JOURNAL *of* ETHICS
& SOCIAL PHILOSOPHY

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.

CAN "MORE SPEECH" COUNTER IGNORANT SPEECH?

Maxime Lepoutre

DEMOCRATIC PUBLIC DISCOURSE is rife with *ignorant speech*—speech that disseminates or promotes falsehoods. Some of this speech comes from speakers who are themselves ignorant and believe the falsehoods they utter. Very often, however, the sources of ignorant speech, as defined above, are not themselves ignorant; rather, they knowingly spread ignorant views and perspectives for personal or political gain.

Ignorant speech is a very broad class. It can spread many different kinds of falsehoods that bear on many different kinds of things. But two species of ignorant speech are particularly noteworthy, as they have proven to be particularly dangerous in contemporary democracies. The first consists in ordinary *political misinformation*, where speakers circulate falsehoods concerning policy-related issues. Whether it is bloggers claiming that the MMR vaccine causes autism, senior politicians asserting that Obamacare would implement life-threatening "death panels," or public figures insisting that exiting the European Union would save the United Kingdom (and its National Health Service) £350 million a week, there is no doubt that policy debates are rife with such misinformation.[1]

The second noteworthy category spreads falsehoods not about policies but about people. Specifically, this species of ignorant speech, which constitutes a type of *hate speech*, puts forward representations of other social groups that falsely deny their basic equality—often by advancing vilifying or degrading characterizations of them. A paradigmatic illustration here is speech that activates distorted stereotypes that present minorities as essentially subhuman or morally inferior: for instance, that refugees are "a plague" of "cockroaches," that Muslims are terrorists, or that Mexican immigrants are "criminals" and "rapists."[2]

The pervasiveness of these two forms of ignorant speech constitutes a serious

1   Godlee, Smith, and Marcovitch, "Wakefield's Article Linking MMR Vaccine and Autism Was Fraudulent"; Holan, "PolitiFact's Lie of the Year"; Sparrow, "UK Statistics Chief Says Vote Leave £350m Figure Is Misleading."

2   Nelson, "Katie Hopkins Reflects on Branding Migrants 'Cockroaches' and 'Feral Humans'";

political problem: policy misinformation and hateful forms of ignorant speech, it has been argued, are both liable to generate grave injustices. The former risks misleading the public into supporting unjust policies or harmful candidates.[3] As for the latter, it is capable of producing a wide array of unjust harms, such as inciting animosity, causing psychological trauma, undermining targets' sense of dignity, subordinating targets, or enacting discriminatory norms.[4]

A celebrated response to this problem recommends countering ignorant utterances with informed and truthful speech. This is of course familiar from debates in democratic theory: "deliberative" theorists of democracy have long advocated inclusive deliberation, and have increasingly done so on epistemic grounds. On their view, exchanging arguments, voicing personal narratives, and soliciting expert testimony are powerful ways of debunking political misinformation and thus correcting false political beliefs.[5]

This response is not specific to ordinary political misinformation: "counterspeech" is also an extremely popular remedy for hateful forms of ignorant speech. In particular, opponents of legal restrictions on hate speech typically insist that we should counter hate speech verbally rather than legally.[6] In this context, counterspeech involves contesting the distorted contents that hate speech asserts, presupposes, or encodes, so as to block whatever harms their diffusion would otherwise give rise to.

Despite this popularity, the counterspeech response to ignorant speech has come under heavy criticism, especially in the context of hate speech. According to one influential objection, advocating counterspeech as the remedy for hate speech is *unfair*, as it "places the burden of the remedy on those targeted (and potentially harmed) by the allegedly harmful speech."[7] Relatedly, others have worried that, because it demands too much of targets of hate speech, counterspeech also risks being *ineffective.* For one thing, there is evidence that targets

---

Asser, "What the Muhammad Cartoons Portray"; Reilly, "Here Are All the Times Donald Trump Insulted Mexico."

3   Brown, "Propaganda, Misinformation, and the Epistemic Value of Democracy."

4   Delgado, "Words that Wound"; Waldron, *The Harm in Hate Speech*; Maitra, "Subordinating Speech"; McGowan, *Just Words*, ch. 7.

5   See, e.g., Young, *Inclusion and Democracy*, chs. 1–3; Anderson, "The Epistemology of Democracy"; and Landemore, *Democratic Reason*.

6   See, e.g., Louis Brandeis's opinion in *Whitney v. California* (274 U.S. 357 (1927)); Post, *Constitutional Domains*; Brettschneider, *When the State Speaks, What Should It Say?* ch. 3; Lepoutre, "Hate Speech in Public Discourse"; and Strossen, *Hate*, ch. 8.

7   McGowan, "Responding to Harmful Speech," 183. See also Maitra and McGowan, "Introduction and Overview," 9.

rarely respond to such speech.[8] And even if targets did speak back, they may lack the authority needed to do so in an effective or compelling way. This is both because targets generally belong to relatively vulnerable social groups, and because, as feminist philosophers of language have argued, hate speech can itself disable (or "silence") its targets' speech.[9]

However, these two preliminary concerns are not decisive. According to an increasing range of defenders of counterspeech, counterspeech can be thought of as a state-driven practice, rather than as something merely to be performed by private individuals. On this understanding, which Corey Brettschneider and Katharine Gelber have prominently articulated, it is primarily, though not exclusively, the state's responsibility to speak out against hate speech. The state has numerous tools for doing this, such as having high-profile politicians denounce vilifying utterances, designing school syllabuses, erecting historical monuments, or funding civil rights groups to challenge hate speakers.[10]

This adjusted conception of who should engage in counterspeech seems a promising way of alleviating the two concerns outlined above. Recommending state-sponsored counterspeech is less *unfair*, as it shifts the burden of responding away from targets of hate speech. Moreover, the state is more likely than vulnerable targets to have the *authority* needed to be taken seriously when it opposes hate speech.

But even with this adjustment, the policy of countering ignorant speech with more speech still encounters a deep problem. In her influential discussion of oppressive speech, Mary Kate McGowan has argued that the counterspeech response operates with a naive conception of how language works. In particular, it overlooks the asymmetric pliability of conversational norms—the phenomenon whereby it is easier to enact conversational norms than it is subsequently to undo those norms. Because of this phenomenon, the damaging effects of ignorant speech on public discourse are difficult to reverse, or "sticky." Accordingly, McGowan concludes, trying to undo the harmful effects of ignorant speech through more speech seems as futile as trying to "unring a bell."[11]

Crucially, this problem of stickiness stems from facts about the dynamics of

---

8   Nielsen, "Power in Public."

9   On the authority problem and silencing, see Maitra and McGowan, "Introduction and Overview," 10; McGowan, "Responding to Harmful Speech," 183–84; Langton, "Speech Acts and Unspeakable Acts." For a positive account of the authority requirement needed to successfully challenge hate speech, see Barnes, "Speaking with (Subordinating) Authority," 251–56.

10   Brettschneider, *When the State Speaks, What Should It Say?*; Gelber, "Reconceptualizing Counterspeech in Hate Speech Policy (with a Focus on Australia)."

11   McGowan, "Oppressive Speech," 403.

language and conversational norms, rather than about who is doing the talking. *Prima facie*, then, it is just as applicable to state counterspeech as to non-state counterspeech. In fact, as we will see, there are reasons to think that it is *especially* problematic for state-driven counterspeech. What this means is that an increasingly popular way of defending counterspeech against criticism—namely, focusing on state-driven counterspeech—will not help with this problem.

In this article, I will examine the stickiness of ignorant speech more closely, with two aims. The first is to show that the asymmetric pliability of conversational norms is an even more general problem than McGowan suggests, and that, as a result, it stands in the way of verbally countering both hateful and non-hateful forms of ignorant speech. The second is to articulate a more sophisticated account of counterspeech, which is distinctively suited to overcoming this problem. Doing so will help us arrive at a normative ideal of public discourse that is more finely attuned to nonideal conditions.

My argument will proceed as follows. After introducing the problem of stickiness, I demonstrate that it is a very general phenomenon, which applies not just to the oppressive or hateful speech that McGowan analyzes, but also to ordinary political misinformation (section 1). I then develop a conception of counterspeech that is directly informed by this difficulty. Specifically, I argue that, by distinguishing between positive and negative forms of counterspeech (section 2) and adopting a more nuanced view of counterspeech's temporality (section 3), we can substantially mitigate the problems generated by sticky ignorance. I conclude by examining the implications of this argument for legal responses to ignorant speech (section 4).

Before proceeding, two important clarifications are in order. First, note that my purpose here is *not* to conclude that counterspeech is a sufficient remedy for ignorant speech, so that legal responses to such speech—such as bans on hate speech or on fake news—are unwarranted. I will be exploring and challenging a specific argument against counterspeech. But this does not necessarily entail opposing legal remedies. To begin, there might be other, as yet uncanvassed, problems for counterspeech. Moreover, my argument should also be useful to those who defend legal remedies. Many proponents of legal remedies accept that counterspeech nonetheless also has a role to play in combatting hateful or deceptive speech.[12] To determine what the division of labor should be between legal and speech-based remedies, such theorists need to determine more precisely what problems are faced by counterspeech but avoided by bans. As will become apparent in section 4, my argument will advance this understanding: in

---

12   E.g., Gelber, "Reconceptualizing Counterspeech in Hate Speech Policy (with a Focus on Australia)," 214; Langton, "Blocking as Counterspeech."

suggesting that a revised conception of counterspeech can substantially weaken the stickiness of ignorant speech, I am debunking one hypothesis regarding what legal remedies can do that counterspeech cannot.

Second, we should pause to examine the appropriateness of considering hate speech and political misinformation together under the heading of ignorant speech. One might be wary of doing so on various grounds. For one thing, one might point out that not all hate speech involves disseminating or promoting falsehoods. In particular, some instances of hate speech attack vulnerable groups via utterances that are true (the leader of a racist party who shouts "We want you out!"), while other instances of hate speech, like imperatives, are not truth-apt ("Get out of this country!"). Thus, strictly speaking, these types of hate speech do not count as ignorant speech, as I define it.

Because my focus is on how we might verbally counter speech that produces injustice or harms in virtue of spreading or promoting falsehoods, I will be bracketing these further types of hate speech for the purposes of this paper. That is, by "hate speech," I will largely be talking about speech that puts forward representations of its targets that incorrectly deny their basic equality (such as racist or xenophobic stereotypes that present their targets as essentially degraded or vile). While this constitutes a limit to the scope of my counterspeech-related recommendations, it does not render that scope insignificant: after all, an important portion of hate speech *does* produce bad effects (e.g., psychological distress, damage to targets' dignity, stirring up animosity) by broadcasting distorted equality-denying stereotypes.[13]

Still, even with this qualification, one might think that hateful and non-hateful forms of ignorant speech remain importantly different: as discussed above, their dissemination of falsehoods may well generate harms or injustices in different ways. Political misinformation typically does so by instilling false beliefs in listeners, which in turn may induce them to embrace unjust policies or dangerous candidates. Although speech that describes vulnerable groups in degrading ways may also induce false beliefs, this is far from the only way in which it can produce injustice: to reiterate, publicizing such hateful representations can also, among other things, assault its targets' dignity, subordinate them, or enact discriminatory norms.

---

13   What is the payoff of this restriction? As I will explain below, part of my aim in considering political misinformation alongside hate speech is to bring the rich philosophical analyses of conversational stickiness developed in the context of hateful speech to bear on our understanding of political misinformation. With this aim in mind, it makes sense to restrict my attention to cases of hate speech that are closer in form to political misinformation, in that—*qua* ignorant speech—they involve spreading or promoting falsehoods.

I do not wish to minimize this difference. My point is rather that, nevertheless, hateful and non-hateful forms of ignorant speech remain similar in at least one important way: as I will explain in section 1, their damaging conversational effects are "sticky," in a way that makes them difficult to rebut via counterspeech. And, when exploring this similarity in section 1.2, I will consider how it may be complicated by the fact that hateful and non-hateful forms of ignorant speech produce harms or injustices via different mechanisms.

This similarity is important, and reveals why it is fruitful to consider hateful and non-hateful forms of ignorant speech side by side. On the one hand, the rich theoretical insights that philosophers of language such as McGowan have developed when thinking about the stickiness of hateful speech can help us understand why and when policy-related misinformation may also be resistant to counterspeech. Because they have traditionally been insensitive to this stickiness, democratic theories that advocate policy-related deliberation would particularly benefit from these insights. But the fruitful relationship also goes in the other direction. Social psychologists have conducted numerous experiments on political misinformation and its correction. Insofar as hateful and non-hateful ignorant speech involve a similar kind of stickiness, this empirical evidence can, I will suggest, inform our verbal responses to hateful as well as non-hateful ignorant speech.

1. THE STICKINESS OF IGNORANT SPEECH

1.1. The Asymmetric Pliability of Conversational Norms

Utterances can affect and alter conversations in countless ways. Among many other things, they can share new information (or bring it into the "common ground"); they can call accepted views into question; they can introduce new hypotheses or ideas, thereby making them more salient or relevant to the conversation; and they can explicitly invoke new conversational standards or rules.

Whenever utterances do one of these things, they alter the norms that govern the conversation. This is clearest in the case of speech that explicitly invokes a new rule for the conversation with the intention of enacting it. Suppose a classroom participant says, "From now on, people must raise their hand before speaking." Provided the speaker has sufficient authority (she is the instructor, rather than an overzealous auditor), her utterance can enact a new conversational norm—namely, the norm that people must raise their hand before speaking.[14]

However, this is only a special case of how utterances alter conversational

---

14 McGowan, "Oppressive Speech," 393–94.

norms. As McGowan has shown, even contributions that neither explicitly invoke nor intend to enact norms nonetheless alter the norms that govern the conversation. When an utterance shares new information, for instance, this makes it more permissible for subsequent utterances to presuppose that information. For example, it is more conversationally appropriate for *A* and *B* to discuss where the ceremony is taking place *after* they have told their interlocutors that they are getting married. Similarly, making a new hypothesis salient partly changes the topic of the conversation: that hypothesis becomes relevant to the conversation, so that participants can appeal to it without seeming out of line. If police officers investigating the disappearance of Sheriff Smith's hat are listing known thieves, mentioning Deputy Jones (who is not a known thief) would seem inappropriate. But doing so becomes more conversationally appropriate if a participant raises the possibility that the theft was an inside job. McGowan's general point, then, is that "our utterances routinely change what is permissible around us."[15]

But even if utterances routinely change conversational norms, not all norm changes are equally easy to bring about. In particular, some norms are easier to introduce than to reverse. More specifically, alongside other philosophers of language, McGowan suggests that the norms enacted by oppressive speech (such as hate speech) are "especially difficult to undo."[16] To illustrate this asymmetry, imagine that a public figure asserts "*X*s are lazy parasites." Verbally countering this assertion—for instance, by saying "That's false! *X*s are *not* lazy parasites!"—may well be ineffective at reversing the initial utterance's nefarious conversational effects. Intuitively, negating the vilifying proposition in this way would only strengthen its conversational relevance: whether or not *X*s are lazy parasites would become even more clearly the topic of conversation and, consequently, the social standing of *X*s would be even more evidently at issue. On the basis of cases like these, McGowan hypothesizes that the norms introduced by oppressive speech are asymmetrically pliable—verbally introducing them is easier than verbally reversing them.

Elaborating on McGowan's hypothesis, Robert Simpson has argued that the asymmetric pliability of the norms enacted by oppressive or hateful speech results from a more general asymmetry, which concerns *salience*. Generally speaking, while we can easily make an association salient by mentioning it and draw-

---

15   McGowan, "Oppressive Speech," 406; see also 394–97.

16   McGowan, "Responding to Harmful Speech," 189; see also "Oppressive Speech," 403. McGowan is building on Lewis's ("Scorekeeping in a Language Game") observation that raising the standards for using vague terms like "flat" is easier than lowering them. Stanley (*How Propaganda Works*, 157–61) and Langton ("Blocking as Counterspeech," sec. 6) echo the point that oppressive conversational norm changes are harder to reverse.

ing people's attention to it, it is much more difficult subsequently to make it less salient.[17] This is because rejecting an association may inadvertently make one's interlocutors pay even more attention to it.

This salience-based asymmetry helps explain why the distorted representations advanced by oppressive hate speech are so resistant to counterspeech. The initial utterance makes a vilifying association salient to listeners (say, between being an *X* and being a lazy parasite). This increased salience adjusts the bounds of the conversation, so that the hateful association becomes relevant to it. In other words, by making this association salient, the hateful utterance introduces the following conversational norm: it makes it more permissible for people subsequently to debate whether or not *X*s are in fact lazy parasites. In this context, trying to repudiate the association risks amplifying its salience and thereby reinforcing the conversational norm initially introduced.[18] That is, exclaiming "*X*s are *not* lazy parasites!" may challenge the claim *that X*s *are parasites.* But even as it does so, it risks magnifying the salience of this stereotype within the conversation. And, in turn, saliently denying the vilifying stereotype reinforces its centrality as a topic of debate: it implicitly reaffirms the harmful norm that it is permissible publicly to discuss whether or not *X*s are lazy parasites.[19]

In sum, the distinctive properties of conversational salience make the harmful effects of hateful speech sticky, in that they are more difficult to reverse than to enact. The natural response to this problem is pessimism about counterspeech. McGowan, as mentioned earlier, famously concludes that counterspeech is as futile as "trying to unring a bell."[20] Just as trying to unring a bell is likely to make it ring even more loudly, so trying to reverse the effects of hateful speech may simply amplify those effects. Simpson likewise expresses pessimism: for activated associations "to become *un*-salient," he observes, "we will simply have to change the subject ... or allow time to elapse."[21]

Importantly, this problem is especially worrisome given how the policy of

---

17   Simpson, "Un-Ringing the Bell," 572.

18   Simpson, "Un-Ringing the Bell," 569–70. While McGowan and Simpson focus on oppressive speech generally, this phenomenon also arises in more specific debates. When discussing hate speech that takes a coded form, Stanley (*How Propaganda Works*, 159) worries that verbally countering such speech requires exposing the code, and thereby raising the salience of its hateful contents. Moreover, examinations of slurs have shown that slur terms preserve their bad effects even when they are negated: "*A* is not an *X*" (e.g., Langton et al., "Language and Race," 756–57). Here, too, salience may have an explanatory role: negating a slur involves repeating it, which increases the salience of the degrading stereotype it expresses.

19   In section 1.2, I explain in greater detail why this conversational norm is harmful.

20   McGowan, "Oppressive Speech," 403.

21   Simpson, "Un-Ringing the Bell," 569.

countering hateful speech with more speech has recently been defended. As discussed above, in response to concerns regarding the fairness and authority of counterspeech, its proponents have increasingly argued that counterspeech should primarily be performed by the state. However, we can now see that embracing state-driven counterspeech does not alleviate the stickiness problem. If anything, it worsens it. In virtue of being highly authoritative, state speech is capable of significantly influencing the norms of public discourse on a large scale. Precisely because state speech is so powerful, what the state says is typically highly salient. What this means, however, is that having the state repudiate vilifying representations risks increasing their salience all the more. Thus, the objection under consideration is so problematic for state-driven counterspeech because it works by harnessing the power of counterspeech against it: since the state has an unparalleled ability to set the agenda of public discourse, it also has an unparalleled tendency to strengthen the salience of vilifying representations when challenging them.

1.2. Generalizing the Objection

Philosophers have almost exclusively discussed the problem of stickiness within debates about oppressive speech or hate speech—i.e., speech that casts its targets as fundamentally inferior.[22] But the linguistic phenomenon that underpins this problem is quite general. To see this, recall that, according to Simpson's analysis, the difficulty of reversing the conversational norms enacted by oppressive speech is partly explained by the fact that it is easier to make an association salient than unsalient. Now, this asymmetry is not inherently sensitive to the specific content of the association.[23] It stems from properties of salience, and any content can in principle be salient. Accordingly, the salience-related asymmetry Simpson identifies applies not simply to hateful stereotypes, but also to associations that spread misinformation about policy.

As a result, it is also true of counterspeech directed at policy misinformation that it risks amplifying rather than reversing the salience of its target utterance. This observation is fairly intuitive. The more the Centers for Disease Control (CDC) asserts that "vaccines do not cause autism," the more conversationally salient the misleading association between vaccines and autism risks becoming.[24]

---

22 Simpson does briefly note when concluding that the asymmetric pliability of conversational norms may apply beyond the context of oppressive speech ("Un-Ringing the Bell," 572–74). But his purpose in doing so is not to establish that ordinary political misinformation, specifically, can be sticky; and his focus in the vast majority of the paper is on oppressive speech.

23 Simpson, "Un-Ringing the Bell," 573.

24 Centers for Disease Control and Prevention, "Vaccines Do Not Cause Autism."

Similarly, insisting, as the anti-Brexit "Remain" campaign did, that the UK did not send the EU £350 million each week, risked drawing attention to that rumor.[25] The point, then, is that insofar as the problem of stickiness results from a general salience-based asymmetry, we have every reason to think that this phenomenon applies to ordinary political misinformation as well.

But more needs to be said. So far, I have suggested that, as with hateful forms of ignorant speech, it is easier to make the associations advanced by ordinary political misinformation salient than unsalient. The second step is to suggest that the salience of political misinformation is connected to its morally problematic effects. Indeed, unless this second step holds, a skeptic might raise the following concern: "Even if it is hard to make political misinformation unsalient through counterspeech, this is not morally problematic. After all, the potential harmful or unjust effects of political misinformation are not strongly connected to, and do not readily ensue from, the mere salience of misinformation. So, even if counterspeech exacerbates the salience of misinformation, it might nevertheless succeed in discrediting it and disabling its harmful conversational effects." To establish that the salience-based asymmetry Simpson diagnoses truly is a problem for countering political misinformation with more speech, we must therefore identify a meaningful connection between the salience of misinformation and its harmful or unjust conversational effects.

Now, because hateful and non-hateful forms of ignorant speech lead to harms or injustices in different ways (and very likely to a different extent), the specific answer we give to this second question is likely to differ from the answer we give in the case of hateful forms of ignorant speech. In the case of hate speech, Jeremy Waldron has powerfully argued that there exists a tight, constitutive connection between conversational salience and injustice. According to Waldron, justice requires that citizens be assured of their standing as social and political equals.[26] Indeed, unless it is visible to citizens that they can count on having their rights as equal members of society respected, they will be unable to participate in social and political life without fear.[27] The problem with hate speech, Waldron emphasizes, is that the mere salience of its vilifying contents in public debate threatens to undermine this very assurance.

To see why, remember that what it means for something to be conversationally salient is that it is relevant to the conversation, or included within the topic of conversation. It is, in other words, partly what the conversation is about. Consequently, when a vilifying stereotype is salient in public debate, this means

---

25 Sparrow, "UK Statistics Chief Says Vote Leave £350m Figure Is Misleading."

26 Waldron, *The Harm in Hate Speech*, ch. 4.

27 Waldron, *The Harm in Hate Speech*, 85.

that the public conversation is *about* whether its contents are accurate. What is at issue in the conversation, then, is whether or not, say, women are submissive, blacks are subhuman, or immigrants are parasites. Since these propositions are clearly inconsistent with the good status of the targeted groups, these groups are no longer assured of their good standing. Instead, their standing is precisely what is at issue in public debate. Therefore, it is clearly problematic if counterspeech reinforces the conversational salience of hate speech: the mere salience of hateful representations constitutes an injustice.

This, however, does not seem to be the case with ordinary political misinformation. Unlike with vilifying stereotypes, the fact that a falsehood about policy matters is at issue does not inherently constitute an injustice. Indeed, that there is a debate about whether MMR vaccines cause autism does not in and of itself call anyone's standing into question. So, if the salience of ordinary political misinformation is not harmful or unjust in the same way that the salience of hate speech is, what nevertheless makes this salience morally problematic?

