do along the lines suggested by Allan Gibbard.[5] On this view, roughly, one does not answer this question by forming a belief about what the world is like (not even in normative respects), but by forming an intention to act in a certain way. We should thus adopt cognitivism about normative questions but noncognitivism about the question that sometimes seems to remain even when all normative questions are answered. A similar noncognitivist view has recently been defended by Justin Clarke-Doane in response to one of the problems that I will discuss, concerning what Matti Eklund calls *alternative normative concepts*, and one contribution of the paper is to argue that this form of noncognitivism is also plausible with respect to several other problems in ethics, metaethics, and metanormativity.[6]

After briefly introducing the problem of alternative normative concepts and the noncognitivist view about the question that it raises (section 1), I will show how similar questions are also raised by an argument against objective consequentialism due to Frank Jackson (sections 2–3), in the normative uncertainty debate (sections 4–5), and by normative conflicts and what Christine Korsgaard calls the "normative question" (section 6).[7] Along the way, I will consider a number of alternative accounts of how this question should be understood, and argue that they all face important challenges. My final argument to that effect focuses on the possibility that the normative truths are dramatically

5 Gibbard, *Thinking How to Live*.

6 Clarke-Doane, *Morality and Mathematics*; and Eklund, *Choosing Normative Concepts*. See also Balaguer, "Moral Folkism and the Deflation of (Lots of) Normative and Metaethics."

7 Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection"; Korsgaard, *The Sources of Normativity*.

different from what we take them to be (section 7). I then return to the question of how the relevant form of noncognitivism is best understood, explain how my preferred version of it differs from "quasi-realism" about normative judgments (section 8) and consider two challenges to it (section 9). Section 10 concludes.

1. ALTERNATIVE NORMATIVE CONCEPTS

The problem of alternative normative concepts can be introduced by noting that, for the most part, historical contingencies are least partially responsible for what concepts we happen to employ—and, more generally, for how we happen to think. If evolutionary processes had shaped our cognition differently, for example, we might well have represented the world using concepts that we do not in fact have. This raises the question: Could the same be true of our normative concepts, such as OUGHT, GOOD, and REASON?[8] That is, are there *alternative* normative concepts that could play the same role in our lives as our actual normative concepts do, but that are true of different actions, attitudes, and so on? If so, is there any suitably neutral way to ask which set of normative concepts we *ought* to use?

Eklund makes the problem vivid by imagining a community of speakers, "Alternative," who use the concept OUGHT* in much the same way that we use the concept OUGHT. That is, while we perform actions that we judge that we ought to perform, they perform actions that they judge that they ought* to perform; whereas we criticize and resent people who do things that we believe ought not to be done, they criticize and resent people who do things that they believe ought* not to be done; and so on. But in the imagined case, OUGHT and OUGHT* are not coextensive—there are some actions that ought but ought* not to be done (or vice versa). If this case is possible, then, as Eklund notes,

> a first thought one might have is that . . . there is some sort of live issue as to whether we or the alternative community get things right. They do what they do based on considerations about what is "good" and "right" in their sense; we do what we do based on considerations about what is "good" and "right" in our sense. Since our normative terms and their normative counterparts aren't coextensive, we then act differently. . . . [But] what set of normative terms ought to be used when we ask ourselves what to do?[9]

In other words, if we learn that there are alternatives to the normative concepts that we actually have, we might want to ask questions like: What ought we to

8 I use small caps to denote concepts.

9 Eklund, *Choosing Normative Concepts*, 22.

do? Should we "go with" what we ought to do or what we ought* to do? Which set of normative concepts and/or terms ought we use? However, as Eklund goes on to note, it is not plausible that the salient further question literally concerns what normative concepts we ought to use (or any other issue that can straightforwardly be put in terms of our actual normative concepts). The reason is that this question might have an answer that is too easy: perhaps we simply ought to use OUGHT—and perhaps we equally ought* to use OUGHT*! Similarly, perhaps we should go with what we should to do, but should* go with what we should* do—and so on.

Several other views about the salient further question are possible. One is that the case is impossible as described (and so the question does not even arise), because all concepts that have the same "normative role" with respect to guiding behavior are also coextensive.[10] Another view is that the question is in some sense "ineffable"—it is genuine but cannot be perspicuously expressed in our language, and perhaps not in any possible language either.[11] A third view is that, although the case is possible, there is no genuine further question at all—there is only what we ought to do and what we ought* to do and that is that. I mention these views only to set them aside. Instead, as already mentioned, the view that I will ultimately go on to endorse is a kind of noncognitivism about this question. Drawing on Gibbard, Clarke-Doane proposes that the salient further question is best understood as a question of *what to do*; e.g., whether to do what we ought or what we ought* to do, or whether to use OUGHT or OUGHT* in deliberation.[12] This question is meant to be noncognitive in the sense that one does not "answer" or "settle" it by forming a belief about some matter of fact (or by forming some other kind of doxastic attitude). Instead, one answers it by forming a noncognitive attitude of some kind. On Gibbard's view, it is a kind of intention.

What is the relation between the question of what to do and the question of what we ought to do? According to Gibbard's noncognitivism (as it is usually understood), they are simply identical, given that his analysis is supposed to be true of the normative concepts that we in fact have.[13] But other views may also

10 This might follow from certain forms of "conceptual role semantics"; see, e.g., Wedgwood, "Conceptual Role Semantics for Moral Terms." For a suggestion along these lines, see Fitzpatrick, Commentary on Matti Eklund, *Choosing Normative Concepts*. The expression "normative role" is from Eklund (e.g., *Choosing Normative Concepts*, 10).

11 For discussion of this view, see Eklund, *Choosing Normative Concepts*, ch. 2.2; and Clarke-Doane, *Morality and Mathematics*, 172.

12 See Gibbard, *Thinking How to Live*; and Clarke-Doane, *Morality and Mathematics*, ch. 6.

13 There are some interpretative complications, however; for instance, Gibbard at one point suggests that any analysis is likely to "strain" the concept that is analyzed, and proposes only that his view strains our actual normative concepts less than competing views (*Wise Choices, Apt Feelings*, 32).

be had. In particular, the view that I will go on to endorse is that the questions are different, in that the question of what we ought to do is a "question of fact" while the question of what to do is not. Thus, the question that I earlier called the "central deliberative question," and that I take Eklund's imagined scenario to highlight, is on this view understood as a nonfactual question, rather than a factual, normative one. This view, I will argue, best explains why similarly puzzling kinds of questions—or, as I will often put it, similarly puzzling kinds of *uncertainty*—are raised by several other problems in ethics, metaethics, and metanormativity. One promising explanation of this commonality is that all these questions (and the corresponding forms of uncertainty) are of the same kind, and that some kind of noncognitivism is thus true in all these cases. Reflection on these other cases accordingly provides independent support for the relevant form of noncognitivism.

2. OBJECTIVE ACT CONSEQUENTIALISM AND CHOICES UNDER EMPIRICAL UNCERTAINTY

A quite different topic in ethics that brings the central deliberative question to the fore concerns a prominent worry about objective act consequentialism, which is the view that we always ought to perform the action that would in fact have the best consequences.[14] The worry is that in most or all real-life situations, it is impossible for us to know which action the view prescribes. This concern is also what motivates Jackson's objection to the view, which departs from the following case:

> Jill is a physician who has to decide on the correct treatment for her patient, John, who has a minor but not trivial skin complaint. She has three drugs to choose from: drug *A*, drug *B*, and drug *C*. Careful consideration of the literature has led her to the following opinions. Drug *A* is very likely to relieve the condition but will not completely cure it. One of drugs *B* and *C* will completely cure the skin condition; the other though will kill the patient, and there is no way that she can tell which of the two is the perfect cure and which the killer drug.[15]

The problem stems from the fact that according to objective consequentialism, Jill ought to give John the perfect cure, even though she does not know which

14 When context does not indicate otherwise, I use "consequentialism" and "objective consequentialism" to refer to objective act consequentialism.

15 Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," 462–63. See also Regan, *Utilitarianism and Co-operation*; and Kolodny and MacFarlane, "Ifs and Oughts."

drug that is. What Jill knows is only that it is *either* objectively best to give John drug *B* or to give him drug *C*. She can thus deduce that it is objectively suboptimal to give him drug *A*. But since Jill does not know whether drug *B* or drug *C* is the perfect cure, she does not know how to realize the best outcome. In view of this fact about her epistemic situation, Jackson writes that

> [the] problem arises from the fact that we are dealing with an *ethical* theory when we deal with consequentialism, a theory about *action*, about *what to do*.... Now, the fact that an action has in fact the best consequences may be a matter which is obscure to an agent. (Similarly, it may be obscure to the agent what the objective chances are.) In the drugs example, Jill has some idea but not enough of an idea about which course of action would have the best results.... Hence, the fact that a course of action would have the best results is not in itself a guide to action.[16]

When Jill is uncertain about what to do, the argument goes, learning that she ought to perform the objectively best action is useless since she does not know which action that is. Jackson thus concludes that consequentialism—which is a theory about what we *ought* to do—fails to answer the question of what *to* do for agents who do not know how to realize the best outcome.[17]

It is extremely common that we do not know which of our alternative actions are objectively best, however, and the point of the Jill and John case is not merely to emphasize that fact. Our ignorance about this is more easily illustrated by the fact that many of our actions have "massive causal ramifications."[18] In particular, seemingly mundane actions (like buying coffee) may affect what germ cells will ever figure in conception, and thus what people will ever exist. As a result, their impact on the total amount of future well-being can be both dramatic and unknowable for us. Arguably, however, these considerations do not pose the same problem for objective consequentialism, since in such cases the view at least *suggests* an answer to the question of what to do: namely, to perform the action that is *most likely* to maximize objective value, or to *try one's best* to do so, or something along those lines.[19] By contrast, in the case of Jill

16 Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," 466–67.

17 Note that while it is not clear what Jackson takes the expression "what to do" to mean, he probably does not accept the noncognitivist interpretation of it that is associated with Gibbard, since he defends a form of cognitivism about normative concepts elsewhere (see Jackson, *From Metaphysics to Ethics*).

18 This expression is from Lenman, "Consequentialism and Cluelessness," 344.

19 Jackson attributes a view along those lines to Peter Railton, though I note that it is unclear to me whether Railton in fact meant to endorse this view. See Jackson, "Decision-Theoretic

and John, such courses of action seem clearly objectionable. Since Jill knows that drug *A* will not be best, the action that is most likely for her to maximize objective value is perhaps to flip a coin and give John either drug *B* or drug *C*, depending on the outcome. Yet it is surely a terrible idea to make the decision in this way. Instead, intuitively, Jill ought to give drug *A* to John, even though she knows that this does not maximize objective value.

Jackson's argument accordingly supports the view that we sometimes ought to perform actions that we know to be objectively suboptimal. If that view is true, then objective consequentialism is false, since objective consequentialism entails that all suboptimal actions are impermissible. Most of the literature on Jackson's argument has thus focused on the question of whether giving John drug *A* really is what Jill ought to do.[20] But in the current context, there is another aspect of Jackson's argumentation that is more important. What ultimately underlies the argument is a widespread and natural view about the role of normative thinking in practical deliberation.[21] It is illustrated by Jackson's claim that, unlike other areas of inquiry, ethics is centrally concerned with the "passage to action":

> It is fine for a theory in physics to tell us about its central notions in a way which leaves it obscure how to move from those notions to action, for that passage can be left to something which is not physics; but the passage to action is the very business of ethics.[22]

In recent discussions about normativity, similar remarks have been frequent. For example, Jacob Ross writes that

> in genuine deliberation, we are guided, at least implicitly, by the question "What should I do?" or "What ought I to do?" And we ask this question not simply in order to satisfy our curiosity, but in order to make up our minds about what to do, that is, in order to form an intention. Thus, the role of the *ought* of practical deliberation is to guide our intentions, and thereby to guide our actions.[23]

Consequentialism and the Nearest and Dearest Objection" 466; and Railton, "Alienation, Consequentialism and the Demands of Morality."

20 See, e.g., Zimmerman, *Living with Uncertainty*.

21 Of course, that is not to say that the view is universally accepted. For opposition, see, e.g., Parfit, *On What Matters*, vol. 2; and Zimmerman, *Ignorance and Moral Obligation*. See also Weatherson, who defends a view on which answers to normative questions need not be guiding (*Normative Externalism*). This view is congenial with my conclusion that the central deliberative question does not concern what we ought to do.

22 Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," 467.

23 Ross, "Rationality, Normativity, and Commitment," 164.

In the same vein, Errol Lord claims that it is "commonly assumed that the answer to the central deliberative question is the thing that you ought to do, full stop"; David Faraci writes that "substantive normative claims answer (or at least entail that there is an answer to) the question of *what to do*"; and Jonathan Way and Daniel Whiting suggest that "in deliberation, we ask ourselves a single question, 'What ought I to do?'"[24] While the idea that all these claims suggest is perhaps somewhat imprecise, it is also highly intuitive. It is plausible that we normally do not engage in normative thinking with the sole aim of learning more about the world. We also do so in order to reach choices in our lives. And the worry that Jackson highlights is that objective consequentialism suggests that this aim is misguided. For while our lives are unavoidably full of uncertainty and ignorance about empirical facts, consequentialism entails that, unless we have knowledge of those facts, we cannot figure out what we ought to do. Perhaps that result would not be so bad if we could instead settle for the action that is most likely to be best. But what Jackson's argument suggests is that we sometimes ought not even to do that—rather, in some situations, we ought to perform actions we know to be objectively suboptimal. And the worry is that, in view of all this, it is hard to see how consequentialism can be reconciled with the role of normative thinking in practical deliberation that Jackson and Ross suggest. Even if objective consequentialism were true and we knew that this was so, our uncertainty about the question of what to do would remain unresolved—and yet this is the very question, the thought goes, that consequentialism and other normative theories seek to answer.

In what follows, I will summarize the above claims by saying that objective consequentialism fails to *address* the central deliberative question for agents who, like Jill, lack the relevant empirical knowledge. If Jackson's and Ross's idea is correct, the fact that consequentialism fails to do so is a serious problem for the view.

24 See Lord, "What You're Rationally Required to Do and What You Ought to Do (Are the Same Thing!)," 1110; Faraci, "On Leaving Room for Doubt," 248; and Way and Whiting, "Perspectivism and the Argument from Guidance," 362. More generally along the same lines, Mark Timmons holds that normative ethics has both a "practical" and a "theoretical" aim (*Moral Theory*, ch. 1), and Michael Smith claims that a metaethical theory must be able to accommodate both the "objectivity" and the "practicality" of moral judgments (*The Moral Problem*, ch. 1). For a recent discussion about using morality as a decision guide under empirical uncertainty, see Holly Smith (*Making Morality Work*). However, Holly Smith does not focus on the issue of fundamental moral uncertainty (or of fundamental normative uncertainty more generally), which will be central in what follows. For an argument that Holly Smith's idea that moral theories should be practically "usable" leads to a noncognitivist view like the one that I endorse in this paper, see Clarke-Doane, "From Non-usability to Non-factualism."

3. DECISION-THEORETIC CONSEQUENTIALISM

In view of the problems for objective consequentialism, Jackson instead endorses "decision-theoretic consequentialism," whose main motivation is its purported ability to avoid those problems. According to decision-theoretic consequentialism, every agent ought to maximize "expected moral utility," where an action's expected moral utility is determined (roughly) by summing the probability-weighed values of its possible outcomes.[25] Notably, while the values in question are meant to be objective, Jackson takes the relevant probabilities to be the agent's subjective ones. Thus, in one respect, the view resembles classical decision theory, as the agent's own mental states partially determine what she ought to do, and in this respect it also differs from objective consequentialism.[26] In another respect, however, Jackson's view resembles objective consequentialism and differs from classical decision theory, as the agent's preferences are not taken to determine the relevant ordering of an action's possible outcomes—instead, that ordering is determined by the objective value facts, whatever they turn out to be.