In answering this question, philosophers can benefit from contemporary advances in cognitive science. Cognitive scientists have shown that there is an intimate connection between the cognitive fluency of an association—roughly, how familiar one is with that association—and one's disposition to believe it. As Lewandowsky et al. summarize, "in general, fluently processed information . . . is more likely to be accepted as true."[28]

This finding helps appreciate what is problematic about the conversational salience of political misinformation. When an association is made conversationally salient, that association is thereby activated for participants: because the association is newly relevant to the conversation, their attention is drawn to it, and they are prone to represent it. Thus, the salience of an association is tightly linked to participants' cognitive familiarity, or fluency, with it. If, as cognitive scientists argue, the fluency of an association in turn disposes people to believe it, then the problem becomes manifest. The conversational salience of ordinary political misinformation is problematic because it increases the likelihood that participants will accept the misinformation. And this, in turn, is plausibly causally related to harm and injustice. Insofar as political misinformation conceals the vices of dangerous policies and candidates, and obscures the virtues of good policies and candidates, the salience-induced acceptance of misinformation may help produce injustice.

The upshot for counterspeech directed at ordinary political misinformation

28 Lewandowsky et al., "Misinformation and Its Correction," 212. See also Alter and Oppenheimer, "Uniting the Tribes of Fluency to Form a Metacognitive Nation," 228; and Berinsky, "Rumors and Health Care Reform," 245–47.

is troubling: insofar as countering misinformation with more speech magnifies its salience, doing so risks bolstering people's disposition to accept the targeted misinformation, together with the attending dangerous consequences. This worry is not merely theoretical. Social psychologists have widely reported a "continued influence effect," whereby attempting verbally to correct falsehoods is ineffective or worse.[29] In one study, CDC flyers distinguishing myths from facts about vaccines are shown to backfire: people who have read the flyer end up being more likely to misidentify myths as facts—and to oppose vaccination—than people who have not.[30] Similarly, Adam Berinsky finds that, although rehearsing and correcting the Obamacare death-panel rumors initially makes people somewhat more likely to reject those rumors, this effect largely disappears after a few weeks.[31]

Thus, the problem of stickiness applies not merely to hateful forms of ignorant speech, but also to political misinformation. As we have seen, there are two steps to establishing this generalization: first, as with hateful forms of ignorant speech, it is easier to increase the conversational salience of political misinformation than it is subsequently to reverse this salience; second, in light of the foregoing theoretical and empirical considerations, the conversational salience of political misinformation is importantly connected to its damaging conversational effects. Accordingly, here also, counterspeech risks being as futile as trying to "unring a bell": instead of undoing the damaging conversational effects of political misinformation, it may amplify them.

This result constitutes a serious problem for deliberative theories of democracy. This family of theories, recall, insists that exchanging information and arguments helps to eliminate false beliefs and fallacious arguments.[32] As Hélène Landemore acknowledges, this position presupposes that people will recognize correct views and sound arguments when presented with them.[33] If, however, having deliberators challenge falsehoods amplifies their salience—which increases listeners' disposition to believe them—then this presupposition fails. Not only are people not necessarily swayed by sound counterarguments, but such counterarguments may instead entrench their false beliefs.

We have arrived at a deep and general problem for counterspeech. Because of the asymmetric pliability of conversational norms—which itself stems from a broader salience-related asymmetry—tackling hateful *and* non-hateful igno-

---

29 For a review, see Lewandowsky et al., "Misinformation and Its Correction," 113–21.

30 Lewandowsky et al., "Misinformation and Its Correction," 115.

31 Berinsky, "Rumors and Health Care Reform," 254–60.

32 See note 5 above.

33 Landemore, "Yes, We Can (Make It Up on Volume)," 220.

rant speech risks being ineffective or worse. We have already seen one possible response to this problem: Simpson suggests that, in the face of the problem of stickiness, we may have to change the subject, or wait for salient falsehoods to become unsalient. Unfortunately, neither option is wholly satisfactory: often, salient falsehoods bear precisely on the subject we are interested in, and the issue at hand may be so pressing that there is no time to wait. In what follows, I will offer an alternative response by demonstrating that the objection at hand relies on an insufficiently sophisticated account of counterspeech.

<center>2. POSITIVE COUNTERSPEECH</center>

2.1. Positive and Negative Counterspeech

In the first place, the objection under consideration overlooks the distinction between negative and positive counterspeech. One way to counter ignorant speech is to explicitly negate its ignorant content. "It is false," a politician might announce, "that Obamacare will implement death panels." Similarly, civil rights activists might respond to hate speech by insisting "*X*s are not lazy parasites!" Call this *negative counterspeech.* Because it involves repeating the ignorant proposition in the process of negating it, negative counterspeech risks strengthening its conversational salience. Accordingly, this kind of counterspeech is highly vulnerable to the stickiness of ignorant speech.

But there is another way of verbally countering ignorant speech. As we have seen, the harder-to-reverse quality of some conversational norms results partly from a salience-based asymmetry: it is easier to make associations salient than unsalient. So far, we have focused on how this general asymmetry applies to incorrect associations, hateful (section 1.1) and non-hateful (section 1.2). However, the asymmetry applies to correct associations as well: *ceteris paribus*, it is also easier to make correct associations salient than unsalient.

As a result, the asymmetric pliability of conversational norms does not suggest that all kinds of counterspeech are bound to be ineffective. Rather, it recommends a distinctive kind of counterspeech. To avoid reinforcing the salience of false associations, we should steer clear of counterspeech whose form is primarily negative. Instead, we should privilege *positive* forms of counterspeech, which aim to put forward and render salient correct associations of ideas. On this approach, countering ignorance is less about directly contesting a distorted vision of the world, and more about affirming a correct vision of the world that is inconsistent with the falsehoods at hand.[34] Because positive speech entails or

---

34 While philosophical and legal debates typically ignore this distinction, it resonates with

implicates, but does not explicitly assert, that the ignorant speech in question is false, it avoids increasing the ignorant proposition's salience—and, by extension, it avoids strengthening the conversational effects of ignorant speech. Moreover, because positive counterspeech makes correct associations salient instead, it harnesses the asymmetric pliability of conversational norms in the interests of truth: once salient, correct representations are comparatively difficult to render unsalient.[35]

---

some discussions in social psychology. Since the pioneering work of Tversky and Kahneman ("The Framing of Decisions and the Psychology of Choice"), social psychologists have investigated how framing messages in different ways—including negative versus positive ways—can alter their impact on listeners. Although Tversky and Kahneman themselves do not explore how framing might affect attempts at correcting previous misinformation, this more specific application has been taken up in the subsequent literature. Lewandowsky et al., for instance, recommend that corrections avoid repeating myths, and suggest that corrections "should be more successful when they can be encoded as an affirmation of an alternative attribute" ("Misinformation and Its Correction," 115–17). Though these comments are brief, and focus exclusively on ordinary misinformation, they seem congruent with my recommendation, inasmuch as they suggest countering falsehoods in an affirmative mode, rather than by negating falsehoods.

35 This last claim might seem too quick. As Leslie has argued, human beings are disposed to generalize strikingly negative information more quickly—and on the basis of less evidence—than neutral or positive information ("The Original Sin of Cognition," 395–99). Accordingly, one might think that negative associations are "stronger" than positive associations, so that positive associations are easier to make unsalient than negative associations. There are two things to say in response to this concern. First, Leslie's argument does not speak directly to the *salience* of positive and negative associations. Put differently, her evidence may indeed show that in *one* respect—the tendency to be generalized—negative associations are stronger than positive associations. But, from this, it need not follow that negative associations are stronger in the respect under consideration—namely, the difficulty of being made unsalient. After all, it could conceivably be the case that (1) *while* negative information is salient, we are more likely to generalize on its basis than on the basis of positive information, but that (2) negative information is no more difficult to render unsalient than positive information. Still, Leslie's argument relating to generalization might be thought to at least suggest the possibility that, by analogy, negative associations might be more difficult to render unsalient than positive associations. However—and this is the second point—even if this hypothesis is true, it remains consistent with my main point here: that, in virtue of how salience works, positive associations are more difficult to render unsalient than to initially render salient. Indeed, we can think that such a salience-based asymmetry applies to positive associations (which creates a hurdle that opposing speech must overcome) while allowing that this asymmetry is even stronger in the case of negative associations. This is congruent with the kind of evidence Leslie finds. She finds that we typically generalize *more quickly* on the basis of negative information than on the basis of positive information, *not* that we do not generalize on the basis of positive information. In fact, she even acknowledges that, although this is less common, the tendency to rapid generalization "arguably applies to strikingly positive information" ("The Original Sin of Cog-

Notice that resorting to positive counterspeech does not amount to simply changing the subject. The truths that positive counterspeech affirms bear on the same issue as negative counterspeech. This is guaranteed by the fact that the content of positive counterspeech is supposed to entail (or at least implicate) the falsehood of the targeted ignorant speech. So, positive counterspeech expresses a similar proposition as negative counterspeech, but under a different mode of presentation. Consider: "The star of *King's Row* was married twice" and "The fortieth US president was married twice" make the same claim about the world (namely, that Ronald Reagan was married twice) while highlighting different facets of that claim (one highlights Reagan's film career, the other his political role). Analogously, positive and negative counterspeech tell us something similar about the state of the world, while emphasizing different features of that state: while negative counterspeech reactivates incorrect associations in the process of negating them, positive counterspeech activates correct associations.

Consider an example. In his "I Have a Dream" speech, Martin Luther King Jr. takes a stand against a deeply racist public culture, which is rife with statements professing that black Americans are morally, culturally, and intellectually inferior to whites. Famously, King does so by vividly depicting a vision of racial equality. In this vision, the American nation "will rise up and live out the true meaning of its creed: 'We hold these truths to be self-evident, that all men are created equal,'" and, accordingly, "the sons of former slaves and the sons of former slave owners will be able to sit down together at the table of brotherhood."[36] King's utterance is very different from saying: "That's false! Black Americans are *not* immoral, uncivilized, or stupid!" While his vision of racial equality entails—or at the very least, strongly implicates—the falsehood of these racist stereotypes, it is first and foremost a positive representation of racial equality. This difference matters. Because it does not repeat the racist stereotypes, King's positive speech does not enhance their salience in the public sphere. Instead, it draws his audience's attention away from these stereotypes, and toward a different—and incompatible—set of associations: namely, between the American dream and the values of equality and interracial unity.[37]

---

nition," 397n10). So the analogy at hand does not endanger my thesis that the salience-based asymmetry might work for correct and positive associations: it suggests, rather, that this effect might work even more strongly when the associations in question are negative.

36 King, "I Have a Dream."

37 As this example indicates, in advocating positive counterspeech—which, when leveled at hateful speech, may well involve putting forward an egalitarian vision—I am in agreement with McGowan's suggestion that speech can be used to enact egalitarian norms (*Just Words*, 185–89). There are, however, at least two reasons why McGowan's discussion of positive speech needs developing. First, because McGowan only discusses the positive potential

Positive counterspeech can also be used to oppose ordinary political misin-formation. In response to a misleading statement X (say, that the UK pays the EU £350 million a week, which could benefit the National Health Service instead), one strategy is to say: "X is false. We do not pay that amount." But another strat-egy would be to conduct a public education campaign that avoids mentioning X and instead draws the public's attention to facts that entail X's falsehood. Such a campaign might publicize facts about what the UK *receives* from the EU (includ-ing, if applicable, how free movement in Europe helps staff the National Health Service).[38] And it might also reveal facts about what the UK actually pays the EU (expressed, perhaps, as a proportion of overall spending, so as to highlight its comparatively small size). Such a public response would entail that the UK does not pay the EU as much as X claimed, and that it is misleading to posit a contrast between funding the EU and funding the National Health Service. In doing so, however, it would avoid repeating the £350 million figure, and would therefore avoid reinforcing its salience.

In fact, there is empirical evidence that such positive responses to political misinformation are more likely to succeed. Notice, first, that the pessimistic evidence mentioned in section 1.2 involved paradigmatic instances of *negative* counterspeech. The CDC "myth versus fact" flyers about vaccines worked by reiterating myths and labeling them "false." Moreover, the result that correct-ing Obamacare death-panel rumors was ineffective obtained primarily in cases where the rumors were explicitly repeated beforehand. By contrast, when sub-jects did *not* repeat the rumor beforehand, corrections actually tended to reduce support for the rumor.[39] Likewise, Lewandowsky et al. report that people are more likely to reject false reports about how a fire started if, instead of simply being told that the initial report was false, they are given an alternative narrative of how the fire began.[40] On this basis, they hypothesize that, when correcting misinformation, it is better to "emphasize the facts you wish to communicate rather than the myth."[41]

of speech by way of concluding *Just Words*, her exploration of this idea is quite brief and suggestive. Second, and more substantively, McGowan does not connect her concluding discussion of the positive potential of speech with her earlier critique of counterspeech ear-lier in *Just Words*. Specifically, in chapter 5, McGowan criticizes the use of counterspeech to respond to oppressive speech by appealing to the asymmetric pliability of conversational norms (*Just Words*, 11–20). What I am suggesting is that, provided counterspeech takes a "positive" form (as I define it here), it can go some way toward alleviating this concern.

38 Blitz, "Brexit and the 'NHS Dividend.'"
39 Berinsky, "Rumors and Health Care Reform," 258.
40 Lewandowsky et al., "Misinformation and Its Correction," 117.
41 Lewandowsky et al., "Misinformation and Its Correction," 123.

In light of the positive conception of counterspeech, counterspeech seems more promising than trying to unring a bell. The conversational norms enacted by ignorant speech are difficult to reverse primarily when counterspeech takes a negative form, which repeats the ignorant contents to debunk them. But positive counterspeech is less vulnerable to this problem. It draws attention away from false associations, and instead renders salient correct associations that entail or implicate the falsehood of ignorant utterances. In doing so, positive counterspeech repurposes the asymmetric pliability of conversational norms, and makes it work for truth rather than falsehood.

2.2. Concerns with Positive Counterspeech

Even though distinguishing between positive and negative counterspeech helps alleviate the problem of stickiness, difficulties remain. The first problem is that, in some cases, positive counterspeech seems unavailable. Positive counterspeech affirms a view that entails or implicates the falsehood of the targeted ignorant speech *without* repeating or invoking its contents. So, positive counterspeech is available as an alternative to negative counterspeech only insofar as there exists a way of presenting the negation of the ignorant contents that does not explicitly invoke those contents. For instance, in King's "I Have a Dream" speech, positive counterspeech is possible because there is a way of presenting racial equality—namely, King's vision of interracial unity and brotherhood—that is not simply the explicit negation of racist stereotypes.

For some instances of ignorant speech, however, there is no way of saying something that entails or suggests the falsehood of the ignorant speech besides invoking and rejecting its ignorant contents. Conspiracy theories seem especially liable to generate this difficulty. By definition, conspiracy theories propound extremely improbable causal stories, which lack credible evidence. *Because* these stories are so improbable, it is difficult verbally to oppose conspiracy theories in a way that does not inadvertently invoke—or "ring the bell" of—the conspiracy theory.

Consider birtherism, the theory that Barack Obama was born in Kenya and is therefore ineligible for the US presidency.[42] Negative counterspeech—saying that Obama was *not* born in Kenya—clearly reactivates the association between Obama and Kenya. But it is not clear what the positive alternative would be. Vigorously affirming that Obama was born in Hawaii would be a highly unusual thing to do, were it not for the concern that Obama was born outside the US. Put differently, because the conversational salience of the birtherist myth is the main reason why Obama's birthplace is politically worth talking about, it is difficult

42   Smith and Tau, "Birtherism."

to say anything on this topic without inadvertently invoking or reactivating the birtherist myth. Given this background fact about conversational salience, insisting that Obama was born in Hawaii risks, like negative counterspeech, calling attention to the concern that he was born in Kenya.

The problem so far has been that genuine positive counterspeech—which opposes falsehoods without reinvoking them—may sometimes be unavailable. But there is a more fundamental problem. Even when we *can* use positive counterspeech to reverse the conversational salience of ignorant associations, counterspeech cannot reverse the harms that ignorant speech brought about while it was salient. McGowan articulates this concern when discussing sexist speech:

> the mere fact that an act of oppression can easily be reversed does not entirely disqualify it as oppressive.... While shorter-lived [oppressive norms] may be *less* oppressive ... they are still oppressive.[43]

The point is straightforward. Even if counterspeech can deactivate vilifying stereotypes, it presumably cannot change the fact *that they had been activated up to that point.* And, consequently, it cannot change the fact that, while they were activated, their salience gave rise to harms.

This concern is not specific to oppressive or hateful types of ignorant speech. Indeed, Laura Caponetto's taxonomy of the ways in which speech can "undo" the effects of prior utterances lends more general support to the worry. According to Caponetto, speech can *recognize* that a past utterance failed to enact the norms we thought it did. For instance, when a newspaper reveals that Father Tom was not actually a priest, and the Roman Rota (or a similarly qualified institution) responds by declaring that the marriages he officiated are null and void, this declaration constitutes a recognition that the validity of these marriages was only purported all along.[44] Alternatively, speech can *amend* or *retract* conversational norms that were enacted by prior utterances, so that these norms are discontinued.[45] But what Caponetto explicitly refrains from saying is that speech can retroactively make it the case that certain conversational norms were never enacted to begin with.[46]

If Caponetto is right to refrain from attributing this retroactive ability to speech, then this highlights a second sense in which counterspeech might seem as futile as trying to unring a bell. Even if you can eventually stop a bell from continuing to ring, you cannot undo the fact that it rang in the first place. Analogous-

43 McGowan, "Oppressive Speech," 404–5.
44 Caponetto, "Undoing Things with Words," 2, 6–8.
45 Caponetto, "Undoing Things with Words," 9–14.
46 Caponetto, "Undoing Things with Words," 10.

ly, even if positive counterspeech can stop the harmful conversational effects of ignorant speech from *persisting*, it can never change the fact that these harmful effects *occurred up to that point*. So, even in the best cases, positive counterspeech seems a limited remedy for ignorant speech: it cannot fully eliminate its harms.

Still, even if it is true that counterspeech is limited in this way, one might respond that it is unfair to hold this against proponents of counterspeech. After all, asking of counterspeech that it somehow fully eliminate the harms of ignorant speech, or that it somehow make it the case that the ignorant speech never occurred, may seem impossibly demanding. And if this demand is unrealistic or unachievable, then it seems misguided to make it.

What this response reveals is that the objection at hand makes sense only in light of a particular understanding of what the alternative to counterspeech should be. Critics of counterspeech are typically not suggesting that, in place of counterspeech, we do nothing. Rather, they commonly recommend implementing legal restrictions.[47] Crucially, the legal regulation of hate speech, or of fake news, is often justified as a deterrent: by imposing penalties, criminal or civil laws get people to refrain from uttering hateful or deceptive contents. When legal remedies succeed as deterrents, they *fully* eliminate the harms that suppressed utterances would otherwise have occasioned. Indeed, insofar as legal remedies prevent ignorant utterances altogether, they prevent any harms these might have generated. Thus, to the extent that legal restrictions succeed as deterrents—an assumption I will revisit in section 3.2—the objection at hand places counterspeech at a comparative disadvantage relative to legal remedies: whereas counterspeech comes in too late to eliminate all of the harms associated with ignorant speech, legal remedies can in principle fully remove those harms. This, in turn, would constitute a reason for preferring legal remedies to counterspeech.

The following section will argue that the two concerns just outlined rely on an overly crude conception of the temporality of counterspeech. My focus will primarily be on the second concern—that counterspeech, even if it takes a positive form, cannot reverse the harmful effects that ignorant speech generated prior to being countered. Nevertheless, in closing, I will also indicate how the view of counterspeech's temporality that I develop can defuse the first concern.

3. THE TEMPORALITY OF COUNTERSPEECH

There are at least two potential ways of rethinking the temporality of counterspeech to address the above objections: the first approach emphasizes the retroactive character of counterspeech; the second underscores its diachronic nature.

47   See, e.g., McGowan, *Just Words*, ch. 7.

While the first strategy is fascinating, I will suggest that, even if we grant its theoretical viability, it ultimately cannot overcome the objections at hand (3.1). The second approach, by contrast, meets with greater success (3.2).

3.1. Retroactive Counterspeech

In the first place, one might object that the objections raised in section 2.2 do not take the *retroactive* potential of counterspeech sufficiently seriously.

It seems perfectly natural to think that the past cannot be changed. By extension, the idea that counterspeech could alter the past may, as Caponetto labels it, seem "magica[l]."[48] Accordingly, one might well conclude that speech can at best stop the conversational effects of prior utterances from persisting, but it cannot make it the case that those effects never happened to begin with—and hence, it cannot fully eliminate the harms of ignorant speech.

However, Rae Langton has recently argued that this skepticism is misplaced. Beyond putting a stop to the preexisting conversational effects of past utterances, Langton suggests, counterspeech can retroactively block the occurrence of these conversational effects altogether. More precisely, retroactive counterspeech "changes a past utterance from the unactualized way it would have been, to the way it actually is."[49] In this way, counterspeech "offers a ticket to a modest time machine."[50] Applied to ignorant speech, the idea is that counterspeech can make it the case that prior utterances failed to ever generate certain damaging conversational effects. If so, then one might think that—*contra* the second objection raised in section 2.2—counterspeech can entirely eliminate the harms of ignorant speech.

The immediate question, however, is why we should accept Langton's counterintuitive claim that counterspeech can retroactively alter the nature of prior utterances. Langton offers two reasons. The first is that, in many other domains, we already recognize that an act's character can retroactively be altered by future happenings. "A stabbing may be a killing," Langton observes, "partly in virtue of what happens later."[51] In other words, the fact that a stabbing is also a killing may depend on the fact that the victim *later* succumbed to her wounds. If doctors had managed to save the victim's life, the past stabbing would arguably have constituted a nonlethal assault instead. Thus, if Langton is correct, we cannot dismiss the idea of retroactive counterspeech simply on the grounds that retroactivity seems "magical"—we already recognize retroactive phenomena in everyday life.

48   Caponetto, "Undoing Things with Words," 10.
49   Langton, "Blocking as Counterspeech," 156.
50   Langton, "Blocking as Counterspeech," 146.
51   Langton, "Blocking as Counterspeech," 157.

Nevertheless, even if some facts can retroactively be made true, one might doubt whether this applies to facts about the effects of past utterances. To establish this stronger claim, Langton offers a second argument. Drawing on Austinian speech-act theory, Langton distinguishes two kinds of things utterances can do. Utterances can *cause* things to happen. For instance, delivering a brilliant argument can cause one's audience to affirm its conclusion. This is roughly what Austin refers to as an utterance's perlocutionary dimension. But utterances can also *constitute* certain acts. When a priest says, "I hereby pronounce you husband and wife," she is not merely causing the couple to be married. Instead, her utterance is part and parcel of, or constitutive of, the act of *marrying* them. The speech act(s) constituted in saying something belong(s) to the utterance's illocutionary dimension.[52]

For an utterance successfully to constitute a particular speech act—marrying a couple, declaring war, etc.—certain background conditions ("felicity conditions") must be satisfied.[53] What kinds of felicity conditions are needed will vary from speech act to speech act. According to McGowan, however, most speech acts depend on the speaker having a particular status or social power.[54] Some speech acts (such as marrying, condemning, or ordering) require that the speaker have a comparatively high or special status, which is commonly referred to as "authority."[55] But, McGowan insists, even speech acts that do not require authority nevertheless require that speakers have a minimal and widespread status (which she calls "standing") in virtue of which they are capable of taking part in conversations.[56] Should key felicity conditions not be satisfied, the utterance will misfire, or fail to constitute the relevant speech act.[57] If the officiant for a religious wedding has not been ordained, for instance, then she lacks the relevant authority and her utterance, "I hereby pronounce you married," will fail to marry the couple. Similarly, if a lowly private (rather than a general) shouts "Attack!" to an assembled army, he will (in normal conditions) fail to order that army into battle.

52 Langton, "Speech Acts and Unspeakable Acts," 300–1. See also Austin, *How to Do Things with Words*, 98–108.

53 Austin, *How to Do Things with Words*, 12–49.

54 McGowan, *Just Words*, 15, 63–68.

55 McGowan, *Just Words*, 63–66. See also Langton, "Speech Acts and Unspeakable Acts," 304–5; and Austin, *How to Do Things with Words*, 23–24.

56 McGowan, *Just Words*, 66–68.

57 Note, however, that not all felicity conditions are like this: some are such that, when they fail to be satisfied, the intended speech act is nonetheless constituted, though it is nonideal. See Austin, *How to Do Things with Words*, 39–52.

Crucially, Langton claims, felicity conditions for speech acts can sometimes be satisfied in the future, after the utterance. By way of illustration, she considers the case of marriage: "for 'I do' to count as a marriage," she suggests, "may require felicity conditions in the future, e.g., the consummation of the marriage."[58] This, on Langton's analysis, is what makes retroactively undoing a speech act possible: if a future felicity condition fails to be satisfied, then it follows that the past utterance misfired. Applied to the marriage case, the idea is that, if some "sad events" keep the couple from consummating the marriage, this makes it the case that the earlier "I do" failed to constitute a speech act of marrying.[59]

The final step in Langton's argument is to emphasize that counterspeech can be a way of challenging the future felicity conditions of speech acts. To see this, consider again that many speech acts require that the speaker have authority. Thus, in attempting to perform such speech acts, speakers tend to presuppose that they have the relevant kind of authority. And, except in cases where their authority is already firmly entrenched, they depend on their audience to accept that presupposition. Consequently, a significant way in which counterspeech can block a speech act is by challenging such presuppositions of authority.[60] A colleague tells you, "A double espresso; make it quick!" and you furiously respond, "Who do you think you are? I don't take orders from you!" The colleague's utterance presupposes that they have the authority needed to give you orders. In denying that presupposition, you prevent their speech from constituting an order. At best, it becomes a (very unsuccessful) request.[61] The thought, then, is that by undermining a future felicity condition of the utterance, your response retroactively changes the nature of that utterance.

Returning to ignorant speech, if this theoretical analysis is correct, it suggests that counterspeech can retroactively make it the case that past ignorant utterances failed to enact problematic conversational norms. Imagine that A tells B: "Your kind are nothing but worthless parasites!" In uttering this falsehood, as we saw in section 1.2, A might be attempting to constitute the following harmful speech act: to refute the assurance that B is a member of society in good standing. Now, suppose that A is part of a minority extremist political group—a group whose authority to determine others' social status is not already firmly established.[62] And suppose, moreover, that C—a spokesperson for the majority—in-

58 Langton, "Blocking as Counterspeech," 157.

59 Langton, "Blocking as Counterspeech."

60 Langton, "Blocking as Counterspeech," 150.

61 Langton, "Blocking as Counterspeech," 156.

62 This stipulation is important. As I specified in the previous paragraph, the ability for counterspeech to block a harmful speech act by challenging a speaker's authority is restricted to

terjects: "You don't speak for us, *A*. And we will *never* condone that kind of view."
In challenging *A*'s presupposed authority, *C* retroactively makes *A*'s utterance
misfire. *A*'s utterance does not refute the assurance of *B*'s standing, because *A*
lacks the requisite authority. In sum, if Langton's theoretical account is correct,
then—*contra* the second objection outlined in section 2.2—counterspeech can
do more than simply stop ignorant utterances from continuing to harm: by chal-
lenging the future felicity conditions of prior ignorant utterances, it can retroac-
tively prevent past utterances from ever constituting harmful speech acts.

However, there are problems with the idea of retroactive counterspeech. In
the first place, Langton's arguments for its possibility might seem contentious.
For one thing, one might deny that the concrete cases Langton points to really
do intuitively involve retroactive phenomena. In the stabbing case, for instance,
one might give *epistemic* significance, rather than constitutive significance, to the
victim's death. On this view, the victim's death merely shows us that the pre-
vious stabbing was also a killing. Alternatively, one might think that the death
does make the stabbing a killing, while denying that this constitutive function is
retroactive: that is, after the death has occurred, the stabbing counts as a killing
from the time of death onward, but not before. As for the marriage case, it is un-
clear that consummation is actually a necessary felicity condition for the success
of speech acts of marrying.[63] If so, then the failure to consummate the marriage
may not actually retroactively affect its validity.

Moreover, the argument for retroactive counterspeech may come at a the-
oretical cost. The stabbing case suggests that the future felicity conditions for
an utterance to constitute a specific illocutionary act may include some of its
perlocutionary effects. As a result, one might worry that, in explaining how ret-
roactive counterspeech is possible, Langton's account ends up blurring the dis-
tinction between illocutionary and perlocutionary dimensions of utterances.[64]

Nevertheless, I want to grant, for the sake of argument, that these concerns
about the possibility of retroactive counterspeech can be overcome. My main
point is that, *even if* we assume that retroactive counterspeech is possible (and

---

cases where the speaker somehow depends on the audience either accepting or at least not
challenging their authority. In cases where the speaker's authority is already firmly estab-
lished (for example, if the chancellor of the Third Reich said the same thing as *A*) counter-
speech may not be able to retroactively block the harmful speech act. In section 4, I briefly
reiterate this limitation and its implications for whether we ought to adopt legal restrictions.

63 Indeed, in Catholic doctrine, consummation may be required to make a marriage *indissol-
uble.* But this is consistent with a marriage being *valid*. Here, and in the previous case, I am
grateful to a reviewer for drawing my attention to these complications.

64 I owe this insight to a reviewer. For further criticisms of Langton's account of retroactive
blocking, see McGowan, *Just Words*, 48.

that it is realized in the cases Langton points to), it cannot entirely defuse the present objection.

Langton's analysis of how speech can retroactively block earlier utterances focuses exclusively on utterances' illocutionary dimension—i.e., on what acts might be *constituted* by those utterances. But, as explained above, utterances also have a perlocutionary dimension: besides constituting acts, utterances also have *causal* effects, such as getting listeners to believe something. The retroactive picture of counterspeech does not establish, and is not intended to establish, that these causal effects can be undone.

First, even if counterspeech prevents an utterance from constituting a particular speech act, that utterance might still causally influence listeners *as if* the speech act had been successful. This is because listeners may not know that the felicity conditions are not satisfied. Suppose Langton is right that, if *A* and *B* fail to consummate their marriage, their "I do's" fail to constitute the speech act of *marrying*. Even so, if people never find out, the utterance "I do" will retain its habitual causal consequences. Friends will congratulate the couple on subsequent anniversaries, the Internal Revenue Service will tax them as a married couple, immigration services will grant *A* a spousal visa when *B* takes a job in a different country, and so on.

Second, there is no way of undoing these causal effects after the fact. To reiterate, the reason why acts constituted by utterances can retroactively be undone is that their performance depends on felicity conditions, and felicity conditions can sometimes be satisfied after the utterance. By contrast, the causal effects of utterances do not depend on felicity conditions in this way. Therefore, once the causal effects have occurred, nothing can change this fact. As Caponetto nicely illustrates: "one can take back a marriage proposal, but one cannot take back the hearer's excitement at hearing the words 'Will you marry me?'"[65]

This spells trouble for counterspeech directed at ignorant speech, because many of the harms attributed to ignorant speech are causal. For instance, one of the ways in which hate speech harms its targets is by causing them to experience psychological distress, particularly when the hate speaker is perceived to have significant social authority. Now, for Alexander Brown, it is "hard to see how any counterspeech could ameliorate some of the[se] psychological harms."[66] Above, I suggested that counterspeech can undermine hate speakers' authority. If the retroactive blocking account is correct, doing so can retroactively prevent hate speech from constituting certain harmful speech acts (since it undermines a felicity condition for such speech acts). Furthermore, it can help targets feel less

---

65   Caponetto, "Undoing Things with Words," 4.

66   Brown, "Hate Speech Law," 260.

distressed by the hate speech than they previously did (once they realize that it only represents the views of a small minority). Even so, Brown's point is that counterspeech cannot undo the distress experienced by targets of hate speech *in the intervening period*, before counterspeech told them that the hate speakers lacked social support.

A similar observation applies to ordinary political misinformation. Counterspeech might, by undermining the authority (or even standing) of a source of misinformation, get the audience at large subsequently to reject that source's claims. But it cannot change the causal harms generated by political misinformation *before* it was verbally countered. Imagine that, following an election, an anonymous blogger publicly and falsely claims that the election was rigged. While the vast majority of people withhold judgment as to whether the blog post is reliable, and some express uncertainty about its reliability, a minority flies into a rage upon hearing this, and responds by destroying private property. Even if counterspeech later repudiates the blogger's credibility in a way that all find convincing, it clearly cannot undo the destruction caused in the intervening period.[67]

The upshot is that, at best, the appeal to retroactive counterspeech can only help address the objection that counterspeech cannot undo the harms generated by ignorant speech before it was countered to a limited extent. If Langton is right, counterspeech can retroactively make it the case that a prior ignorant utterance did not constitute a harmful speech act. But it cannot retroactively undo the past causal harms of ignorant speech.

This limitation matters, once more, in light of the comparison between counterspeech and legal remedies. By contrast with counterspeech, insofar as legal remedies successfully prevent ignorant speech from being uttered, they can entirely eliminate its harms, both constitutive *and* causal. So, even if we were

---

67 A possible complication with this example is that, on Maitra's account of "licensed authority," it is possible for a speaker to acquire authority even if the majority does not accept their claims ("Subordinating Speech," 107). Consequently, one might worry that, in the example, the blogger succeeds in gaining authority prior to the counterspeech, so that the counterspeech fails to retroactively block their authority. A first response is that it is unclear that this example really is a case of licensing in Maitra's sense. In Maitra's discussion, the phenomenon of licensing standardly involves the audience remaining silent, or not making their reservations public ("Subordinating Speech," 107, 116). By contrast, in the above case, even before the challenge to the blogger's presupposed authority occurs, some members of the majority express their uncertainty about the blogger's credibility. Second, and more fundamentally, this concern does not undermine my main point here: that point, once more, is that *even if we assume*, for the sake of argument, that the counterspeech in this example has successfully retroactively blocked the blogger's authority, it still cannot possibly reverse the causal harms produced by the misinformation in the meantime.

to grant Langton's theoretical case for thinking that counterspeech can operate retroactively, counterspeech would still in principle remain at a comparative disadvantage with legal remedies.

3.2. Diachronic Counterspeech

There is a second way in which the objection under consideration—that counterspeech cannot undo the harms generated by ignorant speech before it was countered—depends on an inadequate account of the temporality of counterspeech. Simply put, counterspeech should not merely be interpreted as speech that responds to ignorant utterances after the fact. Rather, counterspeech is better understood as diachronic—as a continuous process, extended over time, that precedes as well as follows ignorant speech. Thus understood, counterspeech might involve educating listeners to inoculate them against future ignorant speech, or preemptively warning listeners about unreliable sources.

A diachronic understanding is implicit in some defenses of counterspeech. When discussing state counterspeech, for instance, Brettschneider recommends having schools teach children about important political matters, such as the Holocaust.[68] This counterspeech is diachronic insofar as it may affect its audience *prior* to their exposure to relevant ignorant utterances: schoolchildren may never have encountered Holocaust denial. Similarly, in her analysis of "toxic speech," Tirrell recognizes that counterspeech can work both as a *post hoc* "antidote" and as a preventive "inoculation."[69]

Nevertheless, the vast majority of discussions of counterspeech articulated by philosophers of language—including the objection currently under examination—cast counterspeech as a *post hoc* response to ignorant utterances. Caponetto's examples of undoing things with words, for example, are invariably verbal attempts at disabling an earlier utterance.[70] Likewise, Maitra and McGowan criticize counterspeech partly by appealing to evidence that targets of face-to-face hate speech rarely respond to that speech then and there. Here too, counterspeech is understood as a *post hoc* response.[71] In recent work, McGowan does briefly acknowledge that counterspeech could take a broader form, such as "a general education campaign." But she insists that her focus, in critiquing counterspeech, is on counterspeech understood as "direct and fairly immediate

---

68 Brettschneider, *When the State Speaks, What Should It Say?* 96–104.

69 Tirrell, "Toxic Speech," 136–39. See also Richards and Calvert, "Counterspeech 2000," 569–74; and Gelber, "Reconceptualizing Counterspeech in Hate Speech Policy (with a Focus on Australia)," 214–15.

70 Caponetto, "Undoing Things with Words." See also Langton, "Blocking as Counterspeech."

71 Maitra and McGowan, "Introduction and Overview," 10.

responses to allegedly harmful utterances."[72] However, limiting her scope in this way substantially limits the significance of her criticisms. Defenders of counterspeech can reply that they are recommending not simply *post hoc* responses, but counterspeech that may both preempt and follow ignorant speech. Indeed, such diachronic counterspeech, I will now suggest, can go some way toward defusing the objection at hand.

When counterspeech occurs *after* ignorant speech, the problem is that ignorant speech can generate harms prior to being countered, and—at least in the case of causal harms—these harms cannot be undone. How does counterspeech that precedes ignorant speech help? The idea is that preemptive counterspeech can condition the conversational setting to make it inhospitable to ignorant speech. Consequently, when ignorant speech is later uttered, it fails to have its damaging conversational effects. This helps address the problem: if preemptive counterspeech disables ignorant speech before it is uttered, then there is no interval during which ignorant speech can, unopposed, generate irreversible harms.

The basic idea is straightforward. What needs elaboration is how precisely preemptive counterspeech might condition the conversational setting so as to disable subsequent ignorant utterances. There are at least two promising strategies. The first is to expound and widely diffuse important politically relevant facts. Advancing such truths alters the conversation in several ways. First, it renders those truths conversationally salient, and thereby familiarizes listeners with them. Relatedly, it facilitates the introduction of these truths into the common ground of shared beliefs. Indeed, preemptively expounding facts can induce listeners to believe those facts for various reasons: because the speaker is perceived as authoritative; because the speaker offers compelling arguments for them; or simply because, as we have seen, the more familiar propositions are to listeners, the more listeners are disposed to believe them.

This conditioning makes it more difficult for falsehoods to subsequently gain a foothold in the conversation. Insofar as listeners are knowledgeable about a subject matter, they are less vulnerable to being swayed by ignorant speech. Holocaust denial or misleading claims about EU funding are less readily accepted if they contradict propositions on these topics that are already part of the audience's common ground. Moreover, as discussed in section 1.1, it is harder to make ideas unsalient than salient. Accordingly, if true propositions have already been made salient by preemptive counterspeech—which, in turn, disposes the audience to believe them—ignorant speech may struggle to render them unsalient.

72 McGowan, "Responding to Harmful Speech," 183.

A second preemptive strategy is to warn the audience about unreliable sources. By undermining the authority of untrustworthy sources, such preemptive counterspeech prevents those sources' utterances from affecting conversations in damaging ways. This idea is akin to the silencing effect that philosophers of language often ascribe to hate speech and pornography. Hate speech and pornography, it is often said, prevent minorities and women from contributing effectively to conversations.[73] One of the ways in which they do so is by stripping these groups of their authority.[74] When hate speech persistently represents racial minorities as inferior, for instance, they may lose the social standing needed for their speech to be taken seriously.

I am suggesting that this silencing effect can also be used for good, by challenging the authority of prospective promoters of ignorance. This usage applies most clearly to political misinformation: if hearers are credibly warned that a political news site has been systematically wrong, that source's subsequent utterances may be taken less seriously. Nevertheless, it also applies to hateful ignorant speech. Suppose a democratically elected head of state affirms: "We categorically reject racist ideologies. Racism is unwelcome here." The hate speaker who *then* pronounces deeply racist views thereby marks himself as a minority voice, who cannot speak for the majority.

Thus, counterspeech can condition the conversation to protect it from subsequent ignorant speech in at least two ways: by entrenching important facts in the conversation's common ground, and by eroding ignorance-promoting speakers' status, so that they no longer count as authorities.

Importantly, this preemptive conditioning is relevant to the constitutive *and* causal effects of ignorant speech. Like retroactive counterspeech, preemptive counterspeech can undermine the felicity conditions (such as authority) that ignorant utterances need to constitute harmful acts. The only difference with

---

73 E.g., Langton, "Speech Acts and Unspeakable Acts"; Maitra and McGowan, "Introduction and Overview"; and McGowan, "Responding to Harmful Speech." Note that I am using "silencing" in a loose sense. First, as I am using the idea, silencing includes both perlocutionary silencing (preventing speech from having its intended causal effects) and illocutionary silencing (preventing speech from constituting its intended speech acts). On this distinction, see Langton, "Speech Acts and Unspeakable Acts," 314–15. Second, my use of the term is broader than typical discussions of silencing in the following way. Typically, silencing involves the speaker being prevented from doing certain things with their words as a result of a recognition failure on the part of the hearer. For example, the speaker might have authority, but the hearer fails to recognize this. However, in the cases I am concerned with here, silencing arguably does not involve a recognition *failure*: the speaker actually lacks pre-established authority, and the hearer is right to think that they do not satisfy the authority condition. I am grateful to a reviewer for this insight.

74 McGowan, "Debate," 491–92.

retroactive counterspeech is that, here, the felicity conditions are undermined *before* the problematic utterance.

In addition, preemptive counterspeech alleviates the difficulties that retroactive counterspeech encountered with ignorant utterances' causal harms. The causal effects of speech depend in significant part on what that speech is *perceived as*.[75] The problem with retroactive counterspeech is that, in the interval before ignorant speech is countered, a mismatch might arise between what the utterance actually *is* and what it is *perceived as*. Before retroactive counterspeech established that the purveyor of the vaccines-cause-autism myth was a crank, people may have believed that he was delivering an authoritative verdict, and followed his dangerous prescriptions. Preemptive counterspeech addresses this causal problem by eliminating the interval in which ignorant speech has been uttered but not yet countered. In doing so, it reduces the likelihood of a temporary mismatch between what the ignorance is and what it is perceived as. If listeners already know that A is a crank when A speaks, they are less vulnerable to being duped into harmful actions.

This theoretical account of how counterspeech might preemptively defuse the constitutive and causal effects of ignorant speech is not merely speculative. Cook, Lewandowsky, and Ecker, for instance, find that telling subjects about the scientific consensus on anthropogenic climate change neutralizes the causal impact of subsequent climate-related misinformation.[76] This corroborates the first preemptive strategy outlined above. By giving people authoritative testimony that climate change is real, the preemptive speech introduces this proposition into the common ground of shared beliefs. This makes it harder for subsequent climate-related misinformation to gain assent.

There is also support for the second preemptive strategy: casting doubt on the authority or credibility of unreliable sources. Lewandowsky et al. report that "misinformation effects can be reduced if people are explicitly warned … that information they are about to be given may be misleading."[77] For example, when people are warned that upcoming information about climate change may come

---

75 Note that, while very many causal effects of utterances depend on how the utterance—and more specifically, its force—is perceived, some are not like this. The utterance "Move aside!" may cause me to move aside if I perceive it as an order. In addition, however, if it is extremely loud, it might also cause my ears to hurt or cause me to be startled. Arguably, neither of these latter two causal effects depends on what I perceive the utterance as. That being said, the harmful causal effects of ignorant speech I am concerned with here generally tend to stem from the perceived force of the utterance.

76 Cook, Lewandowsky, and Ecker, "Neutralizing Misinformation through Inoculation," 10.

77 Lewandowsky et al., "Misinformation and Its Correction," 116.

from nonexperts, and given some indication of how to identify nonexperts, the influence of subsequent misinformation on their beliefs disappears.[78]

So far, I have argued that adopting a diachronic view of counterspeech—where counterspeech can preemptively disarm ignorant speech as well as retrospectively oppose it—helps overcome the main objection considered in section 2.2: that counterspeech can only stop the conversational effects of ignorant speech from persisting, and hence cannot wholly eliminate its harms. But switching to a diachronic view of counterspeech *also* helps counter the other objection raised in section 2.2: that some kinds of ignorant speech—e.g., conspiracy theories—cannot be opposed in a positive way. The only way to counter them, it seemed, is to negate their content. However, doing so risks increasing their salience, and thereby exacerbating their damaging effects.

The idea of preemptive counterspeech reveals a better response. The problem with these utterances is that, once they gain traction, it is difficult to oppose them in a way that does not backfire. To circumvent this difficulty, a more promising approach would be to preemptively condition the conversational context so that they never gain traction to begin with. Now, if the only way to do so were to preemptively deny ignorant propositions ("You may hear next week that Obama was born in Kenya. But rest assured: he was born in Hawaii"), one might worry that this strategy simply reproduces the initial problem: preemptive denial risks inadvertently reinforcing the salience of ignorant associations by repeating them.

Crucially, however, one of the principal preemptive strategies outlined above involves casting doubt on the authority of speakers, rather than directly criticizing the content of their utterances. The recommendation that emerges from the diachronic understanding of counterspeech, then, is that we should preemptively warn people about, say, fake news sites spewing conspiracy theories. We can do this by preemptively identifying specific fake news sites as unreliable; or, alternatively, by warning people that such unreliable sites exist and teaching them how to identify them. As we have seen, such warnings can drastically reduce the audience's vulnerability to misinformation.[79] So, diachronic counterspeech is doubly helpful: not only does it show how counterspeech might wholly forestall the damaging effects of ignorant speech, but it also provides guidance for handling particularly resilient kinds of misinformation.

It is important not to overstate the present section's argument. I am *not* claiming that diachronic counterspeech will always be entirely successful at defusing ignorant speech. Rather, in highlighting diachronic counterspeech, I am making a more restricted point: that, like legal remedies, counterspeech too can

---

78   Cook, Lewandowsky, and Ecker, "Neutralizing Misinformation through Inoculation," 15.

79   Cook, Lewandowsky, and Ecker, "Neutralizing Misinformation through Inoculation," 11–12.

in principle wholly defuse the harms of ignorant speech, and it can do so even in hard cases where positive counterspeech is unavailable.

That counterspeech may not do this perfectly is a problem, but not a comparative problem. The objections raised in section 2.2 suggested that, *insofar as legal remedies succeed* in deterring ignorant speech, they can wholly eliminate the harms that such speech would otherwise have produced. The degree to which they are in fact successful, however, is limited. In their influential study of hate-speech laws in Australia, for instance, Gelber and McNamara find that the enactment of hate-speech laws did not reduce the incidence of verbal abuse.[80] What this indicates is that, just as diachronic counterspeech is imperfect at disabling ignorant speech, so too legal remedies are imperfect at deterring ignorant speech. Thus, as stated, the objections in section 2.2 fail to provide a principled reason for preferring legal remedies to counterspeech.

In what follows, after summarizing my broader argument, I will briefly examine what this means for the division of labor between legal and speech-based responses to ignorant speech.

4. CONCLUSION

Whether it takes the form of hate speech or of ordinary political misinformation, ignorant speech tends to be sticky. Because of the distinctive properties of conversational salience, the damaging effects of ignorant speech on conversational norms are typically easier to enact than to reverse. In fact, verbally countering ignorant speech may simply amplify its salience.

Even so, I have argued that refining counterspeech along two dimensions can substantially mitigate this problem. First, we should distinguish between positive and negative forms of counterspeech. Instead of explicitly negating ignorant speech, positive counterspeech affirms a vision of the truth, which entails or implicates the falsehood of the ignorant utterance without repeating it. Hence, positive counterspeech contradicts ignorant speech without magnifying its salience.

The emphasis on positive counterspeech nevertheless cannot fully address the problem at hand. Not all kinds of ignorant speech can be countered in a positive way. And, more fundamentally, even when positive counterspeech helps to

---

80 Gelber and McNamara, "The Effects of Civil Hate Speech Laws," 644–45. Moreover, while they do find that the language used to express prejudice in newspapers became tamer, this—as Heinze observes—could simply be explained by hate speech taking a more coded form (*Hate Speech in Democratic Citizenship*, 145–48). For more general discussion of this problem, see Heinze, *Hate Speech in Democratic Citizenship*, 145–53; and Mchangama, Review of *The Harm in Hate Speech*, 97.

reverse the conversational effects of ignorant speech, this may not change the fact that, before they were reversed, these effects generated harms.

However, these remaining concerns rely on an unsophisticated understanding of the temporality of counterspeech. One suggestion here is that counterspeech might operate *retroactively* to prevent past ignorant utterances from ever constituting harmful speech acts. But even if we were to grant its theoretical viability, this suggestion is too narrowly focused on the harms constituted by ignorant speech to alleviate the above concern. The more decisive point, then, is that counterspeech should be understood *diachronically*, as an extended process that can both preempt and follow ignorant utterances. Preemptive counterspeech aims to condition the conversational context in a way that disables future ignorant utterances. By disarming ignorant speech before it is uttered, counterspeech can wholly eliminate its attempted harms, causal and constitutive. Moreover, preemptive counterspeech prevents especially resilient forms of ignorance, which could not be countered positively, from taking root.