Decision-theoretic consequentialism thus involves a combination of objective and subjective elements that is striking and seemingly unstable. Indeed, as a result, this view is susceptible to the very same problem that Jackson takes objective consequentialism to face: that it fails to address the central deliberative question for an agent to whom "the fact that an action has in fact the best consequences [is] obscure" (cf. section 1). For decision-theoretic consequentialism also centrally appeals to facts that are often "obscure" to us—namely, the objective value facts about the possible outcomes of our actions.[27] We are often uncertain about what is objectively good, and even when we are not, our views are often mistaken. In particular, the widespread disagreement about value illustrates this point: so many people have conflicting axiological views that, at best, only a few of us can be correct.[28] Hence, while Jackson is right that we

25 See Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," 464.

26 Strictly speaking, classical decision theory states the conditions for representing an individual's preferences with a particular representation function. Decision-theoretic consequentialism adopts the constraints of decision theory on preference orderings for use in a normative consequentialist theory. Thanks to Andrew Reisner for discussion.

27 For similar worries, see M. Smith, "Moore on the Right, the Good, and Uncertainty"; and Bykvist, "How to Do Wrong Knowingly and Get Away with It."

28 Moreover, on many plausible views about the epistemology of disagreement, a subject's true axiological beliefs often or always fail to amount to *knowledge* when they are disputed (at least when the opponents are the subject's epistemic "peers"); see further, e.g., McGrath, "Moral Disagreement and Moral Expertise"; and Risberg and Tersman, "A New

often do not know what action maximizes objective value, he overlooks the fact that we often do not know what action maximizes expected moral utility either.

One might think that the problem is less serious for decision-theoretic consequentialism if the relevant value facts can be known *a priori*, at least in principle. The problem with this suggestion is that the mere *possibility* of axiological knowledge makes no difference to an agent when she does not in fact have it. The case of Jill and John illustrates this point: while Jill does not know whether drug *B* or drug *C* is the perfect cure, it is perfectly possible for her to acquire such knowledge—all she has to do is give John one of the risky drugs and observe the results. Clearly, however, the principled possibility of such knowledge is useless to her when she does not in fact have it. And the point is that the *way* in which she could acquire such knowledge is in this regard irrelevant. Merely possible knowledge, whether *a priori* or otherwise, cannot help us in our decision-making.

4. Choices under Normative Uncertainty

Decision-making under axiological uncertainty is a special case of decision-making under normative uncertainty more generally. We sometimes face hard choices, not because we are uncertain about the relevant empirical or axiological facts, but because we are uncertain about the fundamental normative facts. For example, many people must at some point decide whether to have children. When we try to figure out what we ought to do in such situations, we face difficult problems about our obligations toward future people, present people, merely possible people, and so on.[29] Perhaps we sometimes solve those problems. Very often, however, we fail to do so. For instance, maybe it is just too hard for us to determine whether we are obliged to create a person with a good life rather than a person whose life would be worse but nonetheless worth living, or no person at all. To answer that question, we must take a stance on the many controversial issues in population ethics. Due to their difficulty, some degree of uncertainty about their answers, or perhaps even suspension of judgment, seems to be warranted. Yet even somebody who is uncertain about those questions might one day have to decide whether to become a parent. She cannot wait until the true moral theory has been discovered, since she has to act now. Thus, she will have to deliberate about what to do, even though she has failed to determine what she ought to do.

The recent debate about choices under normative uncertainty has primarily been motivated by the aim of providing some sort of guidance in such

Route from Moral Disagreement to Moral Skepticism," "Disagreement, Defeat, and Higher-Order Evidence," and "Moral Realism and the Argument from Skepticism."

29 For two classic discussions of these problems, see Parfit, *Reasons and Persons*, pt. 4; and Arrhenius, *Future Generations*.

situations.[30] The hope is that, even when we are uncertain about fundamental normative matters, there is a form of normative theorizing that can help us reach actions or decisions. While many different theories about these issues have been proposed, their details need not concern us here.[31] I will instead focus on the question that theories about choices under normative uncertainty are supposed to answer.

Participants to this debate often introduce their topic by noting, as I did above (cf. the introduction), that situations of normative uncertainty make it natural to ask questions like: "I can't figure out what I ought to do; *now* what ought I to do?" However, they then usually note (as I also did above) that it is not clear how this question should be understood. After all, most traditional moral theories, like utilitarianism and Kantianism, entail that we ought to maximize happiness or treat humanity as an end in itself (etc.) whether or not we believe this to be the case. Thus, on those views, what we ought to do is simply independent of our beliefs about the matter. This has led some to think that the question just posed has a trivial answer: we simply ought to do what the true normative theory entails that we ought to do, whether or not we know what that is.[32] On the one hand, this claim seems close to platitudinous and thus hard to deny. On the other hand, however, there is an obvious sense in which this answer fails to address the agent's uncertainty in the situation just considered, just like objective consequentialism fails to address Jill's uncertainty in the case of Jill and John.

30 Michael Zimmerman's work on this topic is an exception. See, e.g., Zimmerman, *Living with Uncertainty*.

31 The currently most popular view is that normatively uncertain agents ought to maximize "expected choiceworthiness." Unlike Jackson's concept of expected moral utility, the concept of expected choiceworthiness is supposed to be sensitive both to the agent's normative probabilities and her nonnormative probabilities. The viability of this strategy is the subject of an ongoing debate. One major concern is that it requires that "inter-theoretical" comparisons of choiceworthiness are meaningful. In other words, the degree to which an action is right according to utilitarianism must be comparable to the degree to which it is wrong according to Kantianism, for example, as its expected choiceworthiness is supposed to be the probability-weighed sum of those values. It is still unclear when, if ever, such comparisons are meaningful; in particular, as William MacAskill notes, the matter is especially complicated for agents who have some (justified) degree of belief in nihilism, on which the moral value of every action is not zero but undefined (see MacAskill, "The Infectiousness of Nihilism"). For an overview of the debate and a discussion of how the strategy of maximizing expected choiceworthiness can be expanded to handle cases that involve incomparability, see MacAskill, Bykvist, and Ord, *Moral Uncertainty*. For an argument that moral uncertainty is not normatively important, see Harman, "The Irrelevance of Moral Uncertainty."

32 See, e.g., Weatherson, Review of *Moral Uncertainty and Its Consequences* and *Normative Externalism*.

To avoid this result, the most popular strategy has been to hold that in the question "What ought I to do when I don't know what I ought to do?" the different occurrences of "ought" have different meanings. Following Andrew Sepielli, let us call this the *dividers'* strategy.[33] Dividers usually take the first occurrence of "ought" in this question to stand for the "objective" ought. Traditional first-order theories like utilitarianism and Kantianism concern what we ought to do in this sense. The second occurrence of "ought," by contrast, is supposed to stand for something that is not the concern of such theories. It is less clear what that ought is like, however, because dividers disagree about how many oughts there are. Some dividers stop at two—on that view, there is just an objective ought and a "subjective" ought and that is that.[34] This has been a minority view, however, and many dividers instead posit a much larger number of oughts. In part, this is due to the fact that an important argument for the dividers' view relies on the idea that seemingly incompatible "ought" sentences can be jointly true relative to different states of information.[35] Since there are clearly many different states of information, dividers are pushed toward positing many different oughts as well. For example, Andrew Sepielli writes that:

> [We] may speak of the belief-relative sense of "ought," the reasonable-belief-relative sense, the degree-of-belief-relative (or credence-relative, or subjective-probability-relative) sense, the evidence-relative sense, and the objective-probability-relative sense, each of which depends for its proper application on the feature mentioned in its label. We could ramify even further. There are, for example, different "interpretations" of objective probability—the long-run frequency interpretation, the propensity interpretation, the logical interpretation, etc.—and there could be an OUGHT corresponding to each interpretation. Finally, there is a subjective OUGHT that I call the minimal-probability-relative OUGHT.[36]

33 Sepielli, "Subjective and Objective Reasons." Sepielli adopts this terminology to distinguish between "dividers" and "debaters" about the question of how we ought to act under uncertainty. For present purposes, we need not consider what semantics for "ought" that dividers should adopt. While it has sometimes been said that "ought" is genuinely ambiguous, like "bat" or "bank," a more plausible view is that the lexical entry for "ought" has an informational parameter that is supplied by context.

34 Harman, "The Irrelevance of Moral Uncertainty," and Parfit, "What We Together Do," both seem to endorse this view (though they also seem to endorse different theories about what we subjectively ought to do).

35 For discussion of this idea, see Kolodny and MacFarlane, "Ifs and Oughts."

36 Sepielli, "What to Do When You Don't Know What to Do," 48. Sepielli uses capital letters to denote concepts, but he also notes that his idea does not strictly require that there are many distinct ought concepts.

While the topic of normative uncertainty does not figure in Jackson's discussion, he nonetheless anticipates the dividers' strategy by positing what he considers "an annoying profusion of 'oughts'":

> I think that we have no alternative but to recognize a whole range of oughts—what [Jill] ought to do by the light of her beliefs at the time of action, what she ought to do by the lights of what she later establishes (a retrospective ought, as it is sometimes put), what she ought to do by the lights of one or another onlooker who has different information on the subject, and, what is more, what she ought to do by God's lights.[37]

The idea is that by God's lights Jill ought to give John the perfect cure, but by her own lights she ought to give him the safe cure. And dividers seek to make sense of the normative uncertainty debate in a similar way: objectively, they think, we ought to satisfy the true first-order normative theory, but when we cannot determine what we objectively ought to do, we can at least try to determine what we ought to do in some other sense of "ought." It is the latter, nonobjective kind of ought that the normative uncertainty debate is taken to concern.

Clearly, regress threatens. While it is true that the traditional, "objective" questions of normative ethics are sometimes hard, the current controversies in the normative uncertainty debate suggest that those questions are not easier.[38] If we cannot figure out what we ought to do in the sense of "ought" that is central to that debate, are we then supposed to try to figure out what we ought to do in yet a new sense of "ought"? But why should we expect that to be easier? Does this ever stop?[39]

However, while the regress problem is important, in what follows I will focus on another problem that (for reasons that will emerge) I take to be more fundamental. The problem concerns the apparent stalemate that arises between all the oughts that dividers posit. Recall that theories about normative uncertainty are supposed to provide some sort of guidance to agents like the potential parent, who must decide whether to have children. The idea is to

37 Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," 471–72.

38 For a convincing argument that maximizing expected value is normally not significantly easier than maximizing objective value, see Feldman, "Actual Utility, the Objection from Impracticality, and the Move to Expected Utility."

39 For further discussion of the regress problem, see Sepielli, who seeks to solve it by distinguishing between "perspectival" and "systematic" notions of rationality and between different "orders" of rationality ("What to Do When You Don't Know What to Do When You Don't Know What to Do . . ."). The discussion in section 5 will indicate why I find this solution unconvincing.

posit many different oughts to make sense of the question that such agents may naturally ask. In the relevant cases, however, these oughts will often prescribe different action—otherwise figuring out what we nonobjectively ought to do would be just as hard as figuring out what we objectively ought to do (since those questions would simply have the same answer). And the existence of such conflicts seems only to give rise to the central deliberative question once again, for we may now also be uncertain about which of all these oughts to satisfy. What should we do when they diverge? Is there any genuine sense in which one of them can be said to be privileged, or more important than the others?

5. The Tie-Breaking Problem

Michael Zimmerman presents the relevant worry when commenting on a variation of the case of Jill and John:

> [Jill] seeks your advice, telling you that she believes that drug *B* would be best for John but that she isn't sure of this. "So," she says, "what ought I to do?" You are very well informed. You know that *A* would be best for John, that Jill believes that *B* would be best for him, and that the evidence available to Jill (evidence of which she is apparently not fully availing herself, since her belief does not comport with it) indicates that *C* would be best for him. You therefore reply, "Well, Jill, objectively you ought to give John drug *A*, subjectively you ought to give him *B*, and prospectively you ought to give him *C*." This is of no help to Jill. It is not the sort of answer she's looking for. She replies, "You're prevaricating. Which of the 'oughts' that you've mentioned is the one that *really* counts? Which 'ought' *ought* I to act on? I want to know which drug I am morally obligated to give John, *period*. Is it *A*, *B*, or *C*? It can only be one of them. It can't be all three."[40]

Of course, Jill's questions here are imprecise. If there are many different oughts, she cannot make progress by asking which ought she really ought to act on. For it is *really the case* that she objectively ought to act on the objective ought. The problem is that it is equally the case that she subjectively ought to act on the subjective ought. Imprecision aside, however, there is surely *some* important, nontrivial form of uncertainty that Jill is trying to express here. It is very similar to what the prospective parent tried to ask about the choice of whether to have children. Jill must decide which drug to give to John, but what she is told in the dialogue above does not take her closer to action. (Compare: being

40 Zimmerman, *Living with Uncertainty*, 7.

told that we ought to do what is objectively best or objectively right similarly fails to take us closer to action when we do not know what is objectively best or objectively right.) In relation to the aim of providing guidance to uncertain agents, this result is a disaster.

Jackson anticipates this problem too. In an attempt to avoid it, he stipulates that by "ought" he means "the ought most immediately relevant to action, the ought which I urged it to be the primary business of an ethical theory to deliver."[41] However, whether this stipulation solves the problem depends entirely on what the word "relevant" is supposed to mean. Since Jackson does not say, let us consider some possibilities.

Jackson probably did not intend his ought to be relevant in some merely descriptive sense of "relevant." The reason is that descriptive facts do nothing to address Jill's uncertainty in the dialogue above. For example, perhaps one of the oughts is the one that we *in fact* tend to focus on in deliberation. In Jill's situation, however, this is surely beside the point—for whatever it is that her uncertainty concerns, it is not which ought we *do* tend to satisfy. Rather, as her questions suggest, it is something closer to the question of what ought she *ought* to satisfy.

For this reason, it is natural to think that Jackson rather intended his ought to be relevant in some *normative* sense of "relevant." While Jackson does not elaborate on this point, he would perhaps agree with Mark Schroeder that there is an "important *deliberative* sense of 'ought,' which is the central subject of moral inquiry about what we ought to do and why."[42] Schroeder mentions several features that, in his view, distinguish this ought from others. The currently most important feature is that the deliberative ought, according to Schroeder, is "the right kind of thing to *close deliberation.*"[43] This seems congenial to what Jackson has in mind.

Importantly, however, to insist that the deliberative ought is the right kind of thing to close deliberation serves in this context only to relocate the problem.[44] For, if there are very many senses of "ought," why should there not also

41 Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," 472.

42 Schroeder, "Ought, Agents, and Actions," 2. Similarly, Lord writes that an agent "doesn't seem to learn what she ought to do, full stop, by learning what she subjectively and objectively ought to do. There is another question that hasn't been answered yet: what ought [she] to do? When we theorize about what answers this question, we theorize about the deliberative ought" ("What You're Rationally Required to Do and What You Ought to Do (Are the Same Thing!)," 1138).

43 Schroeder, "Ought, Agents, and Actions," 9.

44 Of course, Schroeder's characterization of the deliberative ought may still be apt for his own purposes, which is to distinguish the deliberative ought from the "evaluative" ought that is involved when we claim, e.g., that there ought to be world peace.

be very many senses of "right"? In particular, even if it is subjectively right to close deliberation using the subjective ought, it is presumably objectively right to close deliberation using the objective ought. Similarly, and more directly in Jackson's terms, even if his ought is subjectively normatively relevant, it seems undeniable that the question of what Jill objectively ought to do is objectively normatively relevant—normatively relevant, that is, by God's lights.