This, to reiterate, is not to say that counterspeech is infallible. Positive counterparts to negative counterspeech can be difficult to find. Moreover, when an ignorant utterance has already taken root, preemptive counterspeech is no longer an option. And even when preemptive counterspeech remains an option, it is not foolproof. For instance, some speakers may be so authoritative that their verbal influence cannot fully be preemptively disabled.

Nonetheless, the point remains that the refined conception of counterspeech is substantially more resilient in the face of sticky ignorance than it initially appeared. Neither the fact that negating ignorant speech can reinforce its conversational effects, nor the fact that the causal harms of ignorant speech cannot be reversed *post hoc*, suffice to show that counterspeech is inadequate.

Accordingly, while my argument does not preclude thinking that legal remedies may sometimes be warranted to prevent sticky ignorant speech, it does establish that more needs to be said to justify such regulations. First and foremost, advocates of legal regulations need to provide more precise evidence of legal regulations' deterrent effects. If legal regulations are to be defended as a response to the stickiness of ignorant speech, its proponents must (1) offer a more specific account of the contexts in which the refined conception of counterspeech fails, and (2) show that legal regulations could successfully deter ignorant speech in those contexts. For example, they might need to show that, in contexts where the ignorance-promoting speaker's authority is so great that preemptive counterspeech would struggle to disable her speech, legal regulations would by contrast succeed in deterring her utterance. This is a much more specific challenge than simply showing that legal regulations sometimes succeed in deterring hate

speech, and one that has yet to be met. Indeed, as is often observed, even more general evidence of the deterrent effect of legal regulations—i.e., evidence that legal regulations generally tend to deter hate speech—remains sparse.[81]

But that is not all. Even if legal regulations generally succeed in preventing ignorant speech in contexts that prove problematic for counterspeech, it must also be shown that, in the process of doing so, these regulations do not inadvertently amplify the salience of ignorant speech. In other words, it must be shown that legal regulations do not run afoul of the so-called Streisand effect, whereby attempting to hide or censor something unwittingly increases its publicity.[82]

There are at least two reasons why, like negative counterspeech, legal regulations might do so. First, given the powerful expressive force of laws, publicly enacting a law prohibiting, say, degrading speech, risks drawing the public's attention to the fact that there are people who embrace and express degrading views.[83] Second, the enforcement of speech-related legal restrictions often leads to highly publicized trials that put a spotlight on violators, their utterances, and the bad associations they promote. As Simpson notably observes, in the presence of hate-speech laws, the existence of hateful citizens and hateful ideologies is "powerfully conveyed in people's preparedness to express their identity-based contempt even while faced with the threat or reality of prosecution."[84] What this underscores is that the process of implementing and enforcing laws that aim to deter ignorant speech may itself contribute to increasing the salience of such speech.

Accordingly, to establish that legal remedies are more effective ways than counterspeech of overcoming the problem of stickiness, proponents must establish not simply that their deterrent effect applies to the specific contexts where counterspeech fails, but also that the process of enacting and enforcing these laws will not amplify the salience of ignorant speech.

In this light, developing a theoretically refined understanding of counterspeech, whereby counterspeech is positively framed and extended over time, is beneficial in two respects: it offers guidance concerning how verbal responses to ignorant speech might succeed, notwithstanding its stickiness, and it clarifies

81 See note 80 above.

82 I am grateful to a reviewer for pressing me on this point.

83 For discussion of the point that laws are expressively loud, see, e.g., McAdams, *The Expressive Powers of Law*, 123.

84 Simpson, "Dignity, Harm, and Hate Speech," 724. For a similar observation, see Heinze, "Hate Speech and the Normative Foundations of Regulation," 599–600.

the empirical challenge that must be met if we are to ascertain the correct division of labor between counterspeech and legal remedies.[85]

Nuffield College, University of Oxford
maxime.lepoutre@nuffield.ox.ac.uk

REFERENCES

Alter, Adam, and Daniel Oppenheimer. "Uniting the Tribes of Fluency to Form a Metacognitive Nation." *Personality and Social Psychology Review* 13, no. 3 (August 2009): 219–35.

Anderson, Elizabeth. "The Epistemology of Democracy." *Episteme* 3, nos. 1–2 (June 2006): 8–22.

Asser, Martin. "What the Muhammad Cartoons Portray." BBC News, January 2, 2010. http://news.bbc.co.uk/2/hi/middle_east/4693292.stm.

Austin, J.L., *How to Do Things with Words.* Oxford: Oxford University Press, 1962.

Barnes, Michael Randall. "Speaking with (Subordinating) Authority." *Social Theory and Practice* 42, no. 2 (April 2016): 240–57.

Berinsky, Adam. "Rumors and Health Care Reform: Experiments in Political Misinformation." *British Journal of Political Science* 47, no. 2 (April 2017): 241–62.

Blitz, James. "Brexit and the 'NHS Dividend.'" *Financial Times*, March 14, 2018. https://www.ft.com/content/1fcaf8b0-277e-11e8-b27e-cc62a39d57a0.

Brettschneider, Corey. *When the State Speaks, What Should It Say?* Princeton: Princeton University Press, 2012.

Brown, Alexander. *Hate Speech Law: A Philosophical Examination.* Routledge: Abingdon, 2015.

Brown, Étienne. "Propaganda, Misinformation, and the Epistemic Value of Democracy." *Critical Review* 30, nos. 3–4 (2018): 194–218.

Caponetto, Laura. "Undoing Things with Words." *Synthese.* Published ahead of print, May 25, 2018. https://link.springer.com/article/10.1007%2Fs11229-018-1805-9.

Centers for Disease Control and Prevention. "Vaccines Do Not Cause Autism." Centers for Disease Control and Prevention, October 27, 2015. https://www.cdc.gov/vaccinesafety/concerns/autism.html.

Cook, John, Stephan Lewandowsky, and Ullrich Ecker. "Neutralizing Misinformation through Inoculation." *PLOS One* 12, no. 5 (May 2017): 1–21.

Delgado, Richard. "Words that Wound: A Tort Action for Racial Insults, Epithets, and Name-Calling." *Harvard Civil Rights-Civil Liberties Law Review* 17 (1982): 133–82.

Gelber, Katharine. "Reconceptualizing Counterspeech in Hate Speech Policy (with a Focus on Australia)." In *The Content and Context of Hate Speech*, edited by Michael Herz and Peter Molnar, 198–216. Cambridge: Cambridge University Press, 2012.

Gelber, Katharine, and Luke McNamara. "The Effects of Civil Hate Speech Laws: Lessons from Australia." *Law and Society Review* 49, no. 3 (September 2015): 631–64.

Godlee, Fiona, Jane Smith, and Harvey Marcovitch. "Wakefield's Article Linking MMR Vaccine and Autism Was Fraudulent." *BMJ* 342, no. 7788 (January 2011). https://www.bmj.com/content/342/bmj.c7452.

Heinze, Eric. "Hate Speech and the Normative Foundations of Regulation." *International Journal of Law in Context* 9, no. 4 (December 2013): 590–617.

———. *Hate Speech in Democratic Citizenship*. Oxford: Oxford University Press, 2016.

Holan, Angie. "PolitiFact's Lie of the Year: 'Death Panels.'" Politifact, December 18, 2009. https://www.politifact.com/truth-o-meter/article/2009/dec/18/politifact-lie-year-death-panels.

King, Martin Luther. "I Have a Dream." Speech presented at the March on Washington for Jobs and Freedom, Washington, DC, August 1968. https://americanrhetoric.com/speeches/mlkihaveadream.htm.

Landemore, Hélène. *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton: Princeton University Press, 2013.

———. "Yes, We Can (Make It Up on Volume)." *Critical Review* 26, nos. 1–2 (2014): 184–237.

Langton, Rae. "Blocking as Counterspeech." In *New Work on Speech Acts*, edit-

ed by Daniel Fogal, Daniel Harris, and Matt Moss, 144–62. Oxford: Oxford University Press, 2018.

———. "Speech Acts and Unspeakable Acts." *Philosophy and Public Affairs* 22, no. 4 (Autumn 1993): 293–330.

Langton, Rae, Sally Haslanger, and Luvell Anderson. "Language and Race." In *Routledge Companion to Philosophy of Language*, edited by Gillian Russell and Fara Delia Graff, 753–67. Abingdon: Routledge, 2012.

Lepoutre, Maxime. "Hate Speech in Public Discourse: A Pessimistic Defense of Counterspeech." *Social Theory and Practice* 43, no. 4 (October 2017): 851–83.

Leslie, Sarah-Jane, "The Original Sin of Cognition: Fear, Prejudice, and Generalization." *Journal of Philosophy* 114, no. 8 (August 2017): 393–421.

Lewandowsky, Stephan, Ulrich Ecker, Colleen Seifert, Norbert Schwarz, and John Cook. "Misinformation and Its Correction: Continued Influence and Successful Debiasing." *Psychological Science in the Public Interest* 13, no. 3 (December 2012): 106–31.

Lewis, David. "Scorekeeping in a Language Game." *Journal of Philosophical Logic* 8, no. 1 (January 1979): 339–59.

Maitra, Ishani. "Subordinating Speech." In Maitra and McGowan, *Speech and Harm*, 94–118.

Maitra, Ishani, and Mary Kate McGowan. "Introduction and Overview." In Maitra and McGowan, *Speech and Harm*, 1–23.

Maitra, Ishani and Mary Kate McGowan, eds. *Speech and Harm*. Oxford: Oxford University Press.

McAdams, Richard. *The Expressive Powers of Law.* Cambridge, MA: Harvard University Press, 2015.

McGowan, Mary Kate. "Debate: On Silencing and Sexual Refusal." *Journal of Political Philosophy* 17, no. 4 (2009): 487–94.

———. *Just Words: On Speech and Hidden Harm.* Oxford: Oxford University Press, 2019.

———. "Oppressive Speech." *Australasian Journal of Philosophy* 87, no. 3 (2009): 389–407.

———. "Responding to Harmful Speech." In *Voicing Dissent*, edited by Casey Rebecca Johnson, 182–200. Abingdon, UK: Routledge, 2018.

Mchangama, Jacob. Review of *The Harm in Hate Speech*, by Jeremy Waldron. *Policy Review* 176 (2012): 95–102.

Nelson, Sara. "Katie Hopkins Reflects on Branding Migrants 'Cockroaches' and 'Feral Humans.'" *Huffington Post*, July 27, 2015. https://www.huffingtonpost.co.uk/2015/07/27/katie-hopkins-reflects-migrants-cockroaches-feral-humans_n_7877124.html.

Nielsen, Laura Beth. "Power in Public." In Maitra and McGowan, *Speech and Harm*, 148–73.

Post, Robert. *Constitutional Domains: Democracy, Community, Management.* Cambridge, MA: Harvard University Press, 1995.

Reilly, Katie. "Here Are All the Times Donald Trump Insulted Mexico." *Time*, August 31, 2016. https://time.com/4473972/donald-trump-mexico-meeting -insult.

Richards, Robert, and Clay Calvert. "Counterspeech 2000: A New Look at the Old Remedy for 'Bad Speech.'" *Brigham Young University Law Review* 2000, no. 2 (2000): 553–86.

Simpson, Robert Mark. "Dignity, Harm, and Hate Speech." *Law and Philosophy* 32, no. 6 (November 2013): 701–28.

———. "Un-Ringing the Bell: McGowan on Oppressive Speech and the Asymmetric Pliability of Conversations." *Australasian Journal of Philosophy* 91, no. 3 (2013): 555–75.

Smith, Ben, and Byron Tau. "Birtherism: Where It All Began." Politico, April 24, 2011. https://www.politico.com/story/2011/04/birtherism-where-it-all -began-053563.

Sparrow, Andrew. "UK Statistics Chief Says Vote Leave £350m Figure Is Misleading." *The Guardian*, May 27, 2016. https://www.theguardian.com/ politics/2016/may/27/uk-statistics-chief-vote-leave-350m-figure-mislead ing.

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

Strossen, Nadine. *Hate: Why We Should Resist It with Free Speech, Not Censorship.* Oxford: Oxford University Press, 2018.

Tirrell, Lynne. "Toxic Speech: Inoculations and Antidotes." *Southern Journal of Philosophy* 56, no. S1 (December 2018): 116–44.

Tversky, Amos, and Daniel Kahneman. "The Framing of Decisions and the Psychology of Choice." *Science* 211, no. 4481 (January 30, 1981): 453–58.

Waldron, Jeremy. *The Harm in Hate Speech.* Cambridge, MA: Harvard University Press, 2012.

Young, Iris Marion. *Inclusion and Democracy.* Oxford: Oxford University Press, 2000.

INCOME REDISTRIBUTION, BODY-PART REDISTRIBUTION, AND RESPECT FOR THE SEPARATENESS OF PERSONS

Joseph Mazor

THERE IS a broad consensus among both political philosophers and the general public that, though the redistribution of labor income is permissible, the redistribution of body parts is not. The question of why this is so, which I call the *body-income puzzle*, lies at the heart of several important debates regarding the distribution of human organs, just taxation, and the limits of what can be permissibly demanded of the advantaged. It will be my focus in this article.

Some scholars argue that there is no solution to the body-income puzzle. On this view, which is defended by both Robert Nozick and Cécile Fabre, there is no compelling, principled way of distinguishing the redistribution of labor income to the poor from the redistribution of nonessential body parts to the sick.[1] Nozick and Fabre draw opposing conclusions from this position, with Nozick rejecting redistributive taxation and Fabre endorsing (at least in principle) the redistribution of certain nonessential body parts. However, the intuitive implausibility of both of these conclusions recommends the body-income puzzle for further investigation.

A variety of potential solutions to this puzzle have been proposed. Some argue that the solution lies in affirming individuals' self-ownership while denying individuals' ownership of their labor income.[2] Others appeal to the particularly high welfare losses associated with bodily takings.[3] I will argue that these solutions are unsuccessful.

I will argue instead that taking individuals' body parts is more objectionable than taking moderate percentages of their labor income because it constitutes a greater disrespect for their separateness as persons. Respect for the separateness of persons is the idea that each individual's life should be granted a certain

---

1 Nozick, *Anarchy, State, and Utopia*, 167–74, 205–6, 272–73. See also Fabre, *Whose Body Is It Anyway?* ch. 5.

2 For a discussion of this approach and its proponents, see section 2.

3 For a discussion of this approach and its proponents, see section 3.1.

primacy over the demands of the collective good.[4] Such a primacy can be understood in terms of a sphere of moral authority that is protected from others' demands for assistance. The conception of respect for the separateness of persons that I defend holds that a person's body and her labor income both fall within this protected sphere of authority. And it views the strength of an individual's rights against interference in this sphere as non-absolute, variable, and dependent on consequences that the interference would have for the individual's personal identity. Since our body parts—even internal, nonessential ones— are more intimately connected to our personal identity compared to moderate portions of our labor income, our separateness-based rights against body-part takings are particularly strong. Thus, while the weighty claims of the disadvantaged to assistance can justify the redistribution of substantial amounts of labor income, they are generally insufficiently weighty to justify the redistribution of substantial parts of others' bodies.

1. THE BODY-INCOME PUZZLE

Consider the following pair of cases:

> *Rich-Poor*: Rich and Poor live in a society with plentiful arable land and no way of making a living other than farming it. Rich is an excellent farmer who can enjoy high income through just exchanges of his farming produce with others.
>
> Poor is physically unable to farm intensively through no fault or choice of her own. She can eke out a marginal living, leaving her with poor nutrition, inadequate shelter, and no access to healthcare. Receiving 20 percent of Rich's income would enable Poor to lead a minimally flourishing life.[5]
>
> *Healthy-Sick*: Healthy and Sick live in a society similar to that of Rich and Poor. Like Rich, Healthy is an excellent farmer who enjoys a high income. She has two well-functioning kidneys.
>
> Sick suffers from kidney failure through no fault or choice of his own. In addition to a variety of maladies associated with kidney failure, Sick must endure dialysis several times per week.[6] Sick's overall quality of

---

4   For a discussion, see Mack, "Non-Absolute Rights and Libertarian Taxation," 115–19. See also Zwolinski, "The Separateness of Persons and Liberal Theory," 150–53.

5   Fabre endorses a minimally flourishing life as the aim of redistribution (*Whose Body Is It Anyway?* 31). Readers with different views can adjust the examples accordingly.

6   For a description of dialysis, see National Kidney Foundation, "A 'New Normal.'"

life is thus very low—similar to that of Poor's. One of Healthy's kidneys would enable Sick to lead a minimally flourishing life.

This pair of cases brings the body-income puzzle into sharp relief. I take it that the state should redistribute 20 percent of Rich's labor income to Poor but should not redistribute Healthy's kidney to Sick. The puzzling question is this:

> *Body-Income Puzzle*: Why should society condone substantial redistribution of labor income from the economically well-off to the indigent while rejecting the redistribution of nonessential body parts from the healthy to the infirm?

Any solution to this puzzle will have to meet several challenges. First (and most importantly), it will have to explain why the redistribution of nonessential body parts is more objectionable than the redistribution of labor income. Admittedly, an explanation of why the infirm's claims to assistance are weaker than the indigent's would also do. However, since many of those who need others' body parts are at least as heavily disadvantaged (often through no fault or choice of their own) as those who need financial assistance, this route to solving the body-income puzzle is unpromising, and I shall not consider it further here.[7]

Second, a solution to the body-income puzzle must grant body parts strong protection from redistribution. This is no easy task. It is not enough to offer *some* reason in favor of protecting body parts. The reason must be sufficiently weighty to justify resisting the infirm's pressing claims to assistance.

Third, a solution to the body-income puzzle must justify substantial levels of labor-income redistribution to the poor. This, too, is no easy task. After all, many theorists have argued that individuals have strong rights to the fruits of their labor—rights that are particularly difficult to deny while also affirming individuals' strong rights to their bodies.[8]

These are the three challenges posed directly by the body-income puzzle. However, a compelling solution to this puzzle must meet two additional challenges. First, it must place attractive *limits* on the protections offered to individuals' bodies. To see the need for such limits, consider the following case:

> *Drop of Blood to Save a Billion*: A deadly pandemic has killed millions of people and is predicted to kill a billion more. The pandemic can only be stopped by studying a drop of blood from Adam (the sole survivor of the disease). Adam refuses to surrender a drop of his blood.

7   For a discussion, see Fabre, *Whose Body Is It Anyway?* 77.
8   See, for example, Wheeler, "Natural Property Rights as Body Rights."

I take it that Adam can be permissibly coerced into surrendering a drop of blood in this case. Thus, while a theory of distributive justice that condemned all violations of the body, no matter how minor and no matter what else was at stake, could technically solve the body-income puzzle, such a solution would be uncompelling.

A compelling solution to the body-income puzzle must also place attractive limits on labor-income taxation. To see the need for such limits, consider the following policy:

> *Two-Kidney Tax*: In Healthy-Sick, above, the state does not demand that Healthy surrender her kidney. Instead, it requires any person with two healthy kidneys to turn over all of her labor income to the state. Faced with this tax, Healthy transfers one of her kidneys to Sick.

I take it that this Two-Kidney Tax is impermissible. Thus, while a theory of distributive justice that offered no protections whatsoever to individuals' labor income could technically solve the body-income puzzle, such a solution would not be compelling, either.

2. INCOME-EXCLUSION SOLUTIONS

I now turn to considering several proposed solutions to the body-income puzzle, beginning with (what I call) *income-exclusion solutions*. Proponents of these solutions affirm individuals' stringent *self-ownership*—their strong (but non-absolute), extensive, unconditional property rights to their bodies.[9] But they deny that individuals own their labor income due to its reliance on the appropriation of the natural world, transactions with others, and/or government-provided goods.[10] Thus, labor income (but not body parts) can be permissibly redistributed to alleviate others' disadvantage.

Though this type of solution has substantial *prima facie* plausibility, it fails to meet the challenges laid out in section 1. First, the distinctions that income-exclusion solutions draw between labor income and body parts are dubious. As other critics have pointed out, like our labor income, our bodies also depend on natural and social inputs.[11] After all, our bodies constantly regenerate, and the new cells are physically made up of molecules generated from the food we eat—

---

9   See, for example, Otsuka, *Libertarianism without Inequality*, 15.

10 See Otsuka, *Libertarianism without Inequality,* ch. 1; Vallentyne's view in Tideman and Vallentyne, "Left-Libertarianism and Global Justice"; Christman, "Self-Ownership, Equality, and the Structure of Property Rights," 30–31; Murphy and Nagel, *The Myth of Ownership*.

11 See Eyal, "Is the Body Special?" 236.

food that we obtained via uses of the natural world, transactions with others, and/or uses of government-provided goods. If labor income's reliance on these inputs makes it liable to redistribution, it is unclear why the body's reliance on these inputs does not also make it liable to redistribution.

Proponents of income-exclusion solutions might respond by appealing to self-ownership, which protects food molecules from redistribution once they are part of our bodies. However, self-ownership does not preclude placing various redistributive demands on individuals' bodies as a condition of using natural or social inputs. Just as I can demand Healthy's kidney in exchange for, say, my apples without violating Healthy's self-ownership, so too a policy requiring Healthy to surrender her kidney as a condition of appropriating natural resources, transacting with others, or using government-provided goods would not violate Healthy's self-ownership.

Proponents of income-exclusion solutions might respond that, while self-ownership is technically consistent with placing body-part redistribution conditions on the use of natural and social inputs, the moral considerations underlying self-ownership rule out such conditions. However, this response is difficult to evaluate without a comprehensive justification for self-ownership—a justification that many proponents of income-exclusion solutions do not provide.[12] And it is far from obvious that a justification for self-ownership that condemned placing body-part redistribution conditions on the use of natural and social inputs would condone placing labor-income redistribution conditions on the use of these inputs.

Another problem with income-exclusion solutions is that the redistribution of labor-generated wealth seems justified even when this wealth does not depend on natural and social inputs. Consider, for example, the following case:

> *Hairy-Bald*: Hairy and Bald are both created *ex nihilo* with fully developed life plans. They do not need air, water, or external nutrition.
>
> Detached hair is the only material that can be used to make goods. Hair removal is painless but laborious, knowledge for turning detached hair into goods is instinctive, and hair-made goods can be protected without state intervention.
>
> Hairy has sufficient hair to enjoy a high living standard (similar to the standard enjoyed by Healthy and Rich). Hairy refuses to give or trade any of his detached hair.
>
> Bald, who lacks hair through no fault or choice of her own, can only gather sufficient discarded hair to achieve a marginal living standard

12   See, for example, Otsuka, *Libertarianism without Inequality*, ch. 1.

(similar to that of Poor and Sick). Twenty percent of Hairy's detached hair per year would enable Bald to lead a minimally flourishing life.

In this example, labor-generated wealth is produced independently of any social institutions, transactions, or uses of the natural world.[13] If these factors' use is what fundamentally distinguishes labor income from body parts, then the redistribution of Hairy's detached hair should be impermissible. Yet I submit that the redistribution of Hairy's detached hair is permissible (for reasons that I discuss below).[14] If so, this suggests that income-exclusion solutions fail to capture the foundational moral distinction between labor-generated wealth and body parts when it comes to redistribution permissibility.

Another problem with income-exclusion solutions is the insufficient justification they offer for the protection level they grant to the body. The redistribution of Healthy's kidney can be rejected and the redistribution in Drop of Blood to Save a Billion can be endorsed *if* we accept that self-ownership is stringent but non-absolute. Yet proponents of income-exclusion solutions often merely assert this level of self-ownership stringency.[15] It is thus unclear why self-ownership rights are sufficiently weighty to justify resisting Sick's weighty claims to Healthy's kidney but insufficiently weighty to protect Adam's drop of blood from the moral demands of the billion.

Finally, income-exclusion solutions are unable to offer attractive levels of protection to individuals' labor income. It is unclear, for example, how these solutions can reject the Two-Kidney Tax proposed above. After all, this tax does not infringe on Healthy's body. And if labor income is unowned and can therefore be distributed to help the disadvantaged, it is unclear why society should not structure the distribution of income so as to induce Healthy to transfer one of her kidneys to Sick.

Proponents of income-exclusion solutions might respond by appealing to Healthy's *substantive* self-ownership to reject this tax. While formal self-ownership protects individuals from bodily interference, substantive self-ownership grants individuals rights that enable them to make use (or to not have to make use) of their bodies in certain ways. Michael Otsuka, for example, endorses a principle of substantive self-ownership that grants individuals "rights over enough worldly resources to ensure that [they] will not be forced by necessity to come to the assistance of others in a manner involving the sacrifice of one's

---

13   The exception is space, which I will assume is perfectly plentiful.

14   See section 6.4.

15   See, for example, Otsuka, *Libertarianism without Inequality*, 15. For a more detailed argument for the need for a self-ownership stringency justification, see Lippert-Rasmussen, "Against Self-Ownership."

life, limb, or labor."[16] Since Healthy's life would presumably be threatened if she had to surrender all of her labor income to the government, the Two-Kidney Tax would violate Healthy's substantive self-ownership as Otsuka defines it.

However, appealing to substantive self-ownership to reject the Two-Kidney Tax is problematic for three reasons. First, while this tax may well undermine Healthy's substantive self-ownership, *not* implementing this tax seems to undermine Sick's substantive self-ownership to an even greater extent. After all, without Healthy's kidney, Sick will have to live with severe physical limitations as well as a substantial risk of death. Admittedly, Otsuka's conception of substantive self-ownership, which only grants individuals rights to "worldly resources" and is only concerned with protecting individuals from being forced to take certain actions, would not grant Sick claims to have the Two-Kidney Tax implemented. Yet it is unclear why substantive self-ownership should be defined in Otsuka's way. After all, a variety of values that could be used to justify substantive self-ownership (e.g., autonomy, self-determination) seem more heavily compromised by Sick's serious physical limitations and substantial risk of death than by Healthy's being forced by necessity to transfer one kidney to Sick.[17]

Second, even if the Two-Kidney Tax undermines Healthy's substantive self-ownership more than it fosters Sick's, this would still not necessarily justify rejecting this tax. After all, Sick presumably has powerful disadvantage-based moral claims to have this tax implemented (e.g., based on the very low welfare levels he must endure through no fault or choice of his own). Thus, unless it can be established that Healthy's substantive self-ownership has a great deal of moral weight (a task that Otsuka does not undertake), an appeal to Healthy's substantive self-ownership will be insufficient to reject the Two-Kidney Tax.

Finally, even if Otsuka could justify granting substantive self-ownership as he defines it a great deal of moral weight, this would still not ground intuitively compelling protections to individuals' labor income. To see why not, consider the following tax:

> *Two-Kidney Tax with a Necessitarian Cap*: The government makes the Two-Kidney Tax somewhat less extreme by placing a cap on it. The capped tax would leave those who have two kidneys with just enough in-

16 Otsuka, *Libertarianism without Inequality*, 32.

17 We could make this objection even more powerful by assuming that Sick will die without Healthy's kidney. I take it that the Two-Kidney Tax would nevertheless be impermissible. Yet it seems even more difficult in this case to affirm that the effects of the tax on Healthy's substantive self-ownership would be greater than the effects of not implementing the tax on Sick's substantive self-ownership.

come to afford necessities if they refuse to surrender one of their kidneys. To avoid this tax, Healthy transfers one of her kidneys to Sick.

I submit that, even with the cap, this two-kidney-based tax is impermissible. Yet this tax respects Healthy's substantive self-ownership as Otsuka defines it (since Healthy would not be forced *by necessity* to transfer body parts to anyone else). And while Otsuka could respond by adjusting his conception of substantive self-ownership to rule out this tax, it is not clear that substantive self-ownership can be compellingly defined in a way that simultaneously grants Healthy rights against the Two-Kidney Tax with a Necessitarian Cap, does not grant Sick even weightier rights to have this tax implemented, and does not protect Rich from labor-income taxation.

Proponents of income-exclusion solutions could alternatively reject heavy two-kidney-based taxation by setting limits on the amount of labor income that can be demanded of a person due to her use of natural and social inputs. Peter Vallentyne, for example, holds that an individual's redistributive tax burden should be limited by the market value of the natural resources she has appropriated.[18] Since the market value of the plentiful arable land in Healthy-Sick is zero (or close to it), Vallentyne's proposal would protect Healthy from both the Two-Kidney Tax and the Two-Kidney Tax with a Necessitarian Cap.

However, Vallentyne's proposed limit is problematic for two reasons. First, he does not offer a principled justification for this limit. It is therefore unclear why this limit should be respected given the very weighty benefits that heavy two-kidney-based taxation would have for Sick. Second, while Vallentyne's limit rules out the unattractive two-kidney-based taxation, it also rules out attractive levels of redistributive taxation in certain cases. Note, for example, that the land in Rich-Poor is also plentiful and thus would have a market value of zero (or close to it). The level of labor-income redistribution permitted by Vallentyne's theory would therefore be insufficient to meet Poor's intuitively attractive claims to assistance in this case. Thus, it is far from obvious that attractive limits on labor-income taxation can be compellingly defended by proponents of income-exclusion solutions.

In sum, income-exclusion solutions seem incapable of meeting the challenges posed in section 1. The distinctions they draw between body parts and labor income are suspect. The protections they offer the body lack sufficient justification. And the protections they offer to labor income are either implausibly strong or implausibly weak.

18   Tideman and Vallentyne, "Left-Libertarianism and Global Justice."

3. TWO BODY-SPECIALNESS SOLUTIONS: WELFARE AND PERSONAL IDENTITY

Other proposed solutions to the body-income puzzle do not deny that individuals have rights to their labor income. They aim instead to explain what it is about the infringement of individuals' rights to their body parts that is so much more objectionable than the infringement of their rights to their labor income. I call these *body-specialness solutions.* In this section, I criticize two prominent solutions of this sort.

3.1. *The Appeal to Welfare Loss*

The first solution I wish to consider appeals to the particularly high *welfare loss* associated with bodily takings. Note that even the taking of a nonessential, internal body part like a kidney entails considerable welfare loss. The kidney's loss imposes nonnegligible risks and lifestyle restrictions on the individual whose kidney is taken.[19] The recovery from the operation is painful and involves considerable physical limitations.[20] And the thought of someone else cutting deep into one's body to extract an organ is disconcerting and, for many, deeply humiliating.[21] Perhaps, then, taking Healthy's kidney is more objectionable than taking Rich's labor income simply because it entails a greater loss of welfare.

However, this type of welfarist body-specialness solution faces several problems. First, even if welfarist considerations can explain why taking Healthy's kidney is more objectionable than taking Rich's income, these considerations cannot explain why Sick should be denied Healthy's kidney. After all, the welfare loss from having a kidney surgically removed cannot plausibly outweigh the welfare *gain* to Sick from no longer having to live with kidney failure, with the associated severe physical limitations, health problems, and routine bodily intrusions required by dialysis.[22] Those who appeal to a welfarist account of the special objectionability of body-part takings will therefore need a non-welfarist account of individuals' rights to keep their body parts in the face of others' claims to assistance. Yet it is unclear how these two commitments can be compellingly combined.

A second problem with welfarist body-specialness solutions is that, as Fabre points out, the welfare loss associated with having certain body parts surgically removed (at least under modern medical conditions) does not seem to be

---

19   For a list of risks, see Johns Hopkins Medicine, "What to Expect as a Living Donor."

20   For a description of the recovery process, see Johns Hopkins Medicine, "What to Expect as a Living Donor."

21   Eyal, "Is the Body Special?" 237–38.

22   For a description, see National Kidney Foundation, "A 'New Normal.'"

greater than the welfare loss associated with being deprived of, say, 20 percent of one's labor income *annually*.[23] To make matters concrete, assume that Rich and Healthy have identical preferences, have a guaranteed pretax annual income of $100,000, work for fifty years, can save and borrow as much money as they want at the going interest rate, and the interest rate and inflation rate are both zero. I take it that, even in this case, taking Healthy's kidney is more objectionable than taking 20 percent of Rich's annual income. Yet to explain why this is so, a proponent of a welfarist body-specialness solution would have to claim that taking Healthy's kidney entails a greater welfare loss than depriving Rich of $1 million of lifetime earnings—*ten years' salary*. This seems difficult to accept.

Indeed, we need not rely solely on intuition to support this skepticism. Gary Becker and Julio Jorge Elias have estimated what it would take to induce a person to part with her kidney. The figure they defend is a lump sum of $15,200 for a person with an annual income of $35,000 (i.e., under half of one year's salary).[24] Becker and Elias's estimate can admittedly be criticized for excluding several risks and negative consequences associated with surgical removal of a kidney.[25] It will also need to be adjusted upward to take account of Rich's greater income. Yet it is difficult to believe that the appropriate adjustments would raise Becker and Elias's $15,200 figure above $1 million.

Moreover, even if taking Healthy's kidney constitutes a greater welfare loss than taking 20 percent of Rich's income, we can fairly easily imagine other cases in which taking body parts clearly entails a smaller welfare loss while nevertheless seeming more objectionable. Consider, for example, a variation of Rich-Poor in which Rich is an extreme miser who experiences very high welfare losses whenever *any* of his money is taken from him (assume also that Rich did not consciously develop his miserly attitude and cannot help it). I submit that the redistribution of 20 percent of Rich-the-miser's labor income to help Poor would nevertheless be permissible. Yet it is unclear how a proponent of a welfarist body-specialness solution can explain why this is so.

Alternatively, consider the following case:

Dismantling the Bedridden: Ben is suffering from an incurable but non-life-threatening brain tumor that routinely induces weeks-long peri-

---

23 Fabre, *Whose Body Is It Anyway?* 103–4.

24 Becker and Elias, "Introducing Incentives in the Market for Live and Cadaveric Organ Donations," 11.

25 See, for example, the discussion of mental health risks in section 5 below. Becker and Elias's estimate would also need to be adjusted upward due to the special welfare loss associated with a *coerced* taking of a kidney.

ods of unconsciousness and requires Ben to remain hooked up to various machines.

Given his condition, Ben can be subjected to the taking of various body parts—a kidney, portions of his pancreas, liver lobes (multiple times once they regenerate), parts of a lung, and bone marrow (multiple times)—all without significant consequences for Ben's quality of life (as long as the takings occur during his periods of unconsciousness). More- over, removing these body parts while Ben is alive substantially increases their usefulness to others.

Unbeknownst to Ben, his doctors remove these body parts (and re- place them with biosafe materials to ensure the effects are unnoticeable).

I submit that Dismantling the Bedridden is highly objectionable. Yet taking Ben's organs has no discernable effect on Ben's health, quality of life, or self-con- ception. It is thus difficult to argue that Ben's welfare loss is greater than Rich's. If what is done to Ben is nevertheless more objectionable, this suggests that the greater protections due to the human body cannot be fundamentally grounded in the greater welfare loss associated with bodily takings.

3.2. Personal Identity, Body Parts, and Labor Income

A second body-specialness solution that can better contend with these types of cases is proposed by Margaret Radin in "Property and Personhood."[26] Radin ar- gues that a person's entitlement to a resource is stronger the more important the resource is to the individual's *personal identity*—to those features that make her the particular person that she is.[27] In Radin's view, the features that define our personal identity include, not only the characteristics of our physical bodies, but also "our particular commitments and character traits, our particular memories and future plans, and our particular relationships with other people and with the world of external objects."[28]

Radin then argues that a person's money is not as closely connected to her personal identity as her body is. A dollar, Radin points out, is only personally im- portant due to what it can buy—one dollar bill is generally no more personally

---

26 Radin, "Property and Personhood."

27 Radin, "Property and Personhood," 986. Radin uses the term "personhood" to denote what I am calling "personal identity." I do not use the term "personhood" because it could mis- leadingly suggest that the protected resources are metaphysically part of the person. For a general discussion of personal identity in the sense used here, see Olson, "Personal Identity," sec. 1.

28 Radin, "Property and Personhood," 972.

important than any other.[29] On the other hand, we are constituted by our bodies, and our organs are thus intimately connected with our personal identity.[30] The taking of our body parts therefore constitutes a greater *personal violation*—that is, it infringes in an unwelcome way on a right with greater importance to our personal identity—compared with the taking of our money—and is for that reason more objectionable.[31]

Importantly, Radin defends a conception of personal identity that is not purely subjective.[32] This enables her to condone the taking of a miser's income, even if the miser feels an intimate personal connection with every dollar in his bank account. Radin argues that such attachments are best understood as a kind of fetish rather than as a genuine connection of personal identity.[33] The nonsubjective nature of Radin's personal identity conception also enables her to explain the objectionability of policies like Dismantling the Bedridden. Although Ben's self-conception is unaffected by the removal of his organs (since he is unaware of these removals), from an objective perspective, society has taken resources (e.g., several of Ben's internal organs) that are (in Radin's view) important parts of Ben. Radin would thus be able to affirm that what is done to Ben is highly objectionable.

Radin's solution could be challenged by questioning her nonsubjective conception of personal identity or the distinction she draws between fetishes and genuine personal attachments. However, for the purpose of this article, I will take these features of Radin's theory as given. I wish to focus instead on three other problems with Radin's solution to the body-income puzzle.

First, Radin's claim that the taking of body parts constitutes a graver personal violation than the taking of labor income faces important objections that Radin does not consider. As Fabre argues, there are several reasons for doubting the importance that internal, nonessential organs such as a kidney have for our personal identity. For one thing, the consequences of a kidney's loss for individuals' life plans are fairly minor.[34] For another, on purely psychological conceptions of the person, organs that are largely redundant (e.g., a second kidney) would not necessarily even count as part of the person.[35] Finally, even if a kidney is part of the person, it is far less important to her personal identity than, say, a hand, be-

---

29 Radin, "Property and Personhood," 960.

30 Radin, "Property and Personhood," 966.

31 Radin, "Property and Personhood," 1005. Note that the term "personal violation" is mine.

32 Radin, "Property and Personhood," 968–70.

33 Radin, "Property and Personhood," 970.

34 Fabre, *Whose Body Is It Anyway?* 104–5.

35 For a discussion of this view of the person, see Fabre, *Whose Body Is It Anyway?* 12.

cause a kidney is neither necessary for perceiving or interacting with the world nor does its loss affect how the person is seen by others.[36]

On the other side of the ledger, the personal connections between a person and her labor income are potentially stronger than Radin acknowledges. After all, Rich might have used the redistributed income to accomplish important life goals (e.g., pursuing a somewhat expensive but personally meaningful hobby). Moreover, the fact that Rich *worked* for his income—the fact that it is the result of his purposeful mental and physical activity—is a plausible source of substantial personal attachment to it. Thus, while Radin's claim that our body parts are more important than our labor income to our personal identity is not necessarily wrong, it is insufficiently defended.

A second problem with Radin's solution is the dearth of theoretical justifications she offers for granting personal identity fundamental importance in matters of distributive justice. Radin highlights several legal practices that can be plausibly explained by the connections between personal identity and entitlements to resources.[37] But she offers almost no *principled* reasons for viewing personal identity (rather than welfare, autonomy, freedom, capabilities, self-respect, or other forms of advantage) as a key determinant of individuals' entitlements to resources.[38]

Third (and most importantly), Radin's solution cannot protect Healthy's kidney from redistribution. To see why, consider a variation of Healthy-Sick in which Sick will die in one year without Healthy's kidney. I submit that, even in this case, coercively redistributing Healthy's kidney to Sick is impermissible. Yet it is very difficult to see how Radin could argue that Healthy's kidney is more important to Healthy's personal identity than it is to Sick's. After all, Sick will *cease to exist as a person* without Healthy's kidney. Thus, Radin's claim that the strength of our entitlements to resources depends on these resources' importance to our personal identity cannot compellingly solve the body-income puzzle.

4. RESPECT FOR THE SEPARATENESS OF PERSONS

There is, however, an alternative body-specialness solution that can avoid the problems highlighted above. I will argue in the rest of this article that taking Healthy's kidney is more objectionable than taking Rich's income because it entails a greater disrespect for the separateness of persons.

Respect for the separateness of persons (RSP), as I understand it, is the idea that

36   Fabre, *Whose Body Is It Anyway?* 12.
37   Radin, "Property and Personhood," 991–1013.
38   For notable exceptions, see Radin, "Property and Personhood," 960, 968.

each individual's life should be granted a certain primacy over considerations of the collective good.[39] The collective good can be defined in a variety of ways (e.g., in terms of maximizing aggregate advantage, achieving equality of advantage, or ensuring sufficient advantage for every person).[40] The unifying feature of collective-good conceptions is the proposition that greater good can sometimes be achieved by demanding sacrifices from one person for the sake of others. Respect for the separateness of persons serves as a bulwark against such demands.

RSP has been defended in a variety of ways. Warren Quinn, for example, offers the following justification for this value:

> A person [should have] primary say over [his life]—not because such an arrangement best promotes overall human welfare, but because any arrangement that denied him that say would be a grave indignity. In giving him this authority, morality recognizes his existence as an individual with ends of his own—an independent *being*. Since that is what he is, he deserves this recognition. Were morality to withhold it, were it to allow us to kill or injure him whenever that would be collectively best, it would picture him not as a being in his own right but as a cell in the collective whole.[41]

Other justifications for RSP differ from Quinn's in certain details. But they all appeal to the status of human beings as distinct individuals, each with his or her own ends.[42]

RSP can also be defended by highlighting its capacity to make sense of our strongly held judgments about particular cases. Consider, for example, the well-known thought experiment in which five individuals, each suffering from a different type of fatal organ failure, could be saved by killing one healthy person and redistributing her organs to the five. Although a variety of collectivist objectives (e.g., increasing aggregate welfare) would be advanced by killing the healthy person in this case, many have argued that doing so would unacceptably

---

39 There is a much weaker version of respect for the separateness of persons that only demands that we consider the distribution of welfare among different individuals. For a discussion of this weaker version and a defense of the version of RSP that I endorse, see Mack, "Non-Absolute Rights and Libertarian Taxation," sec. 2.

40 "Advantage" can be understood as a positive attribute or combination of attributes of a person or her circumstances that is relevant to distributive justice. For this usage, see Cohen, "On the Currency of Egalitarian Justice," 907, 916–17.

41 Quinn, "Actions, Intentions, and Consequences," 309–10.

42 See, for example, Zwolinski, "The Separateness of Persons and Liberal Theory," 153–54.

eny her the primacy she is due as a separate, independent being with her own mind, body, identity, projects, attachments, and life to lead.[43]

One way of cashing out this primacy is by seeing each person as having a special sphere of moral authority—a *personal domain*—constituted by a set of negative rights—*domain rights*—that protect the person from being sacrificed for others' sake. The healthy individual in the case above, for example, could be understood as having domain rights to her body—rights that protect her from having her organs seized, even when this is the only way of saving five others.

Evaluating RSP's capacity to solve the body-income puzzle will clearly require greater specificity regarding the nature of these rights. However, this value's potential for solving the body-income puzzle is not difficult to see. As the case above demonstrates, RSP can provide substantial protections to the body. Indeed, it can be understood as a justification for formal self-ownership.[44] Moreover, unlike appeals to substantive self-ownership, welfarism, and identity-based claims to resources, RSP protects the body, not by appealing to the moral importance of a well-functioning body, but rather by setting limits on *interference* in individuals' bodily integrity. It can therefore protect healthy individuals from bodily takings without concomitantly strengthening the infirm's moral claims to the body parts they need for *their* bodily functioning. RSP-based solutions to the body-income puzzle therefore clearly merit further consideration.

5. THE ADROIT CONCEPTION AND THE BODY-INCOME PUZZLE

While the idea of a body-income-puzzle solution grounded in respect for the separateness of persons has considerable *prima facie* plausibility, it also faces important challenges. For one thing, both Fabre and Nozick endorse RSP while denying that a solution to the body-income puzzle exists.[45] Thus, if a compelling RSP-based solution to this puzzle can be found, it will need to appeal to a conception of RSP different from the conceptions endorsed by Nozick and Fabre.

Competing conceptions of RSP disagree about two key issues. First, they disagree about the *scope* of the personal domain—the extent of the sphere of individual moral authority protected by RSP. Second, they disagree about the *strength* of individuals' domain rights—the weight individuals' claims against interference in their personal domain should be granted relative to competing moral considerations. I will argue in the rest of this article that there is a plau-

43 Timmons, *Moral Theory*, 144–45, 171.

44 See, for example, Nozick, *Anarchy, State, and Utopia*, 32–33.

45 For Nozick's endorsement of this value, see Nozick, *Anarchy, State, and Utopia*, 32–33. For Fabre's endorsement of this value see, Fabre, *Whose Body Is It Anyway?* 112–13.

sible set of positions in these debates (different from the positions endorsed by Nozick and Fabre) that can offer a compelling solution to the body-income puzzle.

I call the conception of RSP that I will defend the *adroit conception*. "Adroit" is an acronym for "ample domain rights of inconstant tenacity." As the name suggests, this conception of RSP views each individual's personal domain as having *extensive* (i.e., "ample") *scope*. And it views domain rights as having a *variable strength* (i.e., "inconstant tenacity"). More specifically, the adroit conception endorses the following six commitments:

1. Each individual's personal domain extends to her entire body.
2. Each individual's personal domain extends to her labor income.[46]
3. The strength of domain rights ranges on a spectrum from very weak to very strong.
4. A key determinant of a domain right's strength is the magnitude of the personal violation associated with the right's infringement (i.e., the importance of the unwelcome consequences the right's infringement would have for the right-holder's personal identity).
5. At their weakest, an individual's domain rights can be permissibly infringed for the sake of satisfying even one individual's moderate claims to assistance.
6. At their strongest, an individual's domain rights have a strength that *far* outweighs the strength of another individual's claim to lifesaving assistance.

I leave the task of defending these commitments for section 6. My aim in this section is to argue that, *if* this conception of RSP is the right one, it can compellingly solve the body-income puzzle.

Before turning to this argument, a few caveats and clarifications are needed. First, as defined above, the adroit conception is admittedly not fully specified. A full specification would require a complete account of personal identity and a precise account of how domain-right strength varies with the magnitude of personal violations. I will not provide this type of full specification here. Instead, I will take Radin's (admittedly less-than-fully specified) conception of personal identity as given. I will also focus only on those comparisons of personal violations that are central to the body-income puzzle. And I will only argue that there is *some* plausible specification of the relationship between personal violations and the strength of domain rights (consistent with the six commitments above) that would enable the adroit conception to solve the body-income puzzle.

46 This is true under certain conditions that I will specify in section 6.2 below.

Second, I will assume that the adroit conception of RSP is one principle in a *pluralist* theory of distributive justice that is also committed to fostering the collective good. And though I do not specify the relevant collective-good principles (e.g., whether they are aggregative, egalitarian, or sufficientarian), I will assume that these principles grant Poor and Sick *strong pro tanto* claims to assistance. With these clarifications in mind, I now turn to arguing that the adroit conception can indeed meet all five challenges posed by the body-income puzzle.

First, it can explain why taking Healthy's kidney is more objectionable than taking Rich's income. Like Radin's solution, it appeals to the graver personal violation associated with the kidney's taking.[47] Given commitment 4, this implies that Healthy's domain rights against the kidney's taking are stronger than Rich's domain rights against the taking of 20 percent of his labor income, and the kidney's taking is thus more objectionable. Admittedly, the claim that taking Healthy's kidney constitutes a graver personal violation than taking Rich's income is insufficiently defended by Radin. I therefore turn now to defending this claim more rigorously.

Consider first the personal violation associated with taking 20 percent of Rich's labor income. This violation is considerably less grave than it might first appear. For one thing, the life projects curtailed by a 20 percent tax are likely to have a circumscribed personal importance to Rich. After all, the tax is limited to a moderate amount of Rich's labor income, leaves Rich with a very substantial after-tax income with which to pursue his life plans, and allows Rich to *choose* which projects to surrender given this generous post-tax budget constraint. The tax is therefore unlikely to force Rich to surrender projects that are central to his personal identity.

The labor-based connections between Rich and the seized income are also weaker than they might initially appear. First (and most obviously), in taking only 20 percent of Rich's labor income, the state takes only a limited amount of the fruits of Rich's labor. Second, the fact that this wealth is not the *immediate* result of Rich's labor weakens Rich's personal connections to it. Compare, for example, a 20 percent tax on Rich's labor income to a proposal to redistribute 20 percent of the physical farming produce that Rich cultivated by hand. In the case of the physical produce, the long-term contact between Rich's body and the produce and the ways in which this produce has been shaped by Rich's choices could plausibly ground special personal attachments—attachments that do not

---

47 Note that, unlike Radin's solution, the adroit conception offers a *principled* explanation for the importance of personal violations by connecting them to respect for the separateness of persons.

exist between Rich and the money he ultimately receives in exchange for this farming produce.