The upshot is that the stalemate that arose among the different kinds of oughts, and that the appeal to normative relevance was supposed to get rid of, now arises among the different kinds of normative relevance instead. This is an instance of what I will call the "tie-breaking problem." The problem is that if we believe that the deliberative uncertainty that I have highlighted concerns a normative question, it is problematic to think that the normative realm is divided into distinct "domains" or "spheres." For to answer the question that this form of uncertainty concerns, we must somehow single out *one* action as the one to be performed. And since distinct oughts normally prescribe different actions, we face the question of which of these oughts to satisfy. For somebody who is uncertain about this question, it is useless to learn trivialities such as that she objectively ought to satisfy the objective ought, or that she subjectively ought to satisfy the subjective ought.

We might attempt to avoid this stalemate by appealing to some tie-breaking notion that distinguishes one ought from all the others. Zimmerman's Jill tries to do that by asking which ought she *ought* to satisfy. Jackson instead suggests that one of the oughts is "most relevant to action." But all such proposals face a dilemma. On the one hand, if the notion that plays the tie-breaking role is descriptive, then it is beside the point. It is simply plain that Jill's uncertainty does not merely concern whether the objective ought or the subjective ought has a certain descriptive feature. On the other hand, if the tie-breaking notion is normative, then we should expect it to be just as divided as the other normative notions. Thus, rather than breaking the tie, this move only reinstates the stalemate that we faced among the diverging oughts.[45]

45 Peter Graham argues that moral obligations are objective (in the sense, roughly, that they are independent of our evidence) on the ground that it is the objective moral sense of "ought" that concerns a morally conscientious person ("In Defense of Objectivism about Moral Obligation"). However, Graham also holds that, in Jackson's case of Jill and John, Jill is morally conscientious only if she does something (i.e., giving John drug *A*) that she knows that she ought objectively not to do. This makes it hard to avoid the conclusion that there is in fact another kind of ought that tracks what a morally conscientious person *does* (where this may depart from the ought she is *concerned with*), which is (in some sense) the one that we *really* ought to satisfy. At any rate, since Graham assumes that a "morally conscientious person is solely concerned with her moral obligations" ("In Defense of Objectivism about Moral Obligation," 98), his suggestion sheds no light on cases in which moral requirements conflict with other

This also puts us in a position to see why the tie-breaking problem is more fundamental than the regress problem (cf. section 3). Recall that a regress of oughts threatens when we are uncertain about what we ought to do in the sense of "ought" that is supposed to be central to the normative uncertainty debate. It is normally assumed that it would be problematic to simply bite the bullet and accept that such a regress does indeed arise. But why? Regresses are not *intrinsically* problematic; for example, we can all agree that if it is true that *p*, then it is also true that it is true that *p*, and true that it is true that it is true that *p*, and so on. So why would it be so bad to accept a regress in this particular case?

I suspect that the regress of oughts seems problematic only given the further assumption that of all the oughts that dividers posit, one of them is supposed to be the ought that addresses the central deliberative question in cases of normative uncertainty. And it is this assumption that the tie-breaking problem calls into question. *If* that assumption is accepted, then the regress of oughts is a problem because it suggests that we might be ignorant of what we ought to do at each point of the regress. We could try to figure out what we ought to do, fail to do so, move on to figure out what we ought to do in some other sense of "ought," fail again, and so on. Far from being guiding, this process would never result in action. But the tie-breaking problem calls the crucial assumption into question at an earlier stage, before worries about regress even arise.

In other words, while the regress problem is an important epistemological worry for dividers, the tie-breaking problem is a conceptual worry that is prior to it. That problem is to make sense of the question of which ought we *really* ought to act on, as Zimmerman's Jill puts it, rather than the epistemic problem of whether we can know what we ought to do, for some given sense of "ought."

6. THE NORMATIVE QUESTION

So far, I have discussed a number of problems concerning whether normative theories can guide us in choice situations that involve different forms of uncertainty. However, a possible reaction to the discussion so far is to hold that if a normative theory fails to address the deliberative uncertainty of an agent who lacks relevant information, it is (so to speak) the agent and not the theory that is to blame.[46] The idea is that we can acknowledge that normative theories

kinds of normative requirements, which I discuss in section 6. For further critical discussion of Graham's view, see Mason, "Objectivism and Prospectivism about Rightness," sec. 4c.

46 For this suggestion, see, e.g., Bykvist, "Violations of Normative Invariance," 113. In this vein, both Krister Bykvist and Erik Carlson hold that moral theories must be practically useful for "ideal" agents only (see Bykvist, "Violations of Normative Invariance"; and Carlson, "Deliberation, Foreknowledge, and Morality as a Guide to Action"). However, I think we

cannot address the central deliberative question for every agent, no matter their epistemic situation, but insist that they should at least address that question for agents who know all the relevant truths.

A problem for this proposal is that considerations that relate to different states of information are not the only possible reason to posit many oughts. Another possible reason is that different "sources" of normativity, such as morality and prudence, may generate distinct normative requirements.[47] If that is so, the tie-breaking problem arises again when those requirements cannot be jointly satisfied. To illustrate, suppose that prudence requires you to maximize your own well-being while morality requires you to sacrifice yourself for the sake of others. When you ask for advice, you are told only that prudentially you ought to be selfish, but morally you ought to be altruistic. As I suggested above (in the introduction), questions like those from Zimmerman's Jill are natural here too: "Which of the 'oughts' that you have mentioned is the one that *really* counts? Which 'ought' *ought* I to act on?"[48] There is a further salient question about which both you and Jill are uncertain. But once again, it is difficult to argue that your uncertainty literally concerns whether you ought to act morally or prudentially. For, again, it is *really the case* that morally you ought to act morally. The problem is that it is equally the case that prudentially you ought to act prudentially.

The challenge of understanding the salient further question arises particularly clearly in the debate about the "normative question," which is associated with Christine Korsgaard. She formulates this question as follows:

> When we seek a philosophical foundation for morality . . . we are asking what *justifies* the claims that morality makes on us. This is what I am calling "the normative question."[49]

---

should generally be suspicious about appealing to idealized agents in normative theorizing; see further Risberg, "Weighting Surprise Parties" and "The Entanglement Problem and Idealization in Moral Philosophy."

47 There are many possible views about the structure of normative conflicts, however, and not everyone agrees that there are genuinely distinct sources of normativity (for discussion, see Reisner, "Normative Conflicts and the Structure of Normativity"). Philosophers who disagree often hold that there is ultimately only one kind of normative question, such as what we *all things considered* ought to do (for suggestions along these lines, see, e.g., Crisp, *Reasons and the Good*; and Tännsjö, *From Reasons to Norms*). I will return to this suggestion shortly.

48 Interestingly, Zimmerman elsewhere suggests that there is no comprehensible question concerning what one "really" ought to do when the moral ought conflicts with a nonmoral ought (*The Concept of Moral Obligation*, 1–2). He thus seems to take conflicting oughts that are due to different sources of normativity to be less problematic than conflicting oughts that are relative to different states of information. In light of the obvious similarities between the two problems, however, this strikes me as an unattractive view.

49 Korsgaard, *The Sources of Normativity*, 9–10.

Korsgaard continues to write that an answer to this question

> must actually succeed in *addressing* someone in [the "first-person" position from which the normative question is asked]. It must not merely specify what we might say, in the third person, *about* an agent who challenges or ignores the existence of moral claims. Every moral theory defines its concepts in a way that allows us to say something negative about people who do that—say, that they are amoral or bad. But an agent who doubts whether he must really do what morality says also doubts whether it's so bad to be morally bad, so the bare possibility of this sort of criticism settles nothing.[50]

While Korsgaard's reasoning here is supposed to present a problem for moral realism, and for normative realism more generally, there has been a lot of confusion about what the problem is supposed to be.[51] For what question, more exactly, is it that the relevant sort of criticism fails to settle? Surely it is not literally whether moral claims are morally justified or whether it is morally bad to be morally bad. In this vein, Derek Parfit writes:

> According to what Korsgaard calls normative realism, when we know the relevant facts, we are rational if we want, and do, what we have decisive reasons to want, and do. So Korsgaard seems here to suggest that, if realism were true, we might need a reason to want, and do, what we knew that we had decisive reasons to want, and do. That is clearly false. If you should do something, it is not an open question whether you should do it.[52]

While Parfit's claims are undeniable as far as they go, it would be surprising if they were to settle the doubts of the agent that Korsgaard has in mind. For although the question of whether it is so bad to be morally bad is imprecise, there does seem to be an important question that the agent is trying to express. Insofar as that question concerns a nontrivial issue, as it seems to do, it cannot be answered by the trivial facts that Parfit notes.

50 Korsgaard, *The Sources of Normativity*, 16.

51 Dreier helpfully identifies some misunderstandings in the debate ("Can Reasons Fundamentalism Answer the Normative Question?"). However, for reasons that I will present in sections 7–8, I do not share Dreier's view that the problem is that realists cannot explain why it is irrational to act contrary to one's normative judgments.

52 Parfit, *On What Matters*, 2:418. Note that Parfit assumes the controversial view that it is always rational for us to do and want what we have most reason to do and want. In particular, if there are "state-given" reasons for attitudes, we might sometimes have most reason to be irrational; for further discussion, see Reisner, "Is There Reason to Be Theoretically Rational?"

According to a popular view, Korsgaard's question about morality is best understood in terms of a normative concept that is not "indexed" to any particular source of normativity. This concept has variously been suggested to concern either reasons, rationality, correctness, the "favoring-relation," or a special kind of ought (which among other things has been called the "all things considered ought," the "ought full stop," the "ought period," and the "ought *simpliciter*"). For present purposes it does not matter which of those concepts we invoke, so I will focus on the concept ALL THINGS CONSIDERED OUGHT (though I will sometimes omit the "all things considered" qualifier in what follows). The idea is that when an agent faces conflicting normative requirements, like moral and prudential ones, she may acknowledge both that she morally ought to perform a certain action and that she prudentially ought to perform some other action. The salient further question is then what she all things considered ought to do. The all-things-considered ought is supposed to be the tie breaker that resolves her uncertainty.

While it has sometimes been doubted whether the concept ALL THINGS CONSIDERED OUGHT is comprehensible, I will here set this worry aside.[53] Instead, in the next section, I will argue that even if there is such a special ought, the deliberative question that I have highlighted does not concern it. The reason is that even facts about what we all-things-considered ought to do can in principle be subjected to a certain form of practical questioning. That such questioning is comprehensible even when all normative questions are settled shows, I believe, that the central deliberative question is not a special normative question.

7. OUTRAGEOUS NORMATIVE TRUTHS

Many seemingly trivial questions have figured in the discussion so far: whether it is morally bad to be morally bad, for example, and which ought we ought to satisfy. Another trivial question is whether the normative truth will turn out to be normatively outrageous. Of course it will not! However, a nontrivial question in the same neighborhood is whether the normative truth will turn out to outrage *us*, in the purely descriptive sense of striking us as outrageous. That this is at least conceptually possible follows from the commonly accepted view that the thinnest normative concepts, like ALL THINGS CONSIDERED OUGHT, do not have

53 For such doubts, see, e.g., Copp, "The Ring of Gyges"; Tiffany, "Deflationary Normative Pluralism"; and Baker, "Skepticism about Ought Simpliciter." Perhaps Sidgwick's "dualism of practical reason" should also be understood as a version of skepticism about the all things considered ought (see Sidgwick, *Methods of Ethics*). However, another understanding of Sidgwick's view is that while the concept ALL THINGS CONSIDERED OUGHT is itself comprehensible, it is simply not satisfied by any action when morality and prudence conflict.

enough descriptive content to adjudicate between various competing views about first-order normative questions.[54] This view is supported by versions of G. E. Moore's open-question argument, Hume's law, the "is/ought-gap," and several related ideas. For example, even if consequentialism is true—and even true by metaphysical necessity—it is at least conceptually possible that a staunch absolutist theory is true, on which the consequences of our actions are irrelevant to their normative status. On such a theory, what is normatively important is not whether an action has a good outcome. All that matters is that it does not violate a certain set of rules. On this view, it is always forbidden to lie, for instance, no matter the consequences of telling the truth. Indeed, this view is often attributed to Kant.

While the staunch theory about lying is probably not true, it is nonetheless possible for us to reason under the hypothesis that it is true. For instance, it is clear that, given the truth of the staunch theory, the consequences of our actions are normatively irrelevant, and thus, most of us are seriously mistaken about ethics. We can confidently accept such conditionals while rejecting their antecedents. Similarly, it is clearly not the case that, given that the staunch theory is true, the staunch theory is false, so what our intuitions are tracking here is not just the trivial fact that a material implication is true if its antecedent is false. Rather, even when we know that *p* is false, we may nontrivially evaluate claims of the form "given *p*, then *q*." (Or, in the jargon: even when we know that *p* is false, we can still "conditionalize" on *p*.) We may also be uncertain about whether to accept such claims, in the sense that we may be uncertain about whether to accept the consequent given the truth of the antecedent. I will now argue that, in a similar way, we may remain uncertain about the deliberative question even given that all the normative questions are settled.

The argument relies on the following thought experiment. You face a choice situation where you can prevent great suffering by telling a lie. By telling the truth, on the other hand, you will cause even more suffering. Suppose now that it is true that, consequences notwithstanding, you are forbidden to tell the lie. In other words, morally, all things considered, and so on, you ought to tell the truth. Contrary to what you used to think, it has turned out to be normatively

54 What I mean by this, roughly, is that competence with such concepts is not sufficient for knowing which first-order normative theory is true. Note also that what I say here is compatible with the idea of "moral fixed points" that Cuneo and Shafer-Landau endorse ("The Moral Fixed Points"). The reason is that this idea pertains specifically to *moral* concepts, rather than to *normative* concepts in the more inclusive sense, and one of the consequences of this idea is precisely that moral concepts are much "thicker" than what is ordinarily supposed (cf. Cuneo and Shafer-Landau, "The Moral Fixed Points," 406). Indeed, as Cuneo and Shafer-Landau note, the idea of moral fixed points is not supposed to help with the question that arises when morality conflicts with some other source of normativity, such as prudence, or perhaps "shmorality" ("The Moral Fixed Points," 406–7).

irrelevant that you could prevent great suffering by acting otherwise. (To be clear, what I want to imagine is not merely that someone *tells you* that you ought to tell the truth, or that you receive some other type of evidence for that claim; I want to imagine that *it is the case that* you ought to tell the truth. Again, this assumption is surely coherent, even if it is false of metaphysical necessity.)

In this case, at least three reactions are possible. The first is to "go with" the normative truth even though it is outrageous. "If that is what I ought to do," you could say, "then it is also what I shall do," hence proceeding to tell the truth. The second possibility is simply to give up on the commitment to doing what you ought to do. "If *that* is what I ought to do," you might say, "then I shall instead do what I ought not to do," thus going on to lie. The point is not that you may conclude that the ethical truth has turned out to be unethical (or that the normative truth has turned out to be "unnormative")—that remains an incoherent view. The point is rather that you may turn your back on the ethical truth, so to speak, because the trivial fact that the ethical truth is ethical might strike you as no more significant than the fact that immoral actions are legally required in countries whose laws are also immoral. Finally, the third possibility is to remain deliberatively uncertain. If you learn that you ought to cause great suffering, you might try to question ethics itself—"I ought to do something that strikes me as outrageous; *now* what ought I to do?" But it is now clear that this question does not literally express what your uncertainty concerns. You *know* what you ought to do—morally, all things considered, and so on. This is stipulated. Even so, you might remain uncertain about the deliberative question.