On the other side of the ledger, the taking of a kidney constitutes a personal violation that is far graver than skeptics like Fabre have recognized. First, the physical health risks associated with a kidney's removal (e.g., trauma and infection) are more serious than Fabre acknowledges.[48] Second, there is a marked increase in the risk of mental health problems for those who surrender a kidney to others—a risk that Fabre does not consider.[49]

In addition, Fabre's distinctions between the taking of a kidney and the taking of external body parts are overstated. A person's interactions with the world and the way that she is seen by others are, after all, substantially impacted during the weeks-long postoperative recovery period.[50] And the operation leaves an incision scar that serves as a visible reminder of the kidney's loss to the person and to those who can see that part of her body.

The taking of the kidney also constitutes a particularly grave personal violation because it entails cutting deeply *into Healthy* and removing *a part of her*. (This assumes that the person is constituted by her entire body, a claim I defend in section 6.1.) Moreover, what is taken from Healthy is not a single molecule, but rather an entire organ. And since kidneys do not regenerate, the loss is permanent and is for that reason more severe. In addition, the combination of irreversibility and substantiveness implies that, post-taking, Healthy might come to see herself (and might be characterized by others) as irredeemably "incomplete"—a substantial, unwelcome change to her personal identity. Indeed, in one study of those who sold a kidney, this is the way that many kidney vendors characterized themselves, even long after their kidney's loss.[51]

Another reason for the gravity of the personal violation associated with a kidney's taking is the connection between a person's purposeful choices and the condition of her kidney. Healthy may well have made decisions (e.g., in terms of what to ingest or what kind of activities to engage in) aimed at avoiding damaging her kidneys. Just as purposeful choices can ground personal connections between Rich and his labor income, so too the care that Healthy has taken in safeguarding her kidneys constitutes a significant source of personal attachment to them.

---

48 Wilkinson, "The Confiscation and Sale of Organs," 330–31.

49 In one study, psychiatric disorders went from a pre-donation rate of 2 percent to an alarming 31 percent. Switzer and Dew, "Psychosocial Issues in Living Organ Donation." The mental health problems caused by coerced organ takings may well be even greater.

50 See Johns Hopkins Medicine, "What to Expect as a Living Donor."

51 Moazam, Zaman, and Jafarey, "Conversations with Kidney Vendors in Pakistan," 34–35.

A final reason for the strong personal connection between Healthy and her kidney is the important function that Healthy's kidney plays in the overall flourishing of Healthy's body. Healthy's kidney filters harmful (and potentially lethal) waste from her blood. This makes Healthy's kidney far more personally important than, say, Healthy's appendix, which serves no discernable function in Healthy's body.

An obvious rejoinder is that Healthy's kidney is unimportant for the functioning of Healthy's body because the other kidney can take over the function of the removed kidney. However, this rejoinder underestimates the importance of Healthy's kidney in three ways. First, a second kidney is not completely redundant. One kidney is not quite as good at purifying the blood as two.[52] Second, each kidney plays an important *backup function* in case the other kidney fails. Third (and most importantly), even if each kidney was perfectly reliable and fully capable of doing the purification work of two kidneys, it is a mistake to evaluate the value of a contributing factor to a cooperative enterprise using a "marginal product" approach (i.e., by only considering the consequences of the contributing factor's absence given that all other contributing factors are present).

To see why, consider a team of ten individuals who jointly create a product worth $10 million. The team decides to split the proceeds of the product's sale according to the importance of each team member's contribution, judged by what would have happened had each been absent. As it turns out, though, the consequences of any one team member's absence would have been fairly minor—$100,000 loss from the overall value of the project (since the other team members would have picked up the slack). According to this way of estimating each team member's value, then, the sum of the value of the ten team members is only $1 million (rather than the $10 million value of the project).

As this example suggests (and as economists have long recognized), it is implausible to use a factor's "marginal product" to determine its value to some overall cooperative enterprise because the sum of the marginal products of the cooperating factors will often either over-exhaust or (as in the case above) under-exhaust the total value of the enterprise.[53] It is thus a mistake to think about a kidney's value to the body's flourishing by only evaluating what would happen to the body if that kidney was not there. The kidney is part of a two-kidney team with enormous importance to the body's flourishing. On a variety of ways of estimating the value of contributing factors to a cooperative enterprise that are more plausible than the marginal product approach, Healthy's kidney can be

52   Kidney Research UK, "Living with One Kidney."
53   Young, "Individual Contribution and Just Compensation," 268.

judged to have considerable importance to her bodily flourishing and therefore to her personal identity.[54]

Thus, while taking 20 percent of Rich's labor income may well constitute the greater welfare loss, taking Healthy's kidney can reasonably be seen as the graver personal violation. If so, then the adroit conception would grant Healthy's domain rights against the taking of her kidney greater strength than it grants to Rich's domain rights against the taking of 20 percent of his labor income. The adroit conception can therefore judge the kidney's taking to be more objectionable.

The adroit conception can also protect Healthy's kidney from redistribution, even in the face of Sick's pressing claims to assistance. Remember that the adroit conception grants the strongest individual domain rights a strength that *far* outweighs the strength of even another individual's claims to lifesaving assistance. And while Healthy's domain rights against the taking of her kidney are certainly not the strongest possible, I argued above that they are quite strong. There is thus a plausible specification of the relationship between domain-right strength and personal violations (consistent with commitment 6) that would protect Healthy's kidney from redistribution, even if Sick needs Healthy's kidney to survive.

Yet the adroit conception does not grant the body absolute protection. It can, for example, condone the taking in Drop of Blood to Save a Billion. Remember that the adroit conception holds that the weakest domain rights can be permissibly infringed in the face of even moderate claims to assistance (commitment 5). And while Adam's domain rights against the taking of a single drop of blood are certainly not the weakest possible, they are clearly substantially weaker than Healthy's rights against the taking of her kidney. A drop of blood is, after all, an easily replaced, relatively insignificant part of Adam's body. Moreover (and more importantly), *a billion* individuals' claims to lifesaving assistance are immensely strong in aggregate. The adroit conception could thus straightforwardly permit the taking of one drop of Adam's blood in this case.

The adroit conception can also justify redistributing 20 percent of Rich's labor income to Poor. While Rich's domain rights against the taking of 20 percent of his labor income are certainly not the weakest possible, I argued above that they are relatively weak (at least compared to Healthy's rights to her kidney). Since Poor's disadvantage-based claims to financial assistance are *ex hyptothesi* strong, and since the adroit conception holds that the weakest domain rights can

54 A prominent solution to this problem, the Shapley value (see Young, "Individual Contribution and Just Compensation"), calculates the *average* marginal product of the factors, where the average is taken over all possible orderings of the cooperating factors. For any orderings of body parts in which kidney 2 is not already present, kidney 1 would obviously have a very high value. Thus, a kidney's Shapley value would be quite high, despite the other kidney's capacity to make up for its absence.

be justifiably infringed for the sake of satisfying even moderate claims to assistance, there is a plausible specification of the relationship between domain-right strength and the magnitude of personal violations that would permit redistributing 20 percent of Rich's labor income to Poor.

Finally, the adroit conception can place attractive limits on labor-income taxation. Consider again the heavy two-kidney-based taxes proposed earlier. Unlike the 20 percent redistributive tax imposed on Rich, the heavy two-kidney-based taxes infringe on Healthy's rights to *all or nearly all of her labor income*. Since the loss of this income would require Healthy to surrender nearly all of her key life projects, these taxes constitute a very grave personal violation. And since the adroit conception holds that the strongest individual domain rights can *easily* outweigh even lifesaving claims to assistance, it can justify rejecting heavy two-kidney-based taxation, even if such taxation would save Sick's life. Thus, as part of a pluralist theory of justice that is also committed to the collective good, the adroit conception of RSP can meet all five challenges posed by the body-income puzzle.

6. DEFENDING THE ADROIT CONCEPTION

Some readers may dismiss the adroit conception of respect for the separateness of persons as *ad hoc*—specifically designed to solve the body-income puzzle. Others may argue that, whatever its merits in solving this particular puzzle, it is an implausible conception of RSP. However, I will argue in this section that each of the adroit conception's six commitments are compelling, independently of their contributions to solving the body-income puzzle.

6.1. The Entire Body and the Personal Domain

Consider first the adroit conception's commitment to the inclusion of the entire body in an individual's personal domain. This commitment has considerable *prima facie* appeal. After all, it seems reasonable to hold that respect for the separateness of persons should at least protect *persons* from being sacrificed for the collective good. And the idea that the person is constituted by her entire body has been defended by a variety of scholars.[55] It also has considerable intuitive plausibility. If you were asked to divide the universe into two parts, you and not-you, most readers would, I think, make the cut at the boundary of their body.

However, some might object that "a person" is much better understood not in terms of a physical body but rather in terms of purely mental capacities such as being self-conscious, being aware of one's identity and its continued existence

---

55 Radin, "Property and Personhood," 965–66. See also Jensen and Moran, *The Phenomenology of Embodied Subjectivity*.

over time, or having the capacity for moral and rational agency.[56] On this understanding of "the person," an individual's personal domain might be seen as extending only to what Jeff McMahan calls "the embodied mind" (i.e., only to those parts of the body needed to support the mental activities constitutive of personhood).[57]

Proponents of this conception of the person often defend it by appealing to the elements of a person necessary for remaining the same person over time. They compellingly point out that a person can lose large portions of her body (or, in science fiction thought experiments, even her entire body) while still fundamentally remaining the same person.

However, it is a mistake to define what constitutes "a person" by considering the elements that are necessary to ensure that a particular individual remains the same person over time. This intertemporal identity question inquires after what is *essential* to a particular individual's identity as a specific person. This is different from the question of what constitutes the person. For example, a person whose hand is chopped off is still fundamentally the same person. But the attack nevertheless constitutes an assault *on her person*. Her hand is part of her person, even if it is not *essential* to her identity as the particular person that she is. Thus, even if purely psychological answers to the question of what it takes to remain the same person over time are right, this does not challenge the view that "a person" should be defined in ways that include her nonessential body parts. It therefore does not challenge the adroit conception's inclusion of the entire body in the personal domain.

6.2. Labor Income and the Personal Domain

The adroit conception's inclusion of labor income in the laborer's personal domain also has considerable intuitive appeal. There is, after all, a widespread judgment that individuals' pretax income is (in some morally substantive sense) *theirs*.[58] British Prime Minister David Cameron appealed to this judgment when he told his supporters, "I know it's your money. I know you want some of it back.... But we will only cut taxes once we've made government live within its means."[59] Accepting that individuals' labor income falls within their personal domains can make sense of this widespread judgment.

56   For a discussion, see Fabre, *Whose Body Is It Anyway?* 12.

57   McMahan, *The Ethics of Killing*, 66–94.

58   For another example and a brief discussion of the widespread nature of this judgment, see Murphy and Nagel, *The Myth of Ownership,* 35–36. Murphy and Nagel go on to argue that this judgment is misguided. I consider some of their arguments below.

59   Cameron, "David Cameron's Speech in Full."

The inclusion of labor income in individuals' personal domain also has considerable scholarly support, especially among orthodox libertarians (e.g., Nozick).[60] The libertarian defense of this position is complex, and I cannot rehearse it fully here. But the basic idea is that if RSP protects the body it should also protect the fruits of the body's activity (e.g., labor income).

This traditional libertarian argument can admittedly be challenged by appealing to labor income's reliance on natural and social inputs. I already highlighted several general problems with this challenge in section 2. Here I wish to further respond to this challenge by considering *each* of the distinctions made between body-part ownership and labor-income ownership and arguing that none of them justify excluding labor income from the laborer's personal domain.

Consider first labor income's reliance on natural inputs. If an individual's personal domain does not extend to any part of the natural world, then it may be reasonable to exclude wealth generated using her body and the natural world from her personal domain.[61]

However, wholly excluding rights to the natural world from individuals' personal domains is implausible. To see why, consider the following policy:

> *Air Rights for a Kidney*: Healthy is granted full property rights to her body. However, her rights to breathe air are made conditional on her transferring one of her kidneys to Sick. Faced with this air-property-rights regime, Healthy agrees to transfer one of her kidneys to Sick.

I submit that Air Rights for a Kidney violates RSP. It imposes an intuitively unacceptable sacrifice on Healthy for the sake of the collective good. Yet we cannot explain why this is so if we view each individual's personal domain as extending no further than the boundary of her body. On the other hand, we can straightforwardly explain why Air Rights for a Kidney violates RSP if we view Healthy's rights to the air she needs to breathe as falling within her personal domain.

Determining precisely how far into the natural world each individual's personal domain extends is beyond the scope of this article. However, if rights to some of the world's plentiful air fall within Healthy's personal domain, it seems plausible that rights to some arable land could also fall within Rich's personal domain, at least when this land is plentiful. If Rich's body and farmland fall within his personal domain, I take it to be uncontroversial that the wealth Rich gener-

---

60   Nozick, *Anarchy, State, and Utopia,* 169. See also Wheeler, "Natural Property Rights as Body Rights," 172–79. Although Wheeler does not explicitly appeal to RSP, he argues that individuals have the *same type of rights* to their income as they do to their bodies.

61   For a conception of RSP that limits the personal domain to the body, see Dworkin, "Comment on Narveson," 38–39.

ates through the use of these inputs and no others (i.e., in a state of autarky) also falls within his personal domain.

Yet some may nevertheless advocate excluding Rich's labor income from his personal domain by appealing to this income's reliance on *transactions*. One reason suggested by John Christman's work for doing so is that particular transactions are the result of social facts (e.g., the particular distribution of competitors in the economy) for which the transacting agent is not responsible.[62]

However, note that Rich is not responsible for many of the factors that affect his wealth in a state of autarky (e.g., the weather, the lack of insects in the area). He is also not responsible for many features of his body (e.g., being born with two healthy kidneys). If Rich's body and the wealth he produces in a state of autarky nevertheless fall within his personal domain, it is unclear why Rich's lack of responsibility for, say, the distribution of competitors justifies excluding the fruits of his transactions from his personal domain.

Another objection to including the fruits of transactions in the personal domain (also suggested by Christman's work) is that transactions require a certain type of control over others.[63] And it admittedly seems implausible to include resources that are based on "other-control" in the sphere of moral authority protected by respect for the separateness of persons.

In response, I do not deny that *some* transactions are based on control over others. If Rich enslaved others in order to earn his farming wealth, this wealth would clearly be based on other-control. Transactions based on fraud or exploitation could also plausibly be described as grounded in a certain type of control over others. I freely concede that the fruits of *these* transactions should be excluded from Rich's personal domain.

However, when Rich's transactions are just—non-fraudulent, non-exploitative, and uncoerced on all sides—it is difficult to see how the resulting wealth can be plausibly described as based on control over others. It is thus difficult to see why the dependence of Rich's wealth on *just* transactions would warrant excluding the resulting wealth from Rich's personal domain.

One final objection to labor income's inclusion in the personal domain appeals to its reliance on government-provided goods. As Murphy and Nagel point out, without goods like the monetary system, individuals (e.g., Rich) would not have any income whatsoever. Murphy and Nagel conclude that it is thus *incoherent* to grant Rich any type of rights to his pretax labor income.[64]

However, this argument is unconvincing. To see why, consider a case in

---

62 Christman, "Self-Ownership, Equality, and the Structure of Property Rights," 31.

63 Christman, "Self-Ownership, Equality, and the Structure of Property Rights," 34.

64 Murphy and Nagel, *The Myth of Ownership*, 32.

which Rich needs Johnny's seeds to produce farming output. In this case, Rich's farming produce would not exist without Johnny's seeds. And we may not even know what Rich's income would be in the absence of Johnny's seeds. Yet it does not follow that it is logically incoherent to grant Rich domain rights to a substantial portion of his farming produce. Rich could, for example, offer Johnny a percentage of his ultimate farming produce in exchange for using the seeds. As long as Rich's agreement with Johnny is just, Rich could perfectly coherently have domain rights to his *post-seed-payment* farming produce.

Now consider the monetary system. There is no denying that Rich's income fundamentally depends on this system. However, just as Rich can coherently have domain rights to his post-seed-payment farming produce, so too there is nothing incoherent about granting Rich domain rights to the labor income that remains after Rich has paid *his just share* of the costs of the monetary system (and any other government-provided inputs he has used in generating his wealth).

The debate over individuals' rights to their labor income is admittedly complex, and I do not claim to have definitively settled it here. However, when a person's wealth is the result of the use of her body, her just share of the natural world, and just transactions with others, and when she has justly paid for the government-provided goods she has used, I see no compelling reasons for excluding the remaining labor income from her personal domain.

6.3. The Variable Strength of Domain Rights

Consider next the adroit conception's commitment to the variability of domain-right strength. This position enjoys substantial scholarly support.[65] It is also intuitively attractive. Imagine that, rather than a billion lives being at stake in Drop of Blood to Save a Billion, the number of lives threatened by the epidemic is X. Next, imagine that saving X lives requires removing and studying one of the following parts of Adam's body:

1. One drop of Adam's blood,
2. one of Adam's kidneys, or
3. large parts of the frontal lobe of Adam's brain.

I submit that the number of individuals (X) whose lives must be at stake to make the taking permissible varies with the type of taking. If so, and if Adam's entire body falls within his personal domain, this suggests that domain rights are indeed best understood as having variable strength.

---

65   See, for example, Quinn, "Actions, Intentions, and Consequences," 310. For another defense of the variable strength of rights against bodily interference, see Thomson, *The Realm of Rights*, 154–55.

However, this position is subject to two key challenges. First, some might reject *any* differences in domain-right strength. Samuel Wheeler, for example, argues that rights have no underlying dimension. A right is either infringed or it is not.[66] Thus, in Wheeler's view, if I have rights to some ample food supply that I have justly acquired, and I have a right to my body, then the taking of some of my food supply to feed the starving is morally "on a par" with the taking of some of my flesh to feed the starving.[67]

However, Wheeler's argument is unconvincing. It may well be that right infringements do not admit degrees (i.e., a right is infringed or it is not, it cannot be partially infringed). However, it does not follow that *what it takes to justify* the infringement of a right must be the same for all rights (i.e., that all rights have the same strength). Thus, we cannot conclude from the existence of rights to my food supply and rights to my flesh that their takings are morally on a par. It is perfectly possible (and indeed intuitively plausible) that my rights to my flesh are stronger than my rights to (every portion of) my ample food supply. If so, then certain competing considerations (e.g., others' starvation) might make the taking of some of my ample food supply, but not my flesh, permissible.

A more powerful challenge to the adroit conception's position is posed by Fabre. While the adroit conception views domain-right strength as a spectrum, Fabre endorses a single right-strength threshold. A person's right against "minor" domain interference has no strength while a person's right against "more than minor" domain interference has absolute or near-absolute strength (where the threshold between "minor" and "more than minor" lies at the level of interference that substantially threatens a person's ability to lead a minimally flourishing life).[68]

Fabre defends her position by claiming that a passerby has a ten-second duty of justice to physically help her up if she slips on an icy pavement.[69] This intuitively plausible duty of rescue can be straightforwardly explained by accepting Fabre's claim that individuals' rights against sufficiently minor interference with their personal domain have no strength.

However, while Fabre's example does challenge absolutist conceptions of RSP, it does not pose a serious challenge to the adroit conception. Given the very limited nature of what is required of the passerby, the adroit conception need not view the passerby's domain rights against this duty of rescue as having a great deal of strength. Thus, assuming that Fabre's plight of being sprawled on the icy pavement is sufficiently grave (e.g., she cannot get up herself, it is very

66　Wheeler, "Natural Property Rights as Body Rights," 182.
67　Wheeler, "Natural Property Rights as Body Rights," 184.
68　Fabre, *Whose Body Is It Anyway?* 112–118.
69　Fabre, *Whose Body Is It Anyway?* 112.

cold outside, and there is no one else around who will help her up), the adroit conception could affirm the passerby's ten-second duty of justice to help Fabre up. In this case, then, both Fabre's view of domain-right strength and the adroit conception's view have plausible implications.

However, the greater plausibility of the adroit conception's position becomes apparent when we consider domain rights that lie near each other but on different sides of Fabre's threshold. Consider, for example, the following two cases:

> *Major Sacrifice of One for Many*: Andrea would make for an exquisitely talented surgeon, saving the lives of one hundred people over the course of her career who would otherwise die. However, Andrea dislikes surgery and prefers to be a gardener. The government forces Andrea (on pain of jail) to become a surgeon.

> *Moderate Sacrifice of Many for One*: Betsy is struck by a fatal disease. Finding a cure requires testing millions of different chemical combinations. The government forces all one million of society's eighteen-year-olds (on pain of jail) to spend one year after secondary school testing chemical combinations (or supporting this effort administratively) for minimum wage, with waivers for anyone whose ability to lead a minimally flourishing life would be jeopardized by this requirement.

I submit that *both* Major Sacrifice of One for Many and Moderate Sacrifice of Many for One are impermissible. However, while Fabre can reject Major Sacrifice of One for Many because the sacrifice asked of Andrea would undermine her ability to lead a minimally flourishing life, Fabre's account of domain-right strength commits her to endorsing Moderate Sacrifice of Many for One.[70] After all, Betsy has weighty claims to assistance. And the loss of a year of occupational choice for the one million eighteen-year-olds does not jeopardize their ability to lead a minimally flourishing life.[71]

However, what is most implausible about Fabre's view is not its endorsement of Moderate Sacrifice of Many for One. Rather, it is the implication that this case is not even a moral dilemma. A year of occupational choice for *one million individuals*—an opportunity for them to take important, often foundational steps

---

70 Fabre, *Whose Body Is It Anyway?* 55–56. This is so at least as long as no one else's capacity to lead a minimally flourishing life can be fostered by the million 18-year-olds' sacrifice.

71 Fabre endorses a mandatory year of civilian service for the sake of the disadvantaged. See Fabre, *Whose Body Is It Anyway?* ch. 3. Note that Moderate Sacrifice of Many for One is substantially more intuitively unappealing because the relevant sacrifice is made for the sake of only one person (though this does not change the implications of Fabre's theory).

toward fulfillment of key life projects—is granted no protection at all by Fabre's conception of RSP. This seems very difficult to accept.

6.4. Personal Violations and Domain-Right Strength

Consider next the adroit conception's view of the magnitude of personal violation as the key determinant of domain-right strength. This position has considerable scholarly support.[72] It also has substantial theoretical appeal. Respect for the separateness of persons is, after all, fundamentally concerned with safeguarding each person's *individuality* in the face of collectivist demands. It therefore seems reasonable to see the severity of the violations of this value as dependent on how close coerced sacrifices come to the core of who the individual is.

This position can also make sense of our judgments regarding the relative objectionability of different types of domain interference. As Quinn argues, it is because of the fundamental connections between our minds and who we are that our rights against being lobotomized for the sake of the collective good are far stronger than our rights against other types of bodily encroachments.[73]

Consider also the case of Hairy-Bald introduced in section 2. I suggested that the objectionability of redistributing Hairy's detached hair is much more akin to the objectionability of redistributing Rich's labor income than to the objectionability of redistributing Healthy's kidney. And the adroit conception can straightforwardly explain why this is so. Hairy admittedly has somewhat greater personal connections to his detached hair than Rich has to his labor income because the hair used to be a part of Hairy. However, taking Hairy's detached hair does not impose health risks on Hairy, does not require cutting deep into Hairy, and does not require taking any part of Hairy (let alone a substantial, functional part). The adroit conception can thus affirm that taking 20 percent of Hairy's detached hair is only modestly more objectionable than taking 20 percent Rich's labor income.

However, the adroit conception's view of the determinant of domain-right strength is not the only plausible one. Fabre holds that the strength of a domain right is based on the effects its infringement would have on the right-holder's ability to lead a flourishing life—a complex capacity that depends on a variety of factors, including one's self-respect; ability to satisfy current plans, projects, goals, and interests; and ability to reflect and choose one's goals and actions.[74] Like the adroit conception's position, Fabre's view can explain the robustness of

---

72   Quinn, "Actions, Intentions, and Consequences," 311.

73   Quinn, "Actions, Intentions, and Consequences," 311.

74   Fabre, *Whose Body Is It Anyway?* 28–29, 113–18.

our rights against being lobotomized and the similarity of Hairy's detached hair to Rich's labor income in terms of redistribution objectionability.

However, Fabre's view has intuitively implausible implications in other cases. Consider, for example, the following policy:

> *Secret Sexual Access*: The government secretly grants the *sexually deprived*—those unable to find willing sexual partners, even for money—access to the bodies of attractive individuals who have been anesthetized for medical operations. Foolproof protection from disease and pregnancy are used as well as cleaning measures that ensure that the consequences of the sexual activity will be unnoticed by the patients. The patients' faces are also covered so that their identity is kept secret from the sexually deprived.

Though Fabre recognizes that the sexually deprived have weighty claims to assistance, and though she generally demands a great deal from the advantaged, even she balks at demanding that individuals' bodies be made available to the sexually deprived.[75] However, none of the negative effects on the ability to lead a flourishing life that Fabre appeals to in order to reject the sexually deprived's demands for access to others' bodies are relevant in Secret Sexual Access. Secret Sexual Access poses no physical or psychological risks for the patients nor has any effect on their subjective welfare. This policy would also have no consequences for how the patients see their bodies or sex. Fabre's view of the determinant of domain-right strength thus implies that this policy is, at worst, mildly objectionable.[76] Yet I take it that Secret Sexual Access is in fact highly objectionable.

This objectionability can be easily explained by the adroit conception. As many theorists have compellingly argued, our sexual integrity is intimately connected with our personal identity.[77] Given that the adroit conception endorses a nonsubjective account of personal identity, what is done to the patients in Secret Sexual Access can straightforwardly be classified as a very grave personal violation and can thus be judged by the adroit conception to be highly objectionable, despite the benefits for the sexually deprived.

---

75 For a discussion of why forced sexual acts would lead to an unacceptably large reduction in the ability to lead a minimally flourishing life, see Fabre, *Whose Body Is It Anyway?* 118–23.

76 Fabre could appeal to individuals' non-experientialist interest in physical integrity. Yet elsewhere, Fabre argues that such non-experientialist interests do not have a great deal of importance for our ability to lead a flourishing life. See Fabre, *Whose Body Is It Anyway?* 87–88.

77 See Archard, "The Wrong of Rape."

6.5. The Weakest Domain Rights and Moderate Claims to Assistance

Consider next the adroit conception's endorsement of the permissibility of infringing on the weakest domain rights for the sake of meeting others' moderate claims to assistance. To see this commitment's intuitive appeal, assume it is possible to cure Poor's serious, permanent limp by taking *a single cent* of Rich's $100,000 income. Assume also that the redistribution process is costless for all involved. I submit that taking a single cent from Rich is permissible in this case. If so, this suggests that very weak domain rights can indeed be permissibly infringed for the sake of satisfying even moderate claims to assistance.

One challenge to this position could be pressed by orthodox libertarians, who sometimes insist that domain rights are absolute.[78] However, the problems with this orthodox libertarian view are by now familiar. Insisting on the absoluteness of domain rights would not only require the state to forbear from taking a single cent of Rich's income to cure Poor's limp. It would also require near-complete paralysis on the part of humanity.[79] After all, nearly all actions taken in a social setting (e.g., walking to the store) pose *some risk* of infringing on someone else's body. Thus, insisting on absolute domain rights would imply that almost no action taken in a social setting is permissible, regardless of its benefits. Given this patently implausible implication, it is perhaps unsurprising that, despite occasional insistence on domain-right absoluteness, it is difficult to find any libertarian who consistently endorses such a radical position.[80]

A more powerful challenge to the adroit conception's position is posed by Eric Mack. Mack does not deny that sufficiently weak domain rights can sometimes be infringed for the sake of alleviating severe disadvantage. But he denies that the right-holder is ever under an obligation to facilitate the infringement of her domain rights.[81] Thus, even if Poor can permissibly seize a cent from Rich to cure her limp, Mack denies that Rich is obligated to facilitate this taking. If Mack is right, then the adroit conception's position on the infrangibility of domain rights would be insufficient to justify income taxation (which coercively demands taxpayer cooperation) and would thus be insufficient to solve the body-income puzzle.[82]

Mack's defense of his position is complex, and I cannot rehearse it fully here.

---

78   Rothbard, *The Ethics of Liberty*, 60.

79   Mack, "Nozickian Arguments for the More-than-Minimal State," 112–13.

80   For example, Rothbard concedes that air pollution is permissible as long as any damages caused are paid for ("Law, Property Rights, and Air Pollution," 77). Yet *absolute* rights to one's body would imply that explicit consent from those affected is required.

81   Mack, "Non-Absolute Rights and Libertarian Taxation," 120.

82   There is a justification for a kind of tax system that can follow from Mack's premises, though

But his basic point is that a theory that grants others a right to take Rich's income *and also* imposes an obligation on Rich to turn over his income to others makes Rich's rights to the redistributed income effectively meaningless.[83] Mack also suggests that such a theory would impose demands on those in a position to assist others that are intuitively overdemanding.[84]

However, Mack's claim that obligations to facilitate the infringement of one's domain rights are overdemanding is unconvincing. To see why, it is first important to clarify a potential point of confusion about the demands the adroit conception places on those in a position to assist others. Poor's serious limp being sufficiently serious to justify infringing on a cent of Rich's income does not commit the adroit conception to endorsing Rich's coercively enforceable obligation to transfer a cent to Poor. To justify such an obligation, Poor's plight would have to be sufficiently serious to justify infringing on Rich's right to the cent *and* all of the domain rights associated with Rich's facilitation of the cent's transfer. So, if Rich must physically deliver the cent to Poor at a faraway location, the adroit conception could well affirm, in line with Mack, that Rich is not obligated to facilitate the cent's transfer.

However, consider now a case in which Rich can transfer the cent with a few keyboard strokes and the cent is needed to save the lives of one thousand of Rich's compatriots. I submit that coercively requiring Rich to facilitate the infringement of his domain rights to the cent is not overdemanding in this case. Yet, while the adroit conception can straightforwardly endorse this faciliation obligation, Mack is committed to denying it.

Mack's claim that obligations to facilitate the infringement of one's domain rights renders these rights meaningless is also unconvincing. For one thing, accepting that taxpayers have (infringed) rights to the wealth they are obligated to transfer to the state affects our proper *attitude* toward taxation. It implies that those who surrender their labor income for the sake of the indigent properly merit appreciation for their sacrifice in ways that, say, individuals who are paying back a government loan do not.

Recognizing taxpayers' rights to the wealth they transfer to the state can also

---

the argument for it is complicated. See Mack, "Non-Absolute Rights and Libertarian Taxation," 138–41.

83   Mack, "Non-Absolute Rights and Libertarian Taxation," 121–22.

84   Mack appeals to the judgment that it would be overdemanding to coercively require a cabin owner to communicate the location of the cabin's hidden key to a faultlessly freezing hiker who will die without access to the cabin ("Non-Absolute Rights and Libertarian Taxation," 120–21). Though I do not share this judgment, I concede that the rejection of the cabin owner's obligation to facilitate the infringement of his domain rights is plausible in this particular case.

have practical distributive consequences. Imagine, for example, that the state suddenly obtains a large windfall of wealth. Recognizing taxpayers' *pro tanto* rights to the wealth they have previously transferred to the state would imply that there is a case not only for using this windfall to reduce future taxes but also for using some of it to compensate past taxpayers for previous right-infringements. A theory that did not recognize taxpayers' infringed rights to their previously transferred wealth would not have this implication.

Third (and most importantly), recognizing taxpayers' rights to the wealth they turn over to the state has key implications for the *strength* of their claims against taxation. A theory that did not recognize Rich's rights to *his* pre-redistributive-tax income would have great difficulty explaining why we should give special weight to the income's importance for Rich's personal projects when deciding how much of Rich's wealth to take.

Thus, there is nothing implausible about accepting that the advantaged sometimes have coercively enforceable obligations to facilitate the infringement of their own domain rights (e.g., by transferring some of their labor income to the state). If so, then accepting the infrangibility of weak domain rights for the sake of meeting moderate claims of assistance could justify substantial levels of labor-income taxation.

6.6. The Strongest Individual Domain Rights and Lifesaving Assistance

Finally, consider the adroit conception's commitment to viewing the strongest of an individual's domain rights as *far* weightier than even another person's claims to lifesaving assistance. This position's intuitive appeal can be illustrated by the example I used to introduce RSP—the prohibition on killing one person to use her organs to save five others. However, for readers unmoved by this well-known case, let me offer another example from the social sphere:

> *Suicidal's Marriage Proposal*: A woman is offered marriage. Although she likes the suitor and enjoys their physical relationship, she takes marriage seriously and does not wish to be attached to him in this way. However, the woman has recently learned that the suitor will commit suicide (due to mental illness) unless she marries him.

Despite this being a matter of life and death for the suitor, I submit that the woman does not act unjustly by rejecting his proposal. This can be straightforwardly explained by accepting that the strongest of an individual's domain rights can easily outweigh even others' claims to lifesaving assistance.

Some readers might object that I have mischaracterized this moral situation. They might argue that, though justice *does* require the woman to marry the

suicidal suitor, the woman has a personal prerogative (grounded in the special weight she is entitled to give her own plans and commitments) to ignore the requirements of justice when they ask too much of her in certain ways.[85] Appealing to personal prerogatives admittedly makes it possible to give domain rights little or even no weight while also rejecting the woman's obligation to marry the suicidal.

However, as Mack compellingly argues, appeals to prerogatives cannot plausibly obviate the need for accepting RSP as a principle of justice.[86] If the only reason the woman lacks an obligation to marry the suitor is her personal prerogative to ignore the requirements of justice, then it would be permissible for an omniscient, disinterested third party to coerce the woman into marrying the suitor. After all, this third party would have no reason to favor the woman's personal commitments and interest in controlling whom she marries over the suitor's life. Yet I take it that this coercion would be patently impermissible. If so, then we cannot plausibly appeal to personal prerogatives to avoid endorsing (as part of our theory of justice) a commitment to RSP that protects the most intimate and personally important parts of a person's life from others' claims to assistance, even when another person's life is on the line.

7. A SOLUTION IN ALL CASES?

I have argued in the previous two sections that the adroit conception of respect for the separateness of persons (as part of a pluralist theory of distributive justice also concerned with fostering the collective good) is both compelling and capable of solving the body-income puzzle. However, this solution is admittedly subject to an important objection that I wish to consider in this final section. Namely, it does not seem to work in all cases.

To see the problem, consider the following variations of Rich-Poor and Sick-Healthy:

Income for Religious Pilgrimage: Rich wishes to go on a pilgrimage required by his religion. However, the pilgrimage is sufficiently expensive so that a 20 percent redistributive tax would force Rich to abandon this critical life project.[87]

Kidney-Cell Teleportation: Sick can grow a new kidney as long as a *single*

---

85 Scheffler, *Rejection of Consequentialism*. See also Fabre, *Whose Body Is It Anyway?* 34.

86 Mack, "Prerogatives, Restrictions, and Rights."

87 For a real-world example, see Wajihuddin, "Pilgrims Protest 9% Service Tax on Haj, Likely to Move Supreme Court."

cell from Healthy's kidney is implanted in his body. A technology is available that can painlessly and noninvasively teleport a single kidney cell from Healthy's body to Sick's.

In this pair of cases, the adroit conception seems committed to offering greater protections to Rich's labor income than to Healthy's body part because the income's taking seems to constitute the greater personal violation. After all, the income's taking deprives Rich of the ability to pursue a project that is central to his life plans and self-conception. The taking of Healthy's kidney cell, on the other hand, is painless, riskless, and noninvasive and takes a miniscule, inconsequential, and quickly regenerating part of Healthy's body. This pair of cases thus demonstrates that the adroit conception cannot in fact solve the body-income puzzle laid out in section 1.

In response, I freely concede that the adroit conception would grant greater protection to Rich's labor income than to Healthy's body part in this pair of cases. However, I deny that this implies that the adroit conception fails to solve the body-income puzzle. The body-income puzzle (on my understanding of it) does not require granting nonessential body parts greater protection than labor income *in all possible worlds*. Rather, it implicitly asks us to explain the greater protection that should be granted to body parts in the typical circumstances that confront policymakers. Income for Religious Pilgrimage and Kidney-Cell Teleportation are highly atypical cases in which the ordinary ordering of the protections offered to labor income and body parts should, I submit, be reversed. The adroit conception's capacity to explain why this is so is thus a strength rather than a weakness.

8. CONCLUSION

I began this article with a seemingly simple question. Why should society reject the redistribution of nonessential body parts while condoning the redistribution of labor income? I argued that answers to this question that reject individuals' ownership of labor income, that appeal to the greater welfare loss associated with bodily takings, and that appeal to general personal-identity-based entitlements to resources are unconvincing.

I defended instead a solution to this puzzle grounded in a novel conception of respect for the separateness of persons. This conception does not deny that individuals have rights to their labor income. Instead, it recognizes that seizing resources that lie closer to the core of who we are entails a greater disrespect for our separateness as persons. Since bodily resources such as our kidneys are

substantial, non-regenerating parts of us—parts that we have cared for and that play an important role in our bodies' functioning—their taking comes closer to the core of who we are than does the taking of a moderate portion of our labor income, at least when we are left economically well-off. Respect for the separateness of persons, rightly understood, can therefore explain why the weighty claims of the disadvantaged to assistance can justify the redistribution of considerable amounts of labor income but are generally insufficiently strong to justify the redistribution of substantial portions of the human body.[88]

London School of Economics
j.m.mazor@lse.ac.uk

REFERENCES

Archard, David. "The Wrong of Rape." *Philosophical Quarterly* 57, no. 228 (July 2007): 374–93.
Becker, Gary, and Julio Jorge Elias. "Introducing Incentives in the Market for Live and Cadaveric Organ Donations." *Journal of Economic Perspectives* 21, no. 3 (Summer 2007): 3–24.
Cameron, David. "David Cameron's Speech in Full." *The Guardian*, October 1, 2008. https://www.theguardian.com/politics/2008/oct/01/davidcameron .toryconference1.
Christman, John. "Self-Ownership, Equality, and the Structure of Property Rights." *Political Theory* 19, no. 1 (February 1991): 28–46.
Cohen, G.A. "On the Currency of Egalitarian Justice." *Ethics* 99, no. 4 (July 1989): 906–44.
Dworkin, Ronald. "Comment on Narveson: In Defense of Equality." *Social Philosophy and Policy* 1, no. 1 (Autumn 1983): 24–40.
Eyal, Nir. "Is the Body Special?" Review of *Whose Body Is It Anyway? Justice and the Integrity of the Person*, by Cécile Fabre. *Utilitas* 21, no. 2 (June 2009): 233–45.
Fabre, Cécile. *Whose Body Is It Anyway? Justice and the Integrity of the Person.* Oxford: Oxford University Press, 2006.

Jensen, Rasmus, and Dermot Moran, eds. *The Phenomenology of Embodied Subjectivity*, vol. 71. New York: Springer, 2013.

Johns Hopkins Medicine. "What to Expect as a Living Donor." http://www.hopkinsmedicine.org/transplant/living_donors/expect.html.

Kidney Research uk. "Living with One Kidney." https://kidneyresearchuk.org/kidney-health-information/questions-about-kidney-disease/living-with-one-kidney/.

Lippert-Rasmussen, Kasper. "Against Self-Ownership: There Are No Fact-Insensitive Ownership Rights over One's Body." *Philosophy and Public Affairs* 36, no. 1 (Winter 2008): 86–118.

Mack, Eric. "Non-Absolute Rights and Libertarian Taxation." *Social Philosophy and Policy* 23, no. 2 (July 2006): 109–41.

———. "Nozickian Arguments for the More-than-Minimal State." In *The Cambridge Companion to Nozick's Anarchy, State, and Utopia*, edited by Ralf M. Bader and John Meadowcroft, 89–115. Cambridge: Cambridge University Press, 2011.

———. "Prerogatives, Restrictions, and Rights." *Social Philosophy and Policy* 22, no. 1 (January 2005): 357–93.

McMahan, Jeff. *The Ethics of Killing: Problems at the Margins of Life*. Oxford: Oxford University Press, 2002.

Moazam, Farhat, Riffat Moazam Zaman, and Aamir Jafarey. "Conversations with Kidney Vendors in Pakistan." *Hastings Center Report* 39, no. 3 (May–June 2009): 29–44.

Murphy, Liam, and Thomas Nagel. *The Myth of Ownership: Taxes and Justice*. Oxford: Oxford University Press, 2002.

National Kidney Foundation "A 'New Normal': Life on Dialysis—the First 90 Days." New York: National Kidney Foundation, 2007.

Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.

Olson, Eric T. "Personal Identity." *Stanford Encyclopedia of Philosophy* (Summer 2017). https://plato.stanford.edu/archives/sum2017/entries/identity-personal.

Otsuka, Michael. *Libertarianism without Inequality*. New York: Oxford University Press, 2003.

Quinn, Warren. "Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing." *Philosophical Review* 98, no. 3 (July 1989): 287–312.

Radin, Margaret Jane. "Property and Personhood." *Stanford Law Review* 34, no. 5 (May 1982): 957–1015.

Rothbard, Murray. *The Ethics of Liberty*. New York: New York University Press, 2002.

———. "Law, Property Rights, and Air Pollution." *Cato Journal* 2, no. 1 (Spring 1982): 55–99.

Scheffler, Samuel. *Rejection of Consequentialism*. Rev. ed. Oxford: Oxford University Press, 1994.

Switzer, Galen, and Mary Dew. "Psychosocial Issues in Living Organ Donation." In *The Transplant Patient: Biological, Psychiatric and Ethical Issues in Organ Transplantation*, edited by Paula Trzepacz and Andrea DiMartini, 42–66. Cambridge: Cambridge University Press, 2000.

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

Tideman, Nicolaus, and Peter Vallentyne. "Left-Libertarianism and Global Justice." In *Human Rights in Philosophy and Practice*, edited by Burton Leiser and Tom Campbell, 443–58. New York: Routledge, 2011.

Timmons, Mark. *Moral Theory: An Introduction*. Plymouth, UK: Rowman and Littlefield Publishers, 2012.

Wajihuddin, Mohammed. "Pilgrims Protest 9% Service Tax on Haj, Likely to Move Supreme Court." *Times of India*, February 27, 2017.

Wheeler, Samuel. "Natural Property Rights as Body Rights." *Noûs* 14, no. 2 (May 1980): 171–93.

Wilkinson, T. M. "The Confiscation and Sale of Organs." *Res Publica* 13, no. 3 (September 2007): 327–37.

Young, H. Peyton. "Individual Contribution and Just Compensation." In *The Shapley Value: Essays in Honor of Lloyd S. Shapley*, edited by Alvin E. Roth, 267–78. New York: Cambridge University Press, 1988.

Zwolinski, Matt. "The Separateness of Persons and Liberal Theory." *Journal of Value Inquiry* 42, no. 2 (June 2008): 147–65.

ON KEEPING THINGS IN PROPORTION

Adam Lovett and Stefan Riedener

JOANNA IS THE VICTIM of a natural disaster. A drought has caused the harvest to fail. She does not have enough food. You donate one hundred dollars to the Red Cross and that gets her the food she desperately needs. This is a good thing, and for a while it seems you have done something very important. You have saved a life! But then you get some perspective. You realize that Joanna is not the only victim of this natural disaster. The drought has struck her entire community, even her entire nation. Millions of people are at risk of starvation. You have saved a tiny proportion of them. Now the significance of your action seems diminished. It seems to matter somewhat less. We will call this the *Drop in the Ocean Intuition*. The intuition is that the fact that your action makes a tiny dent in the overall extent of suffering diminishes its importance. Your contribution matters less when it is a drop in the ocean.

Some philosophers think this intuition is hopeless. Jonathan Glover calls it a "context illusion." He claims that "it can be dismissed at once."[1] Peter Unger calls it "futility thinking." He says that "when [gripped by this intuition] you're greatly influenced by a consideration which is morally irrelevant."[2] Peter Singer holds that the intuition "doesn't stand up to examination."[3] William MacAskill says it "doesn't make any sense."[4] But we are more sanguine. We want to explore a view that validates some form of this intuition. We think that only when we have such a view on the table can we assess its plausibility.

The view we want to explore is simple. It says that a thing's importance depends on how big a proportional contribution it makes to the world's total value. For example, consider the addition of one happy life. According to this view, the importance of this addition depends on what is going on elsewhere in the

---

1 Glover and Scott-Taggart, "It Makes No Difference Whether or Not I Do It," 172.

2 Unger, *Living High and Letting Die*, 63.

3 Singer, *The Life You Can Save*, 59. In the passage this comes from Singer is referring to the intuition that we should favor saving people we see over people we do not see. But he then says "the same is true" for the Drop in the Ocean Intuition.

4 MacAskill, *Doing Good Better*, 30.

universe. If the universe contains just one happy life, then the addition involves a great proportional improvement. So it is very important. But if instead the universe contains a hundred billion happy lives, then the addition is a tiny proportional improvement. So it is not very important. The number of happy lives in the universe affects the importance of the addition of one more happy life. We will call this view Proportionalism.[5]

Proportionalism implies that the relative insignificance of saving Joanna does indeed diminish its importance. But it does not validate everything involved with the Drop in the Ocean Intuition. Some people, for instance, think that if saving Joanna is just a drop in the ocean, then we may as well indulge ourselves instead. Go buy yourself priority boarding! But this need not follow. Your reason to save Joanna may be diminished by its proportional insignificance. But so is your reason to indulge yourself. And they are diminished in proportion to their original weight. So, the relative dominance of the former is untouched. Other people misapply proportional thinking. They take the importance of something to be the proportional contribution it makes to a specific issue. For instance, suppose two populations are suffering famine. Some people prefer saving 75 percent of the smaller population to saving 25 percent of the much larger one, even if this saves fewer lives overall.[6] According to Proportionalism, these people are making a mistake. It is not the size of the specific problem that matters. It is the size of the whole world. These people are using the wrong reference class. Proportionalism, then, is not meant to validate everything involved with the Drop in the Ocean Intuition. It is only meant to validate its core.

In the rest of the paper we explore this view. In section 2, we formulate Proportionalism precisely. In sections 3–5, we look at its further motivation. In section 6, we see how to extend it. In these later sections we also show how versatile Proportionalism is. We will see an agent-relative version of the view (section 5.1). We will see versions of it restricted to certain kinds of facts (section 5.4)

5   Nick Bostrom talks about this view in his "Infinite Ethics." He says: "We might constitute a minute portion of the whole, but that does not detract from our absolute importance" ("Infinite Ethics," 10). In other words, he dismisses Proportionalism out of hand. We have come across no other prior discussion of Proportionalism.

6   Psychologists call this phenomenon "proportion dominance" (Bartels, "Proportion Dominance"), "psychological numbing" (Fetherstonhaugh et al., "Insensitivity to the Value of Human Life"), or the "reference group effect" (Loewenstein and Jenni, "Explaining the 'Identifiable Victim Effect'"). They have studied the phenomenon extensively. In one study, Daniel Bartels asked subjects to choose between *either* a program that saved 225 lives of 230 at risk, *or* one that saved 230 lives of 920 at risk ("Proportion Dominance"). Almost half chose the former. They took the greater proportion of the at-risk population saved to be of greater moral weight than the extra lives. We think it is a strength of Proportionalism that it can help explain why people make this mistake.

and types of normativity (section 5.6). And we will see how it fits into a much broader phenomenon: proportional thinking (section 6.1). But let us not get ahead of ourselves. We will first locate the notion of importance that matters to Proportionalism.

1. CHARACTERIZING IMPORTANCE

The Drop in the Ocean Intuition involves importance or significance. So we need to give an account of how important or significant something is. Our notion of importance is not the axiological notion. For something to be important, in our sense, is not for it to be worth promoting. Our notion is instead understood in terms of fitting attitudes. How important something is is how much it is fitting to care about it. What is it for an attitude to be fitting? For us, it is just for it to be permissible to have that attitude. So, we characterize our notion of importance with the following principle:

> *Caring*: How important something is just is the most one is permitted to care about it.