It does not matter what reaction we are in fact disposed to have. What matters is just the first reaction is not the only comprehensible one. The possibility of the other reactions shows that the deliberative question is not settled even by the assumption that all the truths, including all the normative truths, are known. On the view that I will go on to suggest in the next section, this is also the type of uncertainty that is made salient, in different ways, by the ethical and metaethical debates I have considered above. However, an underlying assumption in those debates is that this type of uncertainty must concern a special, puzzling normative question that is difficult to express: what we all-things-considered ought to do, for instance, or what we ought to do in a sense of "ought" that is relevant when we do not know what we objectively ought to do. In view of the argument just presented, I believe that we should reject this assumption.

The argument just presented can be helpfully contrasted with two related ones from the literature. First, Clarke-Doane supports his noncognitivist view of the "further question" by appeal to an argument that involves conditionalizing on what he calls "evaluative pluralism," which is roughly the view that there are alternative normative concepts in the sense characterized earlier (cf. section

1).[55] In short, the idea is that under the assumption that we ought to perform some action, *A*, but also ought* not to perform *A*, it seems that we can remain deliberatively uncertain about whether to perform *A*. While I am sympathetic to Clarke-Doane's argument, an important difference is that mine does not involve conditionalizing on pluralism but on the first-order normative claim that we always ought not to lie. This is an advantage since not everyone finds the relevant form of pluralism even intelligible. For instance, William Fitzpatrick writes that "there are no intelligible alternative notions of 'value*' or 'shmalue,' or 'good*' or 'appropriate*.' . . . We shouldn't rush to think we have the foggiest idea what such things would even mean."[56] If Fitzpatrick is right, it is not clear that we can even coherently conditionalize on pluralism. By contrast, as I have emphasized, the view that we should never lie is perfectly comprehensible (albeit implausible).[57]

Another interesting argument has been offered by Matthew Bedke in a critique of metaethical nonnaturalists (roughly, those who think that normative facts are mind independent and different in kind from those that are studied by the sciences).[58] Simplifying somewhat, Bedke's central claim is that nonnaturalists are committed to revising their moral beliefs in immoral ways. He asks us to imagine being told by a reliable oracle that there is no nonnatural property that human pain and nonhuman pain have in common. If we are nonnaturalists and trust the oracle, we are forced to conclude that human pain and nonhuman pain are not both intrinsically bad (at least insofar as we do not abandon our nonnaturalism), since nonnaturalism implies that intrinsic badness is a nonnatural property. And, according to Bedke, being disposed to revise one's moral views on the basis of such "nonnatural information" is morally objectionable. The merits of Bedke's argument need not concern us here, but three differences between his argument and mine are worth noting.[59] First, Bedke focuses on a case in which the information we receive is formulated in nonnormative terms.

55 Clarke-Doane, *Morality and Mathematics*, 167–68.

56 Fitzpatrick, Commentary on Matti Eklund, *Choosing Normative Concepts*, 6.

57 Clarke-Doane claims that it is "hard to see" how pluralism, understood as a metaphysical thesis about normative properties, "could be false," and even that it is "almost trivial" (*Morality and Mathematics*, 166, 163, 175). For criticism of these claims, see Eklund, "The Normative Pluriverse," sec. 3. In particular, as Eklund emphasizes, it is highly nontrivial that the plurality of normative properties are all *instantiable*—especially given a nonnaturalist view of their nature. In more recent work, Clarke-Doane writes that "since properties' identity conditions entail instantiation conditions, there is no doubt about [nonnatural normative] properties being instantiated if they exist" ("From Non-usability to Non-factualism," n12). However, this is also too quick, since it is still a nontrivial question whether the relevant instantiation conditions are satisfiable.

58 Bedke, "A Dilemma for Non-naturalists."

59 For critical discussion of Bedke's argument, see Enoch, "Thanks, We're Good."

He does not discuss a case in which we learn that human pain and nonhuman pain have no *normative* property in common, which would be closer to the case discussed here. With respect to such a case, the argument I presented above can plausibly be run again, since in such a case we may well be deliberatively uncertain about, e.g., whether to care more about human pain than about nonhuman pain—but that is not Bedke's point. Second, while Bedke's argument involves imagining that we *receive evidence* that a certain (nonnatural) claim is true, my argument does not have that epistemic aspect—as emphasized above, it focuses on imagining that a certain (normative) claim is *in fact* true.

Third and finally, whereas Bedke's argument targets nonnaturalists specifically, my argument succeeds also given other views about the nature of normative truths. For example, consider the version of constructivism on which normative beliefs are not "fully representational" in the sense that they seek to represent robust, mind-independent facts, but are instead true just in case (and because) they accord with the rules or procedures that are "constitutive of agency."[60] Surely, the assumption that those rules always require us to tell the truth is at least intelligible—indeed, on some interpretations, it is an assumption that Kant in fact accepted. Thus, we can imagine facing a choice situation where we can prevent great suffering by telling a lie, but in which the rules that are constitutive of agency require us to tell the truth. In such a case, we might still be uncertain about the central deliberative question (or, indeed, even disposed to answer it in a way that the rules of agency forbid). This is so independently of whether normative truths are construed as nonnatural or otherwise mind independent.[61]

8. A Noncognitivist View of Deliberative Uncertainty

If not even the all-things-considered ought is what "settles" the central deliberative question, then what does this question concern? In this section, I will present the account that I favor.

60 This is one way to understand the constructivist view of Korsgaard, *The Sources of Normativity*. As Gibbard notes, Korsgaard can also be read as a noncognitivist (Gibbard, "Morality as Consistency in Living"). On that interpretation, my argument does not clearly work against her view (and is not meant to do so), since it is not clear what it means to conditionalize on a normative claim if such claims express noncognitive attitudes. However, I take it that most constructivists want to distance themselves from noncognitivism; see, e.g., Skorupski, *The Domain of Reason*, 4.

61 Further, as Enoch has emphasized, the status of the rules that are constitutive of agency can also be challenged directly, since we can wonder whether to be an agent rather than a "shmagent" (where a shmagent is an agent-like creature that is governed by different constitutive rules) ("Agency, Shmagency"). For a discussion of the shmagency worry in the context of evaluative pluralism, see Clarke-Doane, *Morality and Mathematics*, 168.

Consider the version of metaethical noncognitivism according to which normative judgments are intentions, plans, decisions, or some similar kind of mental state.[62] According to such a view (most straightforwardly understood at least), to judge that an action ought to be performed is not to have a belief about it—instead, to make such a judgment is, roughly, to intend or decide to perform the action in question if the opportunity should arise.[63] Hence, on this view, uncertainty about what one ought to do does not amount to uncertainty about the truth of any proposition. The reason is simply that, unlike beliefs, the relevant mental states—decisions, intentions, and the like—cannot be true or false.[64] Instead, uncertainty about what one ought to do amounts to a kind of noncognitive uncertainty about what decision to make. It is better characterized as a state of practical indecision (i.e., as a state of not having decided which action to perform in the relevant situation, or not having formed an intention about it, etc.) than as a state of uncertainty about what the world is like.

The noncognitivist account of normative uncertainty just sketched has serious problems.[65] Hence, throughout the paper, I have assumed that it is false. Instead, I have taken for granted the cognitivist views that normative judgments are beliefs that can be true or false, that normative uncertainty amounts to uncertainty about the truth of normative propositions, and so on. However, even assuming the truth of those views, what I wish to propose is that the noncognitivist view just outlined is true of something else: namely, of the deliberative uncertainty, or the "central deliberative question," that we have all along struggled to express literally. On this "divided" view of normative

62 Like many others, I assume that it is true of at least some kind of noncognitive attitude that if I have that attitude toward performing a given action right now, then I will perform that action right now if I can. For instance, Paul Grice endorses this for intention ("Intention and Uncertainty," 263–64), and Gibbard endorses it for planning (*Thinking How to Live*, 152–53). If there are several types of noncognitive attitudes that are related to action in this way, then the differences between them will not matter in what follows.

63 While this view is associated with Gibbard (*Thinking How to Live*), note that whereas Gibbard endorses a quasi-realist version of this view about normative judgments, I do not favor it either as a version of quasi-realism or as a view of normative judgments, for reasons that I will get to in a moment.

64 I am assuming the falsity of extreme forms of cognitivism about intentions, according to which my intention to perform a certain action is simply identical to my belief that I will perform that action. Such views face well-known problems; see, e.g., Bratman, "Intention and Means-End Reasoning."

65 See, e.g., Bykvist and Olson, "Expressivism and Moral Certitude"; and MacAskill, Bykvist, and Ord, *Moral Uncertainty*, ch. 7. Briefly, the problem is that while normative judgments can vary in at least two independent dimensions—how good we judge that something is, for example, and how confident we are that it is good to that degree—paradigm noncognitive states, like desires, vary only in one dimension, i.e., with respect to their strength.

and deliberative uncertainty, uncertainty of the latter sort does not concern a puzzling normative question that is special in some seemingly inexpressible sense. Rather, I suggest, such uncertainty simply concerns what we may call the question of *what to do*. As Jamie Dreier puts it, this type of question is the one that "you answer when and only when you have decided what to do. It is answered with an intention, perhaps, or a plan."[66] It is not answered by a belief or some similar kind of mental state. Thus, *a fortiori*, it is not answered by a belief whose content is a proposition about what you ought to do, or any other normative proposition.[67]

Dreier suggests that the relevant type of question can be expressed by the interrogative sentence, "What shall I do?" Dreier's claim may or may not be correct, but this need not concern us here.[68] What matters is that we can informatively characterize the relevant mental state—it is a separate question whether we can perspicuously express it in English or some other natural language. What is plausible, however, is that many of the not easily understood interrogative sentences that I have discussed in this paper are naturally understood as *attempts at* communicating this type of state, even if they do not do so precisely. These include the following:

(i) I don't know what I ought to do; *now* what ought I to do? (section 3);
(ii) Which "ought" *ought* I to act on? (section 4);
(iii) Are the claims that morality make on us really justified? (section 5);
(iv) What I ought to do strikes me as outrageous; *now* what ought I to do? (section 6).

66 Dreier, "Can Reasons Fundamentalism Answer the Normative Question?" 172.

67 While I find it natural to talk about "questions" and "answers" in this way, it is worth noting that these expressions are ambiguous in ways that can cause confusion. In one sense, the question of whether *p* has two answers (at least disregarding indeterminacy and the like); these are its "possible" or "candidate" answers. One possible answer to this question is that *p*, and the other is that *not-p*. In another sense, however, the question of whether *p* has only one answer—this is its *true* or *correct* answer. If *p* is true, then *p* is the true answer to the question of whether *p*, whereas if *p* is false, then the true answer to that question is *not-p*. To *answer* (verb) the question of whether *p*, moreover, can also mean different things: in one sense, to answer a question is to perform the speech act of asserting a candidate answer to it (in a suitable context), whereas if one ponders the question for oneself, then one answers it by accepting one of its candidate answers, in this case by forming either a belief that *p* or a belief that *not-p*. It is this latter, "first-personal" sense of "[to] answer" that I have in mind in the main text.

68 Perhaps, as suggested to me by Michael Zimmerman (in personal communication), the sentence "What *am* I to do?" better captures the relevant question.

What I propose is that, on a natural interpretation, these questions are all strictly speaking unsuccessful attempts at expressing uncertainty about the question of what to do in the relevant situation, i.e., of what to do when one does not know what one ought to do, of what to do when different "oughts" are in conflict, of whether to act in accordance with the claims that morality makes on us, and of what to do if the normative truths turn out to be outrageous, respectively. Unlike alternative views, this view explains why this type of uncertainty may remain even given that we have knowledge of all the truths, including all the normative ones. Whether we have knowledge of those truths ultimately does not matter, because the relevant question never directly concerned them in the first place. Instead, because the deliberative question does not even concern what is true, it does not have a true answer.

Another advantage is that this view avoids the tie-breaking problem by steering between the horns of the dilemma (cf. section 4). The problem, recall, is that when different oughts diverge (e.g., the subjective and the objective ought, or the moral and the prudential ought, or even OUGHT and OUGHT*), it is hard to make sense of the salient further question in the neighborhood of which ought we *really* ought to act on. On the one hand, descriptive truths seem obviously beside the point. On the other hand, appealing to further normative truths seems only to relocate the problem. This dilemma is avoided if the salient question is neither about some descriptive truth or some normative truth, but is instead noncognitive and thus not about any truth at all.

It is also worth noting that, unlike most contemporary forms of noncognitivism about normative judgments, the noncognitivist view I have developed here is not wedded to the research program of "quasi-realism."[69] While the quasi-realist's position has always been difficult to state precisely, her aim is to explain most or all realist-seeming notions—e.g., normative truths, beliefs, knowledge, reasoning, argumentation, uncertainty—in noncognitivist-friendly terms, and thus avoid the "heavyweight," supposedly problematic commitments that genuine realists incur. Whether this program is successful is at best extremely controversial.[70] And since the form of noncognitivism I have presented does not concern normative judgments proper, it does not require for its truth that the quasi-realist program succeeds. Accordingly, my account allows (though it does not entail) that normative truths, beliefs, and the like

69 The program was first endorsed by Simon Blackburn (see, e.g., *Essays in Quasi-Realism*) and many others have since followed suit; for example, Toppinen even suggests that quasi-realists should endorse nonnaturalism about normative truths ("Non-naturalism Gone Quasi").

70 For two influential critical discussions of quasi-realism, see Dreier, "Meta-Ethics and the Problem of Creeping Minimalism"; and Schroeder, *Being For*.

are best understood in realist terms, so that normative truth requires correspondence with reality, normative beliefs are "fully representational," and so on.

While the noncognitivist view of deliberative uncertainty thus diverges from noncognitivism about normative judgments, I think that it nonetheless captures an important intuition that has often been invoked in support of the latter view. This is the intuition that certain practical questions seem not to answer to matters of fact. No matter how much we learn about the world, those questions may in principle remain open.[71] The divided view vindicates this intuition, since it entails that the question of what to do is not a question of fact. In contrast, whether quasi-realist versions of metaethical noncognitivism ultimately vindicate this intuition as well is far from clear. For what the quasi-realist assumes is that the relevant practical questions are questions about what we ought to do. Thus, when she goes on to try to accommodate the possibility of normative truth, knowledge, and so on, she no longer has the resources to explain why the relevant practical questions could remain open even given that we have knowledge of all the truths, including the normative ones.

9. Two Challenges

Before concluding, I will consider two possible challenges for my view.[72] The first challenge is that it might fail to capture the "normativity" of ought truths and/or ought judgments (or "oughts," for short). This challenge can be spelled out in different ways, depending on how the relevant notion of normativity is understood. However, I will argue that each version of the challenge can be met. For many things that can be meant by "normative," my account allows that oughts are normative. For some possible senses of "normative," the account may well rule out that oughts are normative, but there is also no strong independent support for thinking that oughts are normative in those senses. Either way, then, the challenge fails.

To begin with, one possible idea is that a truth or judgment is normative just in case it is related to some suitable normative notion, such as *ought*, *good*, or

71 For example, this intuition arguably underlies Nowell-Smith's remark that "learning about 'values' or 'duties' might well be as exciting as learning about spiral nebulae or waterspouts. But what if I am not interested? Why should I *do* anything about these newly-revealed objects?" (*Ethics*, 41). It is also illustrated by noncognitivists' frequent reliance upon Moore's open-question argument (cf. Darwall, Gibbard, and Railton, "Toward *Fin de siècle* Ethics") and the assumption that moral disagreements could remain even in "ideal" epistemic conditions (see, e.g., Tersman, *Moral Disagreement*).

72 Thanks to two anonymous referees for presenting the two challenges considered in this section.

reason, in the right way. For instance, Schroeder suggests that what is "distinctive of the normative" is that it is "all about reasons," and John Broome suggests that the term "normative" means "to do with ought," where the relevant "ought" "is a normative one."[73] This may be called the *trivial* sense of "normative," since it implies that at least one normative notion—i.e., reasons for Schroeder and ought for Broome—counts as normative simply by fiat, by being related to itself in the right way. I myself suspect that it is difficult to provide a more informative characterization of normativity than one along these lines, but my response does not rest on this assumption. What I want to emphasize is just that, clearly, nothing in my account excludes that oughts are normative in this sense: oughts may well count as normative because they are analyzable in terms of normative reasons, for instance, or (as Broome's view suggests) simply because they are oughts. What matters is just that even if oughts are normative in this sense, we may still ask what to do with them.

A version of the idea just presented is to take oughts to be normative in the sense of standing in some relation to the normative notions of *rationality* and/or *coherence*.[74] For instance, perhaps oughts count as normative because they figure in some true "enkratic" principle, such as: if a subject judges that she ought to do *A* but does not intend to do *A*, then she is incoherent or irrational. This idea is also compatible with my account—just as we can ask whether to do what we ought to do, we can also ask whether to be incoherent, whether to be irrational, and so on.

Another popular idea is that we should distinguish between *robust* and *merely formal* normativity.[75] This distinction departs from the intuitive difference between the oughts of (e.g.) morality, epistemology, and prudence on the one hand, which are usually taken to be robustly normative, and those of (e.g.) etiquette, chess, and grammar on the other hand, which are usually taken to be merely formally normative. How this intuitive difference should be cashed out in more detail is controversial. One view is that robustly normative requirements differ from merely formally normative ones in that they entail the existence of genuine (or genuinely normative) oughts, reasons, or the like. This view takes us back to the first suggestion considered above—I have already argued that my account allows that some oughts are normative in this sense. Another view is that robustly normative oughts differ from merely formally normative ones in that they are in some suitable sense not "up to us." For instance, maybe the oughts of etiquette, chess, and grammar depend on our

73 Schroeder, "Realism and Reduction," 13; Broome, *Rationality through Reasoning*, 10.

74 A version of this challenge was offered to me by Jonathan Way (in personal communication).

75 See further, e.g., Finlay, "Defining Normativity," sec. 3.2.

attitudes and conventions in a way that the oughts of morality, prudence, and epistemology do not. (This might, but need not, in turn be because robustly normative oughts are "nonnatural.") Nothing in my account rules out that some oughts are robustly normative in this sense either as, again, the supposition that some oughts are not up to us (and perhaps also nonnatural) does not prevent us from asking what to do with them.

A somewhat different idea is that ought judgments are normative in the sense that they are necessarily connected to *motivation*. The most straightforward version of this kind of "motivational internalism" states that if a subject judges that she ought to perform *A*, then she is at least defeasibly motivated to perform *A*. To begin with, I think the arguments in this paper provide at least some reason to deny such a strong, unqualified form of internalism. It seems plausible, for instance, that an ought judgment might leave a subject motivationally "cold" if she finds it outrageous, or if she cares only about what she ought* to do. That said, the account I have presented is consistent with the idea that ought judgments necessarily provide even a very high degree of defeasible motivation to perform the relevant action. The reason is that, even if this is true, we still face the question of whether to act in accordance with the motivation that the ought judgment provides.

The only version of motivational internalism that might pose problems for my proposal is the extremely strong view that ought judgments always provide *overriding* or *indefeasible* motivation, so that if a subject judges that she ought to perform *A*, then she performs *A* (at least if she can). The reason is that this view entails that we always do what we think we ought to do, which in turn could make it hard to see how we might answer the practical question of what *to* do and the normative question of what we ought to do in different ways. However, such extreme forms of internalism are arguably *too* extreme; indeed, as Fredrik Björklund, Gunnar Björnsson, John Eriksson, Ragnar Francén Olinder, and Caj Strandberg note, "in contemporary metaethics, it is regularly assumed" that even the view that ought judgments entail *defeasible* motivation is "too strong," since counterexamples to it seem "possible to conceive."[76] Thus, I am happy to simply assume that this extreme form of motivational internalism, according to which normative judgments always provide overriding motivation, is false.

The final idea I will consider is that ought judgments are normative in the sense that they constitute answers to questions about what to do. As it stands, this suggestion does not amount to much more than the denial of the view that I have offered. I do not deny that this suggestion has often been taken for granted—on the contrary, as I have emphasized, the assumption that

76 Björklund, Björnsson, Eriksson, et al., "Recent Work on Motivational Internalism," 126.

something like this view is true arguably underlies several important debates in ethics, metaethics, and metanormativity. But unless some independent support for this suggestion is presented, the mere fact that my view contradicts it surely cannot itself be seen as an objection.

The second challenge that I will consider is that, if my account is correct, it is not clear why we so often ask ourselves normative questions, especially when we are trying to reach choices. Why do we not simply ask ourselves what to do, rather than what we ought to do, if these questions really are distinct? While this challenge is not identical from the first one, they might still be related, as this challenge can also be understood as an expression of the more general worry that oughts are in some sense more practically significant than my account allows.

The question of why we ask ourselves what we ought to do rather than what to do (when we do so) may well be at least partially an empirical question. Accordingly, what I will have to say about it is bound to be somewhat speculative. That said, three considerations are worth noting before closing.

First, a kind of error theory might in some cases be plausible. After all, philosophical theorizing often allows us to draw distinctions that we do not usually recognize in everyday life. So it might be that we sometimes ask ourselves what we ought to do rather than what to do simply because we have not realized that these questions are distinct. If we were to realize this, perhaps we would care less about what we ought to do and more about what to do. I am not suggesting that this kind of error theory fully explains why we so frequently ask ourselves what we ought to do (rather than what to do), but it may at least play a role in such an explanation.

Second, another partial explanation might be that we are sometimes simply interested in what the normative truths are. In particular, even if Ross is right that we do not ask ourselves normative questions *only* to satisfy our curiosity (cf. section 2), that does not entail that curiosity is never even *part* of the reason why we do so. Indeed, although I have focused on normative and metanormative debates that highlight the question of what to do, many other debates in these fields are less closely connected to action. For instance, it is far from clear how questions such as whether the betterness relation is transitive or whether the good is more fundamental than the right could even have a bearing (except perhaps very indirectly) on the question of what to do. The reason why we investigate them might instead be the same as when we try to find out whether the causation relation is transitive or whether the brain is more fundamental than the mind: we are simply interested in their answers.

Third and finally, as externalists about moral motivation have emphasized, it might be a metaphysically contingent but still quite modally robust fact about

us that we often want to do the right thing.[77] This suggestion interacts with the second one made above, since it sheds further light on why we might often be interested in figuring out answers to normative questions: doing so might not only satisfy our intellectual curiosity but also help us achieve something we want. If this idea is correct, it also helps explain why we often do not consider the question of whether to do what we ought to do—we might simply be happy with figuring out what we ought to do and do our best to act accordingly.

10. CONCLUDING REMARKS

In this paper, I have focused on a distinctive deliberative question that many debates in ethics, metaethics, and metanormativity highlight. I have argued against the common view that this question concerns a special normative notion. Instead, I have offered a combination of cognitivism about normative questions and noncognitivism about the question of what to do. An upshot of this divided view is that even if there are truths about what we all things considered ought to do, the central deliberative question does not have a true answer.

As I have noted (cf. section 8), my view is strictly speaking consistent with the "robust" or "ardent" realist position that there are objective, irreducible, heavyweight truths about how we ought to act.[78] That said, however, I do think that my view threatens to undermine an important argument for normative realism. For if the question of what to do does not even concern the truths that realists posit, then there is one way in which those truths seem much less interesting than we often take them to be. At least in many cases, as Jackson and Ross suggest (section 2), we do not ask ourselves what we ought to do with the sole aim of learning more about the world. We also do so to reach choices in our lives. An attractive feature of ardent realism is that it promises to take such questions seriously, by positing objective truths that are supposed to constitute their answers.[79] However, this argument is undercut if there is an open practical question that remains even when we acquire knowledge of those truths: whether to do what we ought to do.[80]

Uppsala University
olle.risberg@filosofi.uu.se

77 See, e.g., Copp, "Belief, Reason and Motivation," 49–51.

78 The labels "robust realism" and "ardent realism" are from, respectively, Enoch, *Taking Morality Seriously*; and Eklund, *Choosing Normative Concepts*.

79 For example, this idea seems to underlie David Enoch's argument for the view that irreducibly normative truths are indispensable for deliberation (*Taking Morality Seriously*, ch. 3).

80 For very valuable comments on earlier versions of this paper, thanks to Karl Bergman, Daniel Fogal, Anna Folland, Jens Johansson, Simon Rosenqvist, Debbie Roberts, Amogha

REFERENCES

Arrhenius, Gustaf. *Future Generations: A Challenge for Moral Theory*. Uppsala, Sweden: Acta Universitatis Upsaliensis, 2000.

Baker, Derek. "Skepticism about Ought Simpliciter." In *Oxford Studies in Metaethics*, vol. 13, edited by Russ Shafer-Landau, 230–52. Oxford: Oxford University Press, 2018.

Balaguer, Mark. "Moral Folkism and the Deflation of (Lots of) Normative and Metaethics." In *Abstract Objects: For and Against*, edited by José L. Falguera and Concha Martinez-Vidal, 297–318. Cham, Switzerland: Springer, 2020.

Bedke, Matthew. "A Dilemma for Non-naturalists: Irrationality or Immorality?" *Philosophical Studies* 177, no. 4 (April 2020): 1027–42.

Björklund, Fredrik, Gunnar Björnsson, John Eriksson, Ragnar Francén Olinder, and Caj Strandberg. "Recent Work on Motivational Internalism." *Analysis* 72, no. 1 (January 2012): 124–37.

Blackburn, Simon. *Essays in Quasi-Realism*. Oxford: Oxford University Press, 1993.

Bratman, Michael. "Intention and Means-End Reasoning." *Philosophical Review* 90, no. 2 (April 1981): 252–65.

Broome, John. *Rationality through Reasoning*. Malden, MA: Wiley-Blackwell, 2013.

Bykvist, Krister. "How to Do Wrong Knowingly and Get Away with It." In *Neither/Nor: Philosophical Papers Dedicated to Erik Carlson on the Occasion of His Fiftieth Birthday*, edited by Frans Svensson and Rysiek Sliwinski, 31–47. Uppsala, Sweden: Uppsala University, 2011.

———. "Violations of Normative Invariance: Some Thoughts on Shifty Oughts." *Theoria* 73, no. 2 (April 2007): 98–120.

Bykvist, Krister, and Jonas Olson. "Expressivism and Moral Certitude." *Philosophical Quarterly* 59, no. 235 (April 2009): 202–15.

Carlson, Erik. "Deliberation, Foreknowledge, and Morality as a Guide to Action." *Erkenntnis* 57 (July 2002): 71–89.

Sahu, Jonathan Way, and in particular, to Krister Bykvist, Erik Carlson, Justin Clarke-Doane, Matti Eklund, Nils Franzén, Victor Moberger, Wlodek Rabinowicz, Andrew Reisner, and two anonymous referees for this journal. Thanks also to participants at the Uppsala PhD seminar in Philosophy, the Uppsala Higher Seminar in Practical Philosophy, the Future of Normativity conference at University of Kent, and the Philosophy of Logic seminars at Columbia University in 2018. Part of the work on this paper was generously supported by Helge Ax:son Johnson's Foundation; Hultengren's Foundation; the Royal Swedish Academy of Letters, History and Antiquities; the Salén Foundation; Sixten Gemzéus' Foundation; Thun's Grant Foundation; Värmlands Nation in Uppsala; and Grant 2020-01955 from the Swedish Research Council.

Chang, Ruth. "'All Things Considered.'" *Philosophical Perspectives* 18, no. 1 (December 2004): 1–22.

Clarke-Doane, Justin. "From Non-usability to Non-factualism." *Analysis* 81, no. 4 (October 2021): 747–58.

———. *Morality and Mathematics*. New York: Oxford University Press, 2020.

Copp, David. "Belief, Reason and Motivation: Michael Smith's *The Moral Problem*." *Ethics* 108, no. 1 (October 1997): 33–54.

———. "The Ring of Gyges: Overridingness and the Unity of Reason." *Social Philosophy and Policy* 14, no. 1 (Winter 1997): 86–106.

Crisp, Roger. *Reasons and the Good*. New York: Oxford University Press, 2006.

Cuneo, Terence, and Russ Shafer-Landau. "The Moral Fixed Points: New Directions for Moral Nonnaturalism." *Philosophical Studies* 171, no. 3 (December 2014): 399–443.

Darwall, Stephen, Allan Gibbard, and Peter Railton. "Toward *Fin de Siècle* Ethics: Some Trends." *Philosophical Review* 101, no. 1 (January 1992): 115–89.

Dreier, James. "Can Reasons Fundamentalism Answer the Normative Question?" In *Motivational Internalism*, edited by Fredrik Björklund, Gunnar Björnsson, John Eriksson, Ragnar Francén Olinder, and Caj Strandberg, 167–81. New York: Oxford University Press, 2015.

———. "Meta-Ethics and the Problem of Creeping Minimalism." *Philosophical Perspectives* 18, no. 1 (December 2004): 23–44.

Eklund, Matti. *Choosing Normative Concepts*. New York: Oxford University Press, 2017.

———. "The Normative Pluriverse." *Journal of Ethics and Social Philosophy* 18, no. 2 (August 2020): 121–46.

Enoch, David. "Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action." *Philosophical Review* 115, no. 2 (April 2006): 169–98.

———. *Taking Morality Seriously*. New York: Oxford University Press, 2011.

———. "Thanks, We're Good: Why Moral Realism Is Not Morally Objectionable." *Philosophical Studies* 178, no. 5 (May 2021): 1689–99.

Faraci, David. "On Leaving Room for Doubt." In *Oxford Studies in Metaethics*, vol. 12, edited by Russ Shafer-Landau, 244–64. Oxford: Oxford University Press, 2017.

Feldman, Fred. "Actual Utility, the Objection from Impracticality, and the Move to Expected Utility." *Philosophical Studies* 129, no. 1 (May 2006): 49–79.

Finlay, Stephen. "Defining Normativity." In *Dimensions of Normativity: New Essays on Metaethics and Jurisprudence*, edited by David Plunkett, Kevin Toh, and Scott J. Shapiro, 187–220. New York: Oxford University Press, 2019.

Fitzpatrick, William. Commentary on Matti Eklund, *Choosing Normative*

Concepts. PEA Soup. 2018. http://www.sas.rochester.edu/phl/fitzpatrick/PeaSoup.pdf.

Gibbard, Allan. "Morality as Consistency in Living: Korsgaard's Kantian Lectures." *Ethics* 110, no. 1 (October 1999): 140–64.

———. *Thinking How to Live*. Cambridge, MA: Harvard University Press, 2003.

———. *Wise Choices, Apt Feelings*. Oxford: Clarendon Press, 1992.

Graham, Peter. "In Defense of Objectivism about Moral Obligation." *Ethics* 121, no. 1 (October 2010): 88–115.

Grice, Paul. "Intention and Uncertainty." *Proceedings of the British Academy* 57 (1971): 263–79.

Harman, Elizabeth. "The Irrelevance of Moral Uncertainty." In *Oxford Studies in Metaethics*, vol. 10, edited by Russ Shafer-Landau, 53–79. Oxford: Oxford University Press, 2015.

Jackson, Frank. "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics* 101, no. 3 (April 1991): 461–82.

———. *From Metaphysics to Ethics*. Oxford: Clarendon Press.

Kolodny, Niko, and John MacFarlane. "Ifs and Oughts." *Journal of Philosophy* 107, no. 3 (March 2010): 115–43.

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

Lenman, James. "Consequentialism and Cluelessness." *Philosophy and Public Affairs* 29, no. 4 (Autumn 2000): 342–70.

Lord, Errol. "What You're Rationally Required to Do and What You Ought to Do (Are the Same Thing!)" *Mind* 126, no. 504 (October 2017): 1109–54.

MacAskill, William. "The Infectiousness of Nihilism." *Ethics* 123, no. 3 (April 2013): 508–20.

MacAskill, William, Krister Bykvist, and Toby Ord. *Moral Uncertainty*. Oxford: Oxford University Press, 2020.

Mason, Elinor. "Objectivism and Prospectivism about Rightness." *Journal of Ethics and Social Philosophy* 7, no. 2 (March 2013): 1–22.

McGrath, Sarah. "Moral Disagreement and Moral Expertise." In *Oxford Studies in Metaethics*, vol. 4, edited by Russ Shafer-Landau, 87–108. Oxford: Oxford University Press, 2008.

McPherson, Tristram. "Explaining Practical Normativity." *Topoi* 37, no. 4 (December 2018): 621–30.

Nowell-Smith, P. H. *Ethics*. Harmondsworth: Pelican Books, 1954.

Parfit, Derek. *On What Matters*, vol. 2. Oxford: Oxford University Press, 2011.

———. *Reasons and Persons*. Oxford: Oxford University Press, 1984.

———. "What We Together Do." Unpublished manuscript. https://philarchive.org/archive/PARWWT-3.pdf.

Railton, Peter. "Alienation, Consequentialism, and the Demands of Morality." *Philosophy and Public Affairs* 13, no. 2 (Spring 1984): 134–71.

Regan, Donald. *Utilitarianism and Co-operation*. Oxford: Oxford University Press, 1980.

Reisner, Andrew. "Is There Reason to Be Theoretically Rational?" In *Reasons for Belief*, edited by Andrew Reisner and Asbjørn Steglich-Petersen, 34–53. Cambridge: Cambridge University Press, 2011.

———. "Normative Conflicts and the Structure of Normativity." In *Weighing and Reasoning*, edited by Iwao Hirose and Andrew Reisner, 189–206. Oxford: Oxford University Press, 2015.

Risberg, Olle. "The Entanglement Problem and Idealization in Moral Philosophy." *Philosophical Quarterly* 68, no. 272 (July 2018): 542–59.

———. "Weighting Surprise Parties: Some Problems for Schroeder." *Utilitas* 28, no. 1 (March 2016): 101–7.

Risberg, Olle, and Folke Tersman. "Disagreement, Indirect Defeat, and Higher-Order Evidence." In *Higher-Order Evidence and Moral Epistemology*, edited by Michael Klenk, 97–114. London: Routledge, 2020.

———. "Moral Realism and the Argument from Skepticism." *International Journal for the Study of Skepticism* 10, nos. 3–4 (November 2020): 283–303.

———. "A New Route from Moral Disagreement to Moral Skepticism." *Journal of the American Philosophical Association* 5, no. 2 (Summer 2019): 189–207.

Ross, Jacob. "Rationality, Normativity, and Commitment." In *Oxford Studies in Metaethics*, vol. 7, edited by Russ Shafer-Landau, 138–81. Oxford: Oxford University Press, 2012.

Schroeder, Mark. *Being For*. Oxford: Oxford University Press, 2008.

———. "Ought, Agents, and Actions." *Philosophical Review* 120, no. 1 (January 2011): 1–41.

———. "Realism and Reduction: The Quest for Robustness." *Philosophers' Imprint* 5, no. 1 (February 2005): 1–18.

Sepielli, Andrew. "Subjective and Objective Reasons." In *The Oxford Handbook of Reasons and Normativity*, edited by Daniel Star, 784–99. Oxford: Oxford University Press, 2018.

———. "What to Do When You Don't Know What to Do." In *Oxford Studies in Metaethics*, vol. 4, edited by Russ Shafer-Landau, 5–28. Oxford: Oxford University Press, 2008.

———. "What to Do When You Don't Know What to Do When You Don't Know What to Do …" *Noûs* 48, no. 3 (September 2014): 521–44.

Sidgwick, Henry. *The Methods of Ethics*. London: Macmillan, 1874.

Skorupski, John. *The Domain of Reasons*. Oxford: Oxford University Press, 2010.

Smith, Holly. *Making Morality Work*. Oxford: Oxford University Press, 2018.

Smith, Michael. "Moore on the Right, the Good, and Uncertainty." In *Metaethics After Moore*, edited by Terence Horgan and Mark Timmons, 133–48. Oxford: Oxford University Press, 2006.

———. *The Moral Problem*. Oxford: Blackwell, 1994.

Tännsjö, Torbjörn. *From Reasons to Norms: On the Basic Question in Ethics*. Dordrecht, Netherlands: Springer, 2010.

Tersman, Folke. *Moral Disagreement*. Cambridge: Cambridge University Press, 2006.

Tiffany, Evan. "Deflationary Normative Pluralism." *Canadian Journal of Philosophy* 37, supplementary vol. 33 (2007): 231–62.

Timmons, Mark. *Moral Theory: An Introduction*. Lanham, MD: Rowman and Littlefield Publishers, 2002.

Toppinen, Teemu. "Non-naturalism Gone Quasi: Explaining the Necessary Connections between the Natural and the Normative." In *Oxford Studies in Metaethics*, vol. 13, edited by Russ Shafer-Landau, 26–47. Oxford: Oxford University Press, 2018.

Way, Jonathan, and Daniel Whiting. "Perspectivism and the Argument from Guidance." *Ethical Theory and Moral Practice* 20, no. 2 (April 2017): 361–74.

Weatherson, Brian. *Normative Externalism*. Oxford: Oxford University Press, 2019.

———. Review of *Moral Uncertainty and Its Consequences*. *Mind* 111, no. 443 (July 2002): 693–96.

Wedgwood, Ralph. "Conceptual Role Semantics for Moral Terms." *Philosophical Review* 110, no. 1 (January 2001): 1–30.

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

———. *Ignorance and Moral Obligation*. Oxford: Oxford University Press, 2014.

———. *Living with Uncertainty: The Moral Significance of Ignorance*. Cambridge: Cambridge University Press, 2008.

Journal of Ethics and Social Philosophy
VOL. 25, NO. 2 · AUGUST 2023
https://doi.org/10.26556/jesp.v25i2.2616

RESCUE AND NECESSITY

A REPLY TO QUONG

Joel Joseph and Theron Pummer

Suppose that *A* is wrongfully attempting to kill you. *A* is therefore liable to defensive harm: he has forfeited his right not to be proportionately harmed by you. Suppose the only way to stop *A*'s attack is by launching a large grenade at him, blowing off his arms and legs. We take it that this would be proportionate. So, you are permitted to impose this harm on *A* in self-defense.

Next suppose that you can stop *A* by launching either the large grenade or a smaller one that would blow off his left arm only. While each of your defensive alternatives is proportionate, now it is impermissible to launch the large grenade at *A*. Doing so would violate the necessity condition on imposing harm.

Jonathan Quong provides an ingenious account of the necessity condition.[1] According to Quong, even though *A* is liable to defensive harm, *A* retains his right to be rescued. We agree. If while wrongfully attempting to kill you, *A* tripped and fell onto a trolley track, putting him in imminent danger of losing three of his limbs, it would be impermissible not to rescue *A* if this were costless to you. Failing to rescue *A* would violate *A*'s right to be rescued from serious harm. Crucially, Quong holds that this is also why it would be wrong to launch the large grenade at *A* rather than the small grenade: by blowing off four of *A*'s limbs in proportionate self-defense rather than blowing off one of *A*'s limbs in proportionate self-defense, you are failing to costlessly rescue three of *A*'s limbs. The impermissibility of imposing unnecessary harm in self-defense is explained in terms of the violation of the right to be rescued.

While we think there is much to be said for Quong's account of the necessity condition, it has implausible implications. In what follows, we present three related objections to Quong's view. First, consider:

> *Conflict*: Albert is wrongfully attempting to kill you. Meanwhile, Betty has tripped and fallen onto a trolley track, where a trolley is about to sever her right arm and legs. You can (1) press a button that stops Albert

1 Quong, *The Morality of Defensive Force*, ch. 5.

> by severing his left arm (allowing Betty's right arm and legs to be severed); (2) press a button that stops Albert by severing his arms and legs and, separately, moves Betty out of the way of the trolley that was going to sever her right arm and legs; or (3) do nothing, allowing yourself to be killed by Albert and Betty's right arm and legs to be severed.

Quong's account implies there are conflicting rights to be rescued. If you do 1, you contravene Betty's right to be rescued. If you do 2, you contravene Albert's right to be rescued. Either way, the harm that would be prevented is three limbs. So, Quong's account implies either that both 1 and 2 are permissible or else that you are required to toss a coin to decide between 1 and 2 if you are not going to do 3. This, after all, is what holds when there is a conflict between Albert's right to be rescued and Betty's equally stringent right to be rescued.

The problem is that 2 seems impermissible. Notice that 2 does not violate the means principle: it does not harm Albert as a means to saving Betty. Instead, 2 has two causally separate effects: pressing the button causes one effect on Albert, and it causes another effect on Betty.[2]

Intuitively, it is impermissible to do 2 rather than 1 because this involves causing one person (Albert) to lose three limbs while preventing another person (Betty) from losing three limbs. In general, it is impermissible to cause a harm *H* to one person while preventing that same harm *H* to another person. Indeed, it is impermissible to cause harm *H* to one person while preventing a significantly greater harm *H+* to another person. Option 2 would remain impermissible even if it prevented Betty from losing four limbs.

While Quong does not consider cases similar to Conflict, he does consider the objection that his rescue account of the necessity condition is mistaken because "the necessity condition is a constraint against harming, and so it is more demanding than a duty of rescue."[3] In response, he offers the following pair of cases:

> *Attack*: Albert wrongfully attacks Betty. Betty has two ways of averting Albert's attack: using lethal defensive force (which is narrowly proportionate), which will cause no harm to Betty, or jumping to safety at some cost *C* to herself.

2 See Quong, *The Morality of Defensive Force*, 178, for his formulation of the means principle. Note that 2 does not violate other related principles, such as F. M. Kamm's "doctrine of productive purity" (*Intricate Ethics*, 164) or Ketan H. Ramakrishnan's "utility" principle ("Treating People as Tools," 134).

3 Quong, *The Morality of Defensive Force*, 143.

> *Drowning*: Albert is wrongfully attempting to attack Betty. Betty can avert Albert's attack by simply doing nothing, as Albert will then step onto a faulty bridge, causing him to fall into a lake and drown. Alternatively, Betty can jump into an alcove, at some cost *C* to herself. If she jumps into the alcove, Albert will withdraw before reaching the faulty bridge.[4]

Quong writes that "the only apparent difference is that, in Drowning, Betty must provide aid, whereas in Attack Betty must refrain from harming Albert. Intuitively, however, it does not seem to me that the cost Betty is duty bound to bear should be higher in Attack than in Drowning."[5] He concludes that the objection in question is mistaken.

While we accept Quong's view on this pair of cases, it does not follow that the objection in question is mistaken. Even though the duty to save Albert in Drowning is as strong as the duty not to kill Albert in Attack, that does not imply that, in Conflict, the duty to save Betty from losing three limbs is as strong as the duty not to cause Albert to lose three limbs.

The duty to save Albert in Drowning is as strong as the duty to not kill Albert in Attack, because as Quong claims, Albert forfeits his right not to be proportionately harmed in self-defense, yet in both cases he retains his right to be rescued. The level of cost Betty is required to incur is the same in both cases, because in both cases she has only a duty to rescue Albert. But now consider our second objection, based on the following variant of Attack:

> *Attack (Extra Threat)*: Albert is wrongfully attempting to sever Betty's arm. Meanwhile, a runaway trolley is independently threatening to sever her legs. Betty can (1) press a button that stops Albert by severing his left finger, saving Betty's arm but allowing her legs to be severed; (2) press a button that stops Albert by severing his left arm and, separately, moves Betty out of the way of the trolley that was going to sever her legs; or (3) do nothing, allowing herself to lose an arm and both legs.

In this case, 2 seems impermissible. While Albert forfeits his right not to be proportionately harmed in self-defense, intuitively, he retains more than just a right to be rescued. Intuitively, Albert retains a stringent right not to have *additional* harm imposed on him while averting the threat to Betty *from the trolley*. This right is more stringent than a right to be rescued: while Betty can

4 We have modified Quong's original version of Drowning (*The Morality of Defensive Force*, 143) so that allowing Albert to die is what stops the attacker's threat. This makes Drowning closer to Attack.

5 Quong, *The Morality of Defensive Force*, 143–44.

permissibly rescue her legs rather than rescue Albert's left arm, she cannot permissibly cause Albert to lose his left arm as a side effect of rescuing her legs; the former does not wrong Albert but the latter does. Since on Quong's rescue account, Betty's only duty to Albert is a duty to rescue him from the harm corresponding to the difference between losing his left arm and losing his left finger, his account implies that 2 is permissible.

Conflict is different in that 1 and 2 are equally costly to the agent, where 2 involves doing harm to the attacker while preventing harm to a bystander and 1 involves allowing harm to the bystander. But the crucial point is again that while Albert forfeits his right not to be proportionately harmed in self-defense, intuitively he retains more than just a right to be rescued—it is permissible to save three of Betty's limbs rather than save three of Albert's limbs if you choose between them fairly. Intuitively, however, 2 is impermissible because it violates the necessity constraint. Intuitively, Albert retains a stringent right not to have *additional* harm imposed on him while averting the threat to Betty for which he is not responsible.[6] The fact that this right is more stringent than Betty's right to be rescued is what explains why 2 is impermissible. Similar remarks apply to Attack (Extra Threat). This suggests that violating the necessity constraint involves the violation of a right that is more stringent than a right to be rescued.

So, even if cases such as Attack and Drowning fail to show that the necessity condition is more stringent than a duty of rescue, cases such as Conflict and Attack (Extra Threat) plausibly do show this.

Could Quong not respond by conceding that 2 in Conflict and 2 in Attack (Extra Threat) are impermissible, not because they violate the necessity condition but because the attacker is not liable to the additional harm these acts impose? An immediate problem with this response is that these acts *do* seem impermissible because they violate the necessity condition. But even setting this problem aside, this response is unavailable to Quong. According to his rescue account of necessity, even when you gratuitously impose additional yet proportionate harm on an attacker in self-defense, you wrong them only by violating their right to be rescued. The attacker has forfeited their right not to be proportionately harmed in self-defense; this particular harm is proportionate, and they are therefore liable to it. However, the attacker is not *fully* liable to the additional harm because they do not forfeit their right to be rescued. On Quong's view, the attacker is therefore *partially* liable to the additional harm:

6 We mention responsibility for threats primarily for illustrative purposes. Our objection does not essentially rely on the moral responsibility account of liability, which is defended by Jeff McMahan ("The Basis of Moral Liability to Defensive Killing"), among others.

though they have forfeited their right not to be proportionately harmed in self-defense, they still retain the right to be rescued from such harm.[7]

So in Conflict, Quong's account is committed to the claim that Albert is liable to the additional harm of 2 in the sense that he has forfeited his right not to be harmed proportionately—this is true regardless of whether 2 saves Betty as a side effect (i.e., regardless of whether the additional harm of 2 is gratuitous). Albert is not liable to the additional harm of 2 only in the sense that he has not forfeited his right to be rescued. But as we have already seen, the right to be rescued does not explain why 2 is impermissible. Similar remarks apply to Attack (Extra Threat).

Alternatively, Quong could respond to our objection by embracing the implication of his account that 2 in Conflict is permissible. In defense of this claim, he might argue that we can consider the opportunity costs of rescuing a liable agent. For even if a liable agent cannot forfeit her right to be rescued, perhaps this right can diminish in strength relative to others' rights to be rescued. While such a consideration would favor 2 over 1 in Conflict, appealing to it would go significantly beyond Quong's view that the necessity condition is explained simply by the right to be rescued. First, Quong himself rejects the claim that a liable agent's right to be rescued can diminish in cost-requiring strength, and so he cannot offer such a response to our objection based on Attack (Extra Threat).[8] Second, at least without some further explanation, it would seem implausible that a liable agent's right to be rescued cannot diminish in cost-requiring strength *but can* diminish in strength relative to others' rights to be rescued. This suggests Quong cannot defend the permissibility of 2 in Conflict in the way proposed.

Underlying our intuitions about both Conflict and Attack (Extra Threat) is the thought that a liable agent only forfeits their rights against being harmed *for the purpose of* preventing threats for which they are responsible or which they intend. For example, since Albert is not responsible for the independent threat to Betty posed by the trolley in Attack (Extra Threat), he is not liable to the additional harm of averting this threat involved in 2. Conflict is similar in this regard.[9] We have added "or which they intend," as there might be cases in

7 On the connection between necessity and liability, see Quong, *The Morality of Defensive Force*, 145–48.

8 See Quong, *The Morality of Defensive Force*, 143–44.

9 Intuitively, Albert would forfeit this right if he were responsible for the threat to Betty. To see this, consider

> *Conflict (Double Attack)*: Albert is wrongfully attempting to kill you and sever Betty's right arm and legs. You can (1) press a button that severs Albert's left arm, stopping Albert's attack on you without stopping his attack on Betty; (2) press a

which an agent forfeits their rights against being harmed to prevent a threat when they intend to bring about the same harm as this threat, even if they are not responsible for the threat itself.[10] But notice that in Attack (Extra Threat), Albert is not only not responsible for the independent threat to Betty's legs: he also does not intend that harm either (he intends only to sever her arm). Given these facts about Albert's responsibility and intentions in Attack (Extra Threat), 2 is impermissible. Conflict is similar in this regard.

Quong could amend his view to account for intuitions about cases such as Conflict and Attack (Extra Threat). However, this would involve departing from the simple and elegant idea that the duty not to impose additional harm encoded in the necessity condition just consists in a duty of rescue.[11]

In addition to our objections based on Conflict and Attack (Extra Threat), here is a third, related, point against Quong's rescue account of necessity. If the duty not to impose unnecessary harm in self-defense just is a duty of rescue, factors that affect the stringency of duties of rescue would presumably similarly affect the stringency of the duty not to impose unnecessary harm in self-defense. Even if this is plausible with respect to factors such as the magnitude of harm prevented, it does not seem plausible with respect to other factors.

Some have argued, for instance, that if you are physically near someone you can rescue, have a direct personal encounter with them, or are the only person who can rescue them, you have an especially stringent duty to rescue this person.[12] According to some such views, if you can save four of distant *A*'s limbs or two of nearby *B*'s limbs, you are permitted to save *B*, even though if both *A* and *B* were nearby you would be required to save *A*.

But now suppose distant *A* is wrongfully attempting to kill you while a boulder is about to crush two of nearby *B*'s limbs. You can avert *A*'s attack without harming *A* or you can press a button that severs four of distant *A*'s limbs while, separately, saving nearby *B* from losing two limbs. If the duty not to impose

---

button that severs Albert's arms and legs, stopping both of his attacks; or (3) do nothing, stopping neither of Albert's attacks.

In this case, 2 seems permissible, if not required.

10 Suppose that Villain 1 and Villain 2 each independently sent a different hitman to kill Victim. Villain 2's hitman never shows up. However, Victim can prevent Villain 1's hitman from killing him only by using Villain 2 as a shield for the hitman's bullet. Perhaps Victim is permitted to do this. Even though Villain 2 is not responsible for the bullet, he intends to bring about the same harm as this threat.

11 Quong also cannot plausibly say that the right to be rescued explains our intuitions about the necessity condition at least in ordinary cases of self-defense that do not involve a conflict between an attacker's and a third party's rights to be rescued, since Attack (Extra Threat) involves no such conflict.

12 See, for example, Kamm, *Intricate Ethics*; and Woollard, *Doing and Allowing Harm*.

additional harm on *A* in self-defense just is a duty of rescue, and if distance affects duties of rescue as noted, then it would be permissible to sever four of distant *A*'s limbs while saving nearby *B* from losing two limbs. But this seems very implausible. A similar point applies to other factors, such as whether you have a direct personal encounter with those you can rescue or are the only person who can rescue them.

We are not claiming that distance, direct personal encounter, or being a unique potential rescuer do affect the stringency of duties to rescue. We claim only that if Quong's rescue account of necessity is correct, then these factors would presumably similarly affect the stringency of the duty not to impose unnecessary harm in self-defense, and it seems very implausible that these factors would have such effects. At least, this seems significantly less plausible than the view that these factors affect duties of rescue. This is further evidence against Quong's account of necessity.[13]

University of St Andrews
jj73@st-andrews.ac.uk
tgp4@st-andrews.ac.uk

REFERENCES

Kamm, F. M. *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*. New York: Oxford University Press, 2007.

McMahan, Jeff. "The Basis of Moral Liability to Defensive Killing." *Philosophical Issues* 15, no. 1 (October 2005): 386–405.

Quong, Jonathan. *The Morality of Defensive Force*. Oxford: Oxford University Press, 2020.

Ramakrishnan, Ketan H. "Treating People as Tools." *Philosophy and Public Affairs* 44, no. 2 (Spring 2016): 133–65.

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

13 For helpful comments, we are grateful to Joseph Bowen, Cécile Fabre, and an anonymous referee for the *Journal of Ethics and Social Philosophy*.

Journal of Ethics and Social Philosophy
VOL. 25, NO. 2 · AUGUST 2023
https://doi.org/10.26556/jesp.v25i2.2613

CIVIL DISOBEDIENCE AND ANIMAL RESCUE

A REPLY TO MILLIGAN

Daniel Weltman

TONY MILLIGAN argues that some forms of covert nonhuman animal rescue (hereafter "animal rescue"), wherein activists anonymously and illegally free nonhuman animals from confinement, should be understood as acts of civil disobedience.[1] However, most traditional understandings of civil disobedience require that the civil disobedient act publicly rather than covertly and thus rule out animal rescue.[2] Milligan's argument is part of a larger project to widen the scope of civil disobedience.[3] I argue that at least insofar as animal rescue is concerned, we ought not to widen civil disobedience's scope. Animal rescue ought instead to be classed elsewhere under the broad notion of "resistance."

Milligan highlights three reasons why civil disobedience is not supposed to be covert and attacks all three of them. The first reason is that "civil disobedience cannot take the form of action which is intrinsically suspect, and

1 Milligan, "Animal Rescue as Civil Disobedience" and *Civil Disobedience*.

2 Bedau, "On Civil Disobedience," 656; Lang, "Civil Disobedience and Nonviolence," 156; Smart, "Defining Civil Disobedience," 256; Rawls, *A Theory of Justice*, 320–21; Regan, *Empty Cages*, 194; Mancilla, "Noncivil Disobedience and the Right of Necessity"; Celikates, "Rethinking Civil Disobedience as a Practice of Contestation," 38; Edyvane and Kulenovic, "Disruptive Disobedience"; Allen and von Essen, "Is the Radical Animal Rights Movement Ethically Vigilante?" 270; Delmas, *A Duty to Resist*, 42. By "covertly" I mean the actor carries out their actions in secret and does not subsequently reveal their identity. Thus, those who engage in "open animal rescue," which entails rescuing animals covertly but then revealing one's identity, are not acting "covertly" in the relevant sense. This usage of "covert" accords with Milligan's usage of the term, but it is not universal. For instance, Kimberley Brownlee uses the term "covert disobedience" to refer to cases in which one only reveals one's identity after the fact (Brownlee, "Features of a Paradigm Case of Civil Disobedience," 348–49). I believe Brownlee would agree with me that covert disobedience in the sense referred to by Milligan does not count as civil disobedience: she seems not to even compass the possibility of civil disobedience in which the disobedient never reveals their identity. William Scheuerman also notes this point about Brownlee (Scheuerman, *Civil Disobedience*, 146).

3 Milligan, *Civil Disobedience*.

there is always something intrinsically suspect about covertness."[4] The second is that "civil disobedients are protestors who *accept* the consequences of their actions and this requires that they *must* act publicly and *must* disclose their identities."[5] The third is that "civil disobedience is primarily a form of address, a form of communication which is, by its nature, a public act."[6] I will not contest Milligan's responses to the first point. The key disagreements turn on the second point about publicity and the third point about what Milligan calls "the communication thesis."[7]

1. PUBLICITY, CONSEQUENCES, AND RESISTANCE

Some argue that protesters must accept the consequences of their actions. Doing so requires publicity and identity disclosure, which would mean that animal rescue cannot be civil disobedience. Milligan resists this conclusion because accepting the consequences of one's actions "may better capture the outlook and practice of civil disobedience movements (such as [Martin Luther] King [Jr.] and [Mahatma] Gandhi) rather than the approach of ordinary participants."[8] Thus, it is not clear that civil disobedience is a practice such that "civil disobedients *must* act" publicly.[9] We should reject this argument.[10] If we abandon the idea that the civil disobedient must act publicly and accept the consequences of their actions, we risk widening the concept too much. A rejection of the requirement allows actions such as "threats of violence, covert acts of sabotage, blackmail, and even assault" to potentially count as civil disobedience, as Jennifer Welchman argues they do.[11] If civil disobedience as a term applies to this much, or even to some subset of these activities (e.g., animal rescue and similar actions such as tree spiking), it will no longer pick out a relatively distinct,

4 Milligan, "Animal Rescue as Civil Disobedience," 289.

5 Milligan, "Animal Rescue as Civil Disobedience," 291.

6 Milligan, "Animal Rescue as Civil Disobedience," 291.

7 Milligan, "Animal Rescue as Civil Disobedience," 293.

8 Milligan, "Animal Rescue as Civil Disobedience," 291.

9 Milligan, "Animal Rescue as Civil Disobedience," 291.

10 For another argument against Milligan's claim see Weltman, "Covert Animal Rescue," 68.

11 Welchman, "Is Ecosabotage Civil Disobedience?" 105. Welchman is unsure whether violence against persons and threats of such violence ought to count as civil disobedience because they "pose *perhaps* the greatest threat to sociability, so we *might* argue that both violence and threats against persons should be excluded," although she is fine with acts such as tree spiking that can result in injury so long as loggers are adequately warned of the threat ("Is Ecosabotage Civil Disobedience?" 105, emphasis added).

useful category of investigation.[12] It will be almost coextensive with the broader notion of "resistance" as articulated by a number of authors.[13]

This broader notion of resistance encompasses both the traditional categories of disobedience and also some new categories. The traditional categories include what Joseph Raz dubs "revolutionary disobedience," "civil disobedience," and "conscientious objection" and what Michael Martin dubs "conscientious wrongdoing."[14] These categories have recently been expanded to include "uncivil disobedience" and "subrevolution"—the former dispenses with one or more of the traditional requirements of civil disobedience, such as nonviolence or publicity, and the latter covers disobedience that aims to alter only part of a government rather than the entire government.[15]

One ought not to draw distinctions merely because it is possible, but this is a case where distinctions are helpful. A distinction between civil disobedience and concepts describing other forms of resistance, such as uncivil disobedience and subrevolution, helps us think more clearly about differing tactics, justifications, and responses. Resistance broadly speaking need not be nonviolent or public, nor do resistors necessarily need to accept punishment for their actions, whereas many think civil disobedience must be limited in one or more of these ways. There is no reason to think resistance's justifications are limited to sincere justifications, a limit Kimberley Brownlee places on civil disobedience.[16] There is no need to think our approach to legal punishment or penalization for civil disobedience must mirror our approach to legal punishment or penalization for other forms of resistance.[17] It is not clear that justifications for avoiding punishment

12 Milligan thinks tree spiking poses a risk of predictable harm that is "perhaps high enough to rule out any claim of civil disobedience" (*Civil Disobedience*, 114). He also points out that tree spiking "may not be more reckless than driving a car" (115). Notwithstanding this, Milligan argues that the risk is high enough such that tree spiking no longer counts as civil disobedience but could so have counted before we realized its precise degree of recklessness in 1987 (115).

13 Mancilla, "Noncivil Disobedience and the Right of Necessity"; Caney, "Responding to Global Injustice"; Finlay, *Terrorism and the Right to Resist*, 20–21; Delmas, *A Duty to Resist*; Scheuerman, *Civil Disobedience*, 140 and "Why Not Uncivil Disobedience?"; Pineda, "Civil Disobedience, and What Else?"; Lai and Lim, "Environmental Activism and the Fairness of Costs Argument for Uncivil Disobedience."

14 Raz, *The Authority of Law*, 263; Martin, "Ecosabotage and Civil Disobedience."

15 Adams, "Uncivil Disobedience"; Lai, "Justifying Uncivil Disobedience"; Weltman, "Covert Animal Rescue" and "You Say You Want Half a Revolution?"

16 Brownlee, *Conscience and Conviction*, 19–20.

17 Lefkowitz, "On a Moral Right to Civil Disobedience" and "In Defense of Penalizing (but Not Punishing) Civil Disobedience"; Brownlee, "Penalizing Public Disobedience" and "Two Tales of Civil Disobedience."

that apply to civil disobedience on its own ought to apply to resistance more broadly.[18] Revolutions and subrevolutions are appropriate in cases where civil disobedience might be inappropriate, and vice versa.[19] The same applies for any considerations one might adduce about which we might reach differing judgments with respect to other forms of resistance versus civil disobedience. Thus, although Milligan is right to claim that the mere fact that civil disobedients often act publicly is not alone a reason to think that civil obedience must be public, in light of how wide the concept becomes if we abandon the publicity requirement, we should retain it unless we have some further active reason to eliminate it.

2. COMMUNICATION

Milligan's third defense of covert civil disobedience hinges on his rejection of the communication thesis.[20] He notes that Rawls's influential approach to civil disobedience was novel mostly for its treatment of civil disobedience as "a form of communication."[21] The communication thesis is also perhaps the most enduring feature of the Rawlsian approach, as most theorists of civil disobedience have attacked one or another of Rawls's other commitments.[22] Milligan

18 Moraro, "On (Not) Accepting the Punishment for Civil Disobedience"; Weltman, "Must I Accept Prosecution for Civil Disobedience?" Moreover, Moraro's justification for dropping the consequences requirement for civil disobedience would still not allow animal rescue to count as civil disobedience, because he still accepts the publicity condition ("On (Not) Accepting the Punishment for Civil Disobedience").

19 Weltman, "You Say You Want Half a Revolution?"

20 Milligan, *Civil Disobedience*, 18–21.

21 Milligan, *Civil Disobedience*, 18. Other classic accounts, such as Habermas's (*Between Facts and Norms*, 148, 383), endorse the communication thesis. Communication is also key to many contemporary accounts. See Smith, "Civil Disobedience and the Public Sphere" and *Civil Disobedience and Deliberative Democracy*; Scheuerman, *Civil Disobedience*, 118; Lai, "Justifying Uncivil Disobedience"; Lai and Lim, Environmental Activism and the Fairness of Costs Argument for Uncivil Disobedience." The disjunctive obligation to "persuade or obey" articulated in Plato's *Crito* perhaps also presages the thesis. (Cf. Kraut, *Socrates and the State*; Irwin, "Review: Socratic Inquiry and Politics," 400–7; Penner, "Two Notes on the *Crito*," 155–66). As noted by William Herr, already in 1972 Elliot M. Zashin claimed that "a study of recent academic writing on civil disobedience . . . yields a rough consensus" on the requirement that civil disobedience "be done with intent primarily to educate or persuade the majority" (Zashin, *Civil Disobedience and Democracy*, 110; Herr, "Thoreau," 88). Zashin's and Rawls's accounts were contemporaneous, and thus it seems the communicative approach was influential even before Rawls.

22 Milligan, *Civil Disobedience*, 18. Opponents of the communication thesis include Milligan, *Civil Disobedience*, 18–21; Welchman, "Is Ecosabotage Civil Disobedience?"; and Bedau, "On Civil Disobedience." Raz has a disjunctive account of civil disobedience according to which communication is only required for one of the disjuncts (*The Authority of Law*, 263).

drops the communication thesis for various reasons.[23] But if the communication thesis is the core of the Rawlsian approach, what is left of civil disobedience once we drop it? We are left with something like Hugo Bedau's pre-Rawls account: "anyone commits an act of civil disobedience if and only if he acts illegally, publicly, nonviolently, and conscientiously with the intent to frustrate (one of) the laws, policies, or decisions of his government."[24] Bedau's account is nearly equivalent to the first disjunct in Raz's disjunctive view, according to which civil disobedience is "a politically motivated breach of the law designed either to contribute directly to a change of a law or of a public policy or to express one's protest against, and dissocation [*sic*] from, a law or a public policy."[25] Bedau includes, whereas Raz omits, the nonviolence requirement.

One worry is that, for reasons defended by others, we may want to drop some of these requirements, such as nonviolence.[26] Milligan himself wants to drop publicity and nonviolence. Because Bedau's definition includes publicity, it rules out animal rescue. But let us grant that we can salvage something like the Bedau account and apply it to animal rescue. What makes this disobedience civil? For Bedau, "the pun on 'civil' is essential; only nonviolent acts thus can qualify."[27] Nonviolence as Bedau understands it rules out property damage of the sort Milligan explicitly compasses, and Milligan explicitly aims to avoid ruling out "surprising forms of violence that were not envisaged when we accepted the claim that civil disobedience must be non-violent or largely and aspirationally non-violent."[28] So, what can civility amount to if it does not amount to the communication thesis, or to nonviolence, or to publicity, or to a combination of these things, as Bedau himself thought?

23 He describes his view as "a *civility-focused* account by contrast with a *communication-based* account" (Milligan, *Civil Disobedience*, 37).

24 Bedau, "On Civil Disobedience," 661.

25 Raz, *The Authority of Law*, 263. Raz specifies that he focuses only on "morally motivated," or in other words conscientious, disobedience (263).

26 Raz, *The Authority of Law*, 268; Morreal, "The Justifiability of Violent Civil Disobedience"; Moraro, "Violent Civil Disobedience and Willingness to Accept Punishment," "Respecting Autonomy Through the Use of Force," and "Is Bossnapping Uncivil?"

27 Bedau, "On Civil Disobedience," 656.

28 Milligan, *Civil Disobedience*, 150. Milligan thinks that "premeditated violence" disqualifies something from counting as civil disobedience, although it is unclear why (22). Perhaps it is because premeditated violence is "difficult to reconcile with any familiar understanding of civil disobedience" (55). But Welchman's approach compasses blackmail and other actions equally difficult to reconcile, and Milligan endorses Welchman's arguments (Milligan, *Civil Disobedience*, 20). So, it is not clear how Milligan can reject her conclusion. See also Milligan, *Civil Disobedience*, 135–36.

It is not quite clear. Milligan *is* clear that there are "basic norms" of civility that a protest "must not violate or break *beyond a certain point* if it is to stay within civil bounds," including respect for others, the rejection of hate speech, "the largely successful commitment to *try* to avoid violence and threats of violence," and others, although it is not obvious what the "certain point" is or how successful one must be in order to be "largely" successfully committed to "trying" to avoid violence.[29] He is also clear about wanting civil disobedience to encompass more than just "*indirect* civil disobedience," which aims at communication and which thus forms the basis of the Rawlsian approach.[30] It should also include "a certain kind of direct action in which communication plays (at most) a subordinate role," as it did in "the Civil Rights Movement," which "primarily involved what King openly referred to as direct action (not lovingly addressed to the conscience of the opponent but aimed instead at embarrassing the Federal Government into enforcing its laws)."[31] Similarly, the Indian independence movement "involved both indirect protest … as well as direct action."[32] Thus, Milligan asks, "why not, for example, embrace a disjunctive account such that civil disobedience can be *either*" communicative or direct action?[33] But it is not clear how rhetorical the question is. Milligan says that "a disjunctive approach to the concept looks promising," but whether it looks promising enough to adopt is left unstated.[34] Such a disjunctive account would return us to Raz's view, which Milligan does not discuss.

At other points, however, Milligan, unlike Raz, seems to want to abandon the communicative part entirely: "the retention of the [communicative] thesis risks turning civil disobedience into an endangered concept" because it lends weight to "the argument that civil disobedience is overly deferential to authority," such that activists may abandon the concept entirely.[35] (Thus, perhaps the question about the disjunctive approach was a genuine question, and the answer is that we should discard the communicative requirement.) Why would we worry about activists abandoning the concept? That is, what is wrong if activists by definition turn out to not be engaging in civil disobedience when they engage in animal rescue? The answer is that "no other concept carries the

29 Milligan, *Civil Disobedience*, 36. For an objection about whether this is an alternative to the communicative approach see Scheuerman, *Civil Disobedience*, 145.

30 Milligan, "Animal Rescue as Civil Disobedience," 295.

31 Milligan, "Animal Rescue as Civil Disobedience," 295.

32 Milligan, "Animal Rescue as Civil Disobedience," 295.

33 Milligan, "Animal Rescue as Civil Disobedience," 295.

34 Milligan, "Animal Rescue as Civil Disobedience," 296. It is also in tension with his earlier claim (noted above in note 23) that his approach opposes communicative views.

35 Milligan, "Animal Rescue as Civil Disobedience," 296.

moral authority of 'civil disobedience', and none is likely to do so for the foreseeable future. . . . What is then in danger of being lost is the ongoing relevance of a concept of protest that still has a great deal of work left to do."[36]

Milligan's argument here begs the question, because whether the concept has a great deal left to do with respect to animal rescue hinges on whether it accurately describes animal rescue.[37] But, more relevantly, if we drop the communicative thesis and everything else Milligan seems to want to drop, it is hard to know what is left in the "civil" part of civil disobedience. Milligan thinks that civility and communication are contrasting focuses rather than that the latter constitutes an elaboration of the former.[38] But Milligan's notion of civility allows rather wide leeway for engaging in what we might think of as uncivil behavior, such as theft and property destruction, because civility only entails respecting people "as persons" rather than "as racists or as anti-Semites" or as other sorts of things.[39] Given that one need only be *largely* successful in trying to avoid violence, the door is open for the occasional violent failure to count as civil disobedience. And Milligan is wary even of using this minimal notion of civility for categorization purposes, because "to couch matters in terms of civility . . . may be a questionable basis upon which" to determine what *counts* as civil disobedience.[40] Given that we have the wider category of resistance, we do not need to widen civil disobedience as much as Milligan (or Raz) would have us do. We have space for uncivil forms of resistance that we can use to discuss all sorts of behavior rather than labeling it civil disobedience. So, we might think that if the civility requirement is abandoned, then we have changed the subject to other forms of resistance rather than enlarged our concept of civil disobedience.

However, Milligan still thinks we should retain the civility requirement: he suggests that if civil disobedience were about communication rather than civility, it would be hard to explain the actions of those who, like Henry David Thoreau, display "an unwillingness to suspend illegal activism in return for a proper hearing as, perhaps, they *ought* to do if they view civil disobedience as communication."[41] As I have argued elsewhere, it is not clear that the sort of hearings available to disobedients are "proper."[42] But even granting this point,

36 Milligan, "Animal Rescue as Civil Disobedience," 296.

37 See also my arguments in Weltman, "Covert Animal Rescue," 69–70.

38 Milligan, *Civil Disobedience*, 13.

39 Milligan, *Civil Disobedience*, 17. It is not clear how exactly we are meant to slice people into the parts we have to respect and the parts we do not.

40 Milligan, *Civil Disobedience*, 17.

41 Milligan, *Civil Disobedience*, 146.

42 Weltman, "Must I Accept Prosecution for Civil Disobedience?"

it works equally well against Milligan in this case. The fact that animal rescuers would not be satisfied with a proper hearing suggests they are not engaged in civil disobedience.

3. CONCLUSION

Milligan's arguments for labeling animal rescue as civil disobedience are undermotivated and face powerful objections. It would therefore be better to label animal rescue as some form of resistance other than civil disobedience and to reserve the term for actions that more clearly fit the bill (e.g., open rescue). Whether this spells trouble for Milligan's attempts to widen civil disobedience more broadly is a further topic, but insofar as we are concerned with animal rescue, we ought to refrain from widening the concept.[43]

Ashoka University
danny.weltman@ashoka.edu.in

REFERENCES

Adams, N. P. "Uncivil Disobedience: Political Commitment and Violence." *Res Publica* 24, no. 4 (November 2018): 475–91.

Allen, Michael, and Erica von Essen. "Is the Radical Animal Rights Movement Ethically Vigilante?" *Between the Species* 22, no. 1 (Fall 2018): 260–85.

Bedau, Hugo A. "On Civil Disobedience." *Journal of Philosophy* 58, no. 21 (October 1961): 653–65.

Brownlee, Kimberley. *Conscience and Conviction: The Case for Civil Disobedience*. New York: Oxford University Press, 2012.

———. "Features of a Paradigm Case of Civil Disobedience." *Res Publica* 10, no. 4 (December 2004): 337–51.

———. "Penalizing Public Disobedience." *Ethics* 118, no. 4 (July 2008): 711–16.

———. "Two Tales of Civil Disobedience: A Reply to David Lefkowitz." *Res Publica* 24, no. 3 (2018): 291–96.

Caney, Simon. "Responding to Global Injustice: On the Right of Resistance." *Social Philosophy and Policy* 32, no. 1 (October 2015): 51–73.

Celikates, Robin. "Rethinking Civil Disobedience as a Practice of Contestation—Beyond the Liberal Paradigm." *Constellations* 23, no. 1 (March 2016):

43 I thank at least two anonymous reviewers and an associate editor for this journal for their very helpful questions and comments.

37–45.
Delmas, Candice. *A Duty to Resist: When Disobedience Should Be Uncivil*. New York: Oxford University Press, 2018.
Edyvane, Derek, and Enes Kulenovic. "Disruptive Disobedience." *Journal of Politics* 79, no. 4 (October 2017): 1359–71.
Finlay, Christopher J. *Terrorism and the Right to Resist*. Cambridge: Cambridge University Press, 2015.
Habermas, Jürgen. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Translated by William Rehg. Cambridge, MA: MIT Press, 1996.
Herr, William A. "Thoreau: A Civil Disobedient?" *Ethics* 85, no. 1 (October 1974): 87–91.
Irwin, T. H. "Review: Socratic Inquiry and Politics." *Ethics* 96, no. 2 (January 1986): 400–15.
Kraut, Richard. *Socrates and the State*. Princeton, NJ: Princeton University Press, 1984.
Lai, Ten-Herng. "Justifying Uncivil Disobedience." In *Oxford Studies in Political Philosophy*, vol. 5, edited by David Sobel, Peter Vallentyne, and Steven Wall, 90–114. Oxford: Oxford University Press, 2019.
Lai, Ten Herng, and Chong-Ming Lim. "Environmental Activism and the Fairness of Costs Argument for Uncivil Disobedience." *Journal of the American Philosophical Association* (forthcoming). Published online ahead of print, January 16, 2023. https://doi.org/10.1017/apa.2022.15.
Lang, Berel. "Civil Disobedience and Nonviolence: A Distinction with a Difference." *Ethics* 80, no. 2 (January 1970): 156–59.
Lefkowitz, David. "In Defense of Penalizing (but Not Punishing) Civil Disobedience." *Res Publica* 24, no. 3 (August 2018): 273–89.
———. "On a Moral Right to Civil Disobedience." *Ethics* 117, no. 2 (January 2007): 202–33.
Mancilla, Alejandra. "Noncivil Disobedience and the Right of Necessity: A Point of Convergence." *Krisis*, no. 3 (2012): 3–15.
Martin, Michael. "Ecosabotage and Civil Disobedience." *Environmental Ethics* 12, no. 4 (Winter 1990): 291–310.
Milligan, Tony. "Animal Rescue as Civil Disobedience." *Res Publica* 23, no. 3 (August 2017): 281–98.
———. *Civil Disobedience: Protest, Justification, and the Law*. New York: Bloomsbury Academic, 2013.
Moraro, Piero. "Is Bossnapping Uncivil?" *Raisons Politiques* 69, no. 1 (2018): 29–44.
———. "On (Not) Accepting the Punishment for Civil Disobedience." *The*

Philosophical Quarterly 68, no. 272 (July 2018): 503–20.
———. "Respecting Autonomy through the Use of Force: The Case of Civil Disobedience." *Journal of Applied Philosophy* 31, no. 1 (February 2014): 63–76.
———. "Violent Civil Disobedience and Willingness to Accept Punishment." *Essays in Philosophy* 8, no. 2 (June 2007): 270–83.
Morreal, John. "The Justifiability of Violent Civil Disobedience." In *Civil Disobedience in Focus*, edited by Hugo Adam Bedau, 130–43. London: Routledge, 1991.
Penner, Terry. "Two Notes on the *Crito*: The Impotence of the Many, and 'Persuade or Obey.'" *The Classical Quarterly* 47, no. 1 (May 1997): 153–66.
Pineda, Erin R. "Civil Disobedience, and What Else? Making Space for Uncivil Forms of Resistance." *European Journal of Political Theory* 29, no. 1 (January 2021): 157–64.
Rawls, John. *A Theory of Justice*, rev. ed. Cambridge, MA: Harvard University Press, 1999.
Raz, Joseph. *The Authority of Law: Essays on Law and Morality*. Oxford: Oxford University Press, 1979.
Regan, Tom. *Empty Cages: Facing the Challenge of Animal Rights*. Lanham, MD: Rowman and Littlefield, 2004.
Scheuerman, William E. *Civil Disobedience*. Cambridge: Polity Press, 2018.
———. "Why Not Uncivil Disobedience?" *Critical Review of International Social and Political Philosophy* 25, no. 7 (December 2022): 980–99.
Smart, Brian. "Defining Civil Disobedience." *Inquiry* 21, no. 1–4 (1978): 249–69.
Smith, William. *Civil Disobedience and Deliberative Democracy*. New York: Routledge, 2013.
———. "Civil Disobedience and the Public Sphere." *Journal of Political Philosophy* 19, no. 2 (June 2011): 145–66.
Welchman, Jennifer. "Is Ecosabotage Civil Disobedience?" *Philosophy and Geography* 4, no. 1 (2001): 97–107.
Weltman, Daniel. "Covert Animal Rescue: Civil Disobedience or Subrevolution?" *Environmental Ethics* 44, no. 1 (Spring 2022): 61–83.
———. "Must I Accept Prosecution for Civil Disobedience?" *The Philosophical Quarterly* 70, no. 279 (April 2020): 410–18.
———. "You Say You Want Half a Revolution? The Ethics of Subrevolution." Unpublished manuscript.
Zashin, Elliot M. *Civil Disobedience and Democracy*. New York: Free Press, 1971.

JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY
http://www.jesp.org
ISSN 1559-3061

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.