The relevant type of permissibility is that which rules attitudes of caring. This is not the same as moral, prudential, or epistemic permissibility. It is a distinctive normativity that governs how much we care about things. We think there is such a type of normativity because, intuitively, one's attitudes of caring can be mistaken in distinctive ways. We all know people who care more than anything about long-gone sporting victories. Their attitudes are out of whack; they are making a mistake. But this need not be a moral or prudential mistake. Still less must it be an epistemic mistake. Rather, it is a mistake parochial to attitudes of caring. So, there seems to be a distinctive normativity that governs such attitudes.[7] Of course, there may still be interesting connections between this type of normativity and other types of normativity. For example, we are inclined to think that one has moral reason to care fittingly about things that matter morally. If you care about football more than famines, then your attitudes are not just unfitting: they are morally defective. But there remains a distinction between moral permissibility and that which governs attitudes of caring. It is this second type of normativity that matters for our notion of importance.

7   We admit that there are hard questions about how to demarcate the reasons relevant to this kind of normativity. But these questions do not relate to the phenomenon we are concerned with. So we will put them aside for the rest of the paper. For these questions see, e.g., D'Arms and Jacobson, "The Moralistic Fallacy"; Rabinowicz and Rønnow-Rasmussen, "The Strike of the Demon"; Bykvist, "No Good Fit."

What is it to care about something? For us, "caring" refers to any emotional response. Adoration and adulation are types of caring. But so too are abhorrence and apprehension: you can care about something by hating it just as much as by loving it. Such attitudes have a phenomenological character. There is a certain way it feels to care about something. It seems plausible that they also have an evaluative character. Perhaps caring about something involves valuing (or disvaluing) it. And maybe caring is constitutively connected to actions. Perhaps adoring something involves being disposed to promote or preserve it. Perhaps despising something involves being disposed to prevent or destroy it. We take no stand on this final issue. We are neutral about whether there are such connections between caring and behavior. The account of importance we give would be tenable were there such connections. But it would still be interesting without them. The key thing is that, when we talk about caring about something, we are talking about whatever is involved in emotionally responding to that thing.

Caring defines importance in terms of how *much* one can care about something. So what is it to care some amount about something? Here we have a choice. The first option conceives of caring as an absolute quantity. How much one cares about something is how intensely and for how long one cares about it. The second option conceives of caring as a share. How much one cares about something is what share of one's emotional life one devotes to it. The two come apart when people have different emotional capacities. Consider a person who is incapable of caring a large quantity about anything. They might nonetheless devote their entire emotional life to their family. Then they will not care much about their family in the first sense, but will in the second. We are neutral between these options. We think Proportionalism is tenable on either option.

Caring also does not settle *who* is doing the relevant caring. Here we have a second choice. One option is to relativize importance to each individual. On this interpretation, how important something is *for Adam* is how much it is permissible *for Adam* to care about that thing. We have a multitude of importance rankings, one for each person. But one can also (or instead) define a non-relativized notion of importance. There are a couple of ways to do this. First, one might define such a notion out of the relativized notions. This is easiest if we assume everyone's importance ranking is the same. That is, everyone is permitted to care to the same degree about everything. Then one can say that how important something is (*simpliciter*) is how much it is permissible for *anyone* to care about that thing. Second, one could invoke the emotional life of some ideal observer. So how important something is is how much it is permissible for a fully

informed, imaginative, and impartial observer to care about it.[8] Again, we think Proportionalism is tenable on any of these options. For now, what matters is that we cash out our notion of importance in terms of maximum permissible degrees of caring. How important something is is the maximum emotional response it is fitting or permissible to have toward that thing.

This is not the only notion of importance. We have mentioned the axiological notion already. But, more saliently, there is also a notion of importance involving minimum permissible degrees of caring. On this notion, how important something is is the minimum one ought to care about it. We think the maximum-permissible notion is more interesting. This is because we doubt there is often a minimum degree to which one ought to care about things. We will explain why in section 5.3. But, for now, let us just note that very little we say depends on this. Almost everything we say could be changed to fit the minimum-permissible notion.[9] So, we go forward with the maximum-permissible notion. This paper is about the most one is permitted to care about things.

2. PROPORTIONALISM

We now express Proportionalism a bit more precisely. Let us first characterize a notion of *normative mass*. To do this we assume that some kinds of intrinsic properties invite an emotional response. These are the properties that intrinsically matter. How much normative mass something has is how much of these properties it contains. For example, one might think that joy and suffering were all that intrinsically mattered. That is, one might think that these are the only kinds of intrinsic properties you can permissibly care about. Then the normative mass of something is just how much joy it contains and how much suffering it contains. Compare a long marriage to a short fling. The former might contain a lot of joy and a lot of suffering. So it might have relatively great normative mass. The fling hopefully contains some joy but perhaps also some suffering. Yet probably much less of both. So it has relatively little normative mass. Of course, joy and suffering are likely not the only things that matter intrinsically. Maybe beauty matters. Maybe fearsomeness matters. Maybe it matters that you violated someone's rights. Then these too will contribute to normative mass. But joy and suffering are likely contributors to normative mass.

What kind of things have normative mass? Not much swings on this. But we will treat facts as the bearers of normative mass. We think of facts as proper-

---

8   There is a venerable tradition that invokes such observers in ethics. See, e.g., Firth, "Ethical Absolutism and the Ideal Observer"; Railton, "Moral Realism"; Smith, *The Moral Problem*.

9   We note the only substantial change in note 16, below.

ty instantiations and conjunctions thereof. Your marriage, for example, can be thought of as the conjunction of some of the facts involving you and your partner. We will say the normative mass of a fact is how weighty are the properties that make it up. A weighty marriage may be one that is constituted by much joy and suffering. One important bearer of normative mass is the entire universe. We will treat this as the conjunction of all the property instantiations. The normative mass of the entire universe is how much of what matters there is anywhere. So, if there is a lot of joy and suffering, we have a heavyweight universe. If there is not much of either, our universe may be lightweight. The key thing is that the normative mass of a fact is how much of the properties that intrinsically matter it contains.

We can now state our view. Let $I(f)$ be how important f is. Let $NM(f)$ be f's normative mass. Let $NM(U)$ be the normative mass of the entire universe. Then *Proportionalism* says:

$$I(f) = \frac{NM(f)}{NM(U)}.$$

In other words, how important f is is determined by f's share of the normative mass of the whole universe. Remember you can interpret $I(f)$ either as an absolute quantity of caring or as a share of your emotional life.[10] So this says how much you can care about something depends on the proportion of the universe's total normative mass it makes up. For example, the fling probably makes up a smaller proportion of the universe's total normative mass than the marriage. So you can care about it less. Your headache probably makes up a lesser proportion than both of these things. So you can care about it still less. And getting priority boarding? You can care hardly at all about that.

This view captures the Drop in the Ocean Intuition. This is because the weightier the universe is, the less important is an event of given normative mass. So, seeing how much joy and suffering there is in the world should make one's own contribution seem of diminished importance. In particular, seeing that Joanna's suffering is just a drop in a whole ocean of sorrow makes alleviating it less important. This alleviation is still good. But it makes up a smaller share of the total joy and suffering in the world. This is the initial motivation for the view. We now explore Proportionalism a bit further. In particular, we will bring out some

---

10   On the first interpretation, you need some way of mapping numbers to absolute quantities of caring. To do this, start by saying that 0 maps to no caring at all. Then pick some n to map to some given amount of caring. Perhaps, for instance, 0.001 maps to a mild tinge of homesickness. Then, for any k, $I(f) = k$ just in case the amount of caring you can devote to f is $k/0.001$ times that of the homesickness.

interesting features of the view by comparing it to two alternative accounts of importance.

2.1. Intrinsicism

Let us first see how our view relates to the most natural alternative. This alternative says that the importance of a fact just is how much of what intrinsically matters inheres in the fact. In other words:

$$I(f) = NM(f).$$

Call this alternative *Intrinsicism*. Within a given universe, the importance ranking of our view and that of Intrinsicism perfectly coincide. If any two facts, f and g, occur in some universe, then Intrinsicism says f is at least as important as g if and only if our view says f is at least as important as g. We think this is a nice feature of our view.

But Intrinsicism and Proportionalism are otherwise quite different. According to Intrinsicism, all that matters to how important something is is its intrinsic properties. Proportionalism says that how it relates to something else—the entire universe—also matters. The most severe consequences of this are modal. Intrinsicism implies that intrinsic duplicates keep the same importance across possible worlds. Consider a world in which nothing much happens but a headache. Suppose you are permitted to care to some degree about the headache in this world. Now consider a world in which a duplicate of the headache happens. But suppose many other things that matter also happen. Intrinsicism implies that you are permitted to care to exactly the same degree about the headache in this other world. But Proportionalism implies the opposite. Proportionalism implies that you cannot care as much about the headache in the second world. This is because it makes up a lesser proportion of that world's normative mass. So these views differ on the importance of intrinsic duplicates.

This difference might seem detrimental to Proportionalism. Compare the lonely headache to a murder in a world full of joys and suffering. The headache makes up a large proportional share of the normative mass of its universe. The murder does not. So, on Proportionalism, the former is more important than the latter. But, one might think, surely a murder *must* be more important than a headache? If so, Proportionalism gets these cases wrong. But we think not. Now we do think there is *something* right about this thought. When a murder and a headache happen in the same world, you should care about the murder more. But the thought is not just that. It is that, even if a headache were all that ever happened, you should care less about it than you should about a murder in a well-populated world. To us, this does not seem intuitively obvious. Indeed,

there is reason to doubt it. It is in tension with the Drop in the Ocean Intuition. The murder is just a drop in the ocean. The headache *is* the ocean. So if drops are less important the less of an ocean they make up, the murder must be less important than the headache. So, we doubt this difference is detrimental to Proportionalism. But there is more to be said, so we return to Intrinsicism in section 5.

2.2. *Averagism*

Now let us see how our view relates to another alternative. This alternative says that how important a fact is is how much more (or less) normative mass that fact has than the average fact. In other words:

$$I(f) = \left| NM(\bar{f}) - NM(f) \right|,$$

where $(\bar{f})$ is the mean normative mass. Call this alternative *Averagism*. Both our view and Averagism say that the importance of a fact depends on the importance of other facts. But Averagism says that such issues can change the relative importance of two facts. Proportionalism never goes that far. According to Proportionalism, such issues only ever change the absolute importance of a fact.

   How does this make a difference? Consider the ancient Egyptians.[11] Suppose facts about the past have normative mass. Now imagine that the ancient Egyptians had a far more flourishing society than we currently believe. Suppose they lived glorious lives and there were a lot of them to boot. According to Averagism, this would make currently splendid lives less important. They would be closer to the average. But it would make currently mediocre lives much more important. They would be further below the average. So, on Averagism, facts about ancient Egypt can affect which things are more important than which other things. But that is not so on Proportionalism. According to Proportionalism, the Egyptians' fate cannot affect the relative importance of things. It can make some things of less absolute importance than they would otherwise be. But it cannot affect which things are more important than which other things. We think the consequence of Averagism here is implausible. But that of Proportionalism is tenable. But again there is more to be said. So we will return to the issue in section 4.

2.3. *Empty Worlds*

A final interesting way these views relate lies in how they treat empty worlds. These are worlds that contain nothing that matters: everything in them has zero normative mass. According to both Intrinsicism and Averagism, everything in

---

11   Jeff McMahan first discovered that ancient Egyptology mattered for some moral theories—specifically, average utilitarianism ("Problems of Population Theory," 115). But Derek Parfit made the point prominent (*Reasons and Persons*, 420).

an empty world has zero importance. Intrinsicism says this because it says the importance of a thing is just its normative mass, which is zero. Averagism says this because it says that the importance of a thing is the difference between its normative mass and the average normative mass. In an empty world the average normative mass is zero and everything has exactly this normative mass. So both views attribute everything in an empty world zero importance. But according to Proportionalism, the importance of anything in an empty world is undefined. That is because the normative mass of an empty world is zero. And you cannot divide by zero. So you cannot assess the importance of anything in an empty world.

We think this is a feature of Proportionalism, not a bug. To see why, consider what must be the case in an empty world. By definition, no facts have any normative mass in such a world. But now consider facts about the permissibility of attitudes. Suppose such facts had normative mass. Then they could not be in an empty world. But we think such permissibility facts do have normative mass. We think this because it matters whether there is a maximum amount you can care about something. This matters even—or especially—if this maximum amount is 0. This is because you can care that 0 is the most you can care about something.[12] So no such permissibility facts must obtain in the empty world. But this just amounts to the claim that the importance of everything in such a world is undefined. So, our view gets the right results in this case.

But this might be denied. Perhaps we are wrong to think that facts about what is important themselves have normative mass. Even so, we do not think this would be a significant problem. This is because we doubt it is metaphysically possible for a universe to be normatively empty. We doubt this because we think various modal facts have normative mass. Consider the fact that *there could have been intelligent life.* We think that this fact matters. Indeed, if there is no intelligent life this fact is grounds for regret. But plausibly every possible universe must contain facts like this. So every possible universe has at least some normative mass. So, how Proportionalism treats empty worlds is a feature, not a bug.[13]

12 Does it matter that you cannot care more than some amount about the fact that you cannot care more than some amount about some fact? And so on? We think so. But such facts about permissibility plausibly matter less and less the higher up you go, so do not matter infinitely much in total. For instance, the normative mass of these permissibility facts may be characterized by the sequence ½, ¼, ⅛, . . ., such that their total normative mass is 1 (since $\sum_{n=1}^{\infty}(1/2)^n = 1$). So we do not think that this kind of regress presents a problem.

13 But perhaps this is wrong. Nonetheless, Proportionalism could easily be modified. We could treat facts in non-empty worlds as dictated by Proportionalism, but stipulate that the importance of everything in an empty world is 0. This is a little inelegant. But a little inelegance ain't fatal.

3. THE INSIGNIFICANT SPECK INTUITION

We now turn to providing further motivation for the view. So far, we have just claimed it captures a widespread intuition: the Drop in the Ocean Intuition. But we think it can also capture another common intuition:

> *The Insignificant Speck Intuition:* Look up at the stars. There are a lot of them. In fact, there are 10^{13} of them in our supercluster alone. That is one followed by thirteen zeros. And our supercluster is just one in ten million superclusters. That is a lot of stars. The universe is vast and awe-inspiring. You, your actions, everyone you know, the whole history of your civilization and species is just a tiny speck in this vast universe. This realization can make all these things seem less important. The fact that the universe is vast beyond comprehension reduces the significance of your own actions.

Here is how Proportionalism can capture this. Above, we suggested that the normative mass of a world might be the amount of joy and suffering it contains. But this is quite a limited view. We already mentioned that beauty or fearsomeness might contribute to the normative mass of the world. And the list need not stop there: in particular, we think that *awesomeness*, in the sense of worthiness of awe, may contribute to normative mass. And it is plausible that the sheer size of our universe is awesome. So, given our account of importance, the vastness of the universe should diminish the importance of our actions. This is the Insignificant Speck Intuition.

This is not to say that these are the only considerations at play in the Insignificant Speck Intuition. Like with the Drop in the Ocean Intuition, other things may also play a role. Nagel, for instance, claims that the Insignificant Speck Intuition conveys the absurdity of our lives. He takes this to amount to "the collision between the seriousness with which we take our own lives and the perpetual possibility of regarding everything about which we are serious as arbitrary."[14] Perhaps. And if so, our view will not capture this aspect of the intuition. But we think this is not all there is to the intuition. We are not merely aware of the arbitrariness of our aspirations; we are aware of the insignificance of our concerns when measured against the vastness of the universe. Our view can capture this aspect of the Insignificant Speck Intuition.

14   Nagel, "The Absurd," 718.

4. A DEEPER MOTIVATION

The evidence we have just given for our view is intuitive fit. It captures various intuitions we think are widespread. This type of motivation is not to be sneezed at. Many good philosophical theories are justified by intuitive fit alone. But we can also give Proportionalism a deeper motivation. This motivation is not essential: you can accept the view without accepting this motivation. But we ourselves think this motivation provides strong support for Proportionalism. It shows that the view follows from some assumptions that we find compelling.

Roughly, the deeper motivation goes like this: to have fitting attitudes, your attitudes have to be proportional to how much things matter. A person suffering in front of you matters much more than the fact that your paper got rejected. So it would be unfitting to care more about the paper than the person. But you are a finite being with finite conative resources. There is only so much you can care about things. So consider what happens when you move through universes where more and more things matter. Because of your conative limits you cannot keep expanding how much you care about things willy-nilly. So, to keep your attitudes in proportion, you need to reduce how much you care about your paper's rejection. So how much you can care about the rejection is proportional to the normative mass of the universe. In universes where more matters, you have to care about it less. But this, if you squint right, is Proportionalism.

That is a rough statement. Now let us go a bit more carefully. We first define what it is to have overall fitting degrees of caring. This means that every collection of your degrees of caring is permissible. We now state our first assumption. This is the assumption that your degrees of caring are overall fitting if and only if they are proportional to the normative mass of facts. This assumption is made up of a necessary and a sufficient condition. The necessary condition says that it is necessary for overall fittingness that your attitudes are in proportion to the normative mass of things. So if you care much more about the paper than the person, your degrees of caring cannot be fitting. The sufficient condition says that this is also sufficient for overall fittingness. So if you care proportionally about the paper and the person (and everything else), then your degrees of caring are fitting. We can put this as follows:

Congruity: Necessarily, your overall degrees of caring are fitting if and only if, for any facts f and g, if the normative mass of f is n times greater than the normative mass of g, you care n times more about f than about g.

We expect Congruity to be controversial. We will discuss objections to it in sections 5.1–5.4. But we think it has strong *prima facie* plausibility. And we think

we can answer the objections. So we think it grounds a credible motivation for proportional thinking.

We also rely on a second assumption. This assumption is twofold. First, we assume that there is a maximum degree you are able to care about things. The idea here is just that you are a finite being with finite emotional resources. If you were a god perhaps you could care without bound. But you are not a god: there is a limit to the intensity and duration you can care about things. Second, we assume that you can have any collection of caring attitudes that together do not exceed this limit. In the relevant sense of "can," the only constraint on your caring attitudes is that you have limited conative resources. To put this another way:

> *Limits*: There is a maximum amount, T, you are able to care about things, but any collection of attitudes consistent with this limit is possible.

In what follows, we will stipulate that $T = 1$. This is required to get the precise formulation of our view. But nothing deep swings on this: it is just an arbitrary choice about which number represents the limit. This aside, when it comes to Limits we think the claim that we are not gods will be uncontroversial. But the claim that you could have any attitudes that do not exceed the limit might be controversial. We will discuss this in section 5.5.

These are our two primary assumptions. These assumptions alone have a very interesting consequence. They entail that for any fact, f, the maximal degree you are able to care about f consistent with you having overall fitting attitudes is:

$$\frac{NM(f)}{NM(U)}.$$

This just is our formula for $I(f)$. Thus the maximal degree you can care about f while having overall fitting attitudes is the proportion of the universe's total normative mass that f represents. This means that these assumptions alone get us very close to Proportionalism.

We will just give an informal proof that this entailment holds.[15] Consider any fact, f. If the share of your maximal total degree of caring that you devote to f is

---

15 Whenever we give an informal proof, we have a formal proof that we do not include. In this footnote we will outline the relevant formalization of these assumptions. But we will not provide the whole proof. We begin by setting up a framework. Remember we defined a fact as a property-instantiation or a conjunction thereof. We will represent a conjunction of property-instantiations as a set. So let $\Omega = \{f_1, f_2, \ldots, f_n\}$ be the set of property-instantiations. This means Ω represents the entire universe. Let $\wp(\Omega)$ be the set of all subsets of Ω (i.e., its power set). So this represents the set of all facts. Let NM be a non-negative function of facts representing the normative mass of a fact. Let c be a non-negative function of facts representing how much you care about a fact. Finally, let fit represent the fact that you have

greater than f's share of the normative mass of the universe, then you cannot care proportionally about every other fact. This is because if you cared proportionally about every other fact, your total degrees of caring would exceed 1. So your overall attitudes cannot be fitting. If the share of your maximal total degree of caring that you devote to f is precisely equal to f's share of the normative mass of the universe, then you can care proportionally about every other fact: you will do so if the share of your maximal total degree of caring that you devote to any other fact g is precisely equal to g's share of the normative mass of the universe. So your overall attitudes can then be fitting. So this formula describes the maximum degree you can care about f consistent with having overall fitting attitudes.

This is an interesting result. But it does not quite yet give us our account of importance. This is because we said how important something is is the most you can fittingly care about it. We just showed that caring too much about something entails your overall attitudes cannot be fitting. But this does not mean that this very attitude is unfitting. Having it might ensure some other attitude (or collection thereof) is unfitting, but it need not be unfitting itself. So, to get our account, we need the following:

> *Necessary Means*: If p necessarily implies that you violate some of your obligations, then you ought to see to it that $\neg p$.

This is a necessary means principle. It says you ought to see to the necessary means to not violating your obligations. We think such a principle explains why you should not make promises you cannot keep or have children you cannot care for. Doing so ensures you will violate an obligation. So we find this principle very plausible. Yet it is not beyond dispute. We discuss an objection to it in section 5.6. But for now, let us see how this leads to Proportionalism.

Here is an informal proof. Recall that $I(f)$ denotes the normative mass of

---

overall fitting degrees of caring. In this framework, we can formalize our two principles as follows:

> *Congruity (formal)*: fit $\leftrightarrow$ for all F and G in $\wp(\Omega)$ and any real number t such that $NM(F) = t \times NM(G)$, $c(F) = t \times c(G)$.

> *Limits (formal)*: $c(\Omega) \le 1$.

Two notes about this formalization. First, we should interpret these principles as necessarily true. Second, Limits (formal) just captures the first conjunct of Limits. The second conjunct is captured by the fact that this is the only constraint we assume on c. Assuming that $NM(\Omega) > 0$ it can be shown that for any F in $\wp(\Omega)$, $NM(F)/NM(\Omega)$ is the maximal value $c(F)$ can take while Congruity and Limits hold and fit is true. That is, in light of these assumptions, F's share of the universe's total normative mass represents the maximum you can care about F while having overall fitting degrees of caring.

f over that of the entire universe. Suppose this is the maximum one can care about f consistent with having overall fitting degrees of caring. So, if you care more than $I(f)$ about f, you must lack overall fitting degrees of caring. But that you lack overall fitting degrees of caring just means that some collection of your attitudes is impermissible. So, you violate one of your obligations. So by Necessary Means, you ought not care any more than $I(f)$ about f. Next, recall that it is possible for you to care exactly to degree $I(f)$ about f while still having overall fitting degrees of caring. But this—caring to degree $I(f)$ about f—is itself a degree of caring. So, if you can have this attitude and have overall fitting degrees of caring, this attitude can itself be permissible. So it is possible for your caring exactly to degree $I(f)$ about f to be permissible. So $I(f)$ is the maximum you can permissibly care about f.

This completes the deeper motivation for our view.[16] Roughly, Proportionalism follows from the need to keep our attitudes in proportion while being constrained by our human limitations. With this in hand, we can see better where Proportionalism and its competitors come apart.[17] If you deny Proportionalism, you must reject Limits, Congruity, or Necessary Means. Let us make this vivid

16 This is where the distinction between the maximum-permissible and the minimum-permissible notions of importance (from section 1) makes a difference. We could have characterized importance as a measure of the minimum one ought to care about something. For this notion of importance, our exact argument does not go through. This is because the right-left of Congruity implies there are no such minimums. But a closely related argument does. The related argument endorses the left-right of Congruity and endorses *Lower Limits*. Lower Limits says you must use some positive minimum proportion of your emotional resources. You must use, let us say, at least one percent of your emotional resources. Now consider any fact, f. Identify the fraction of the universe's total normative mass f makes up. Suppose you devoted less than 1 percent of this fraction of your emotional life to f. Then you could not care about all other facts proportionally. For suppose you did. Then you would not end up using even 1 percent of your emotional resources. In other words, in order to have proportionate attitudes that exceed the lower limit, the minimum degree you care about anything must be proportionate to the share of the universe's total normative mass that thing makes up. Why would one endorse Lower Limits? Well, humans are feeling animals. We may be able to numb ourselves: we may be able to care about very little. But caring about nothing whatsoever? Perhaps this we just cannot do. Motivated in this way, Lower Limits is akin to Limits: it is a principle about human capability. Alternatively, we could construe Lower Limits as a normative principle. So construed, it says that it is impermissible to use less than a certain proportion of our emotional resources. Why would that be? Well perhaps it is *wasteful* for such richly endowed beings as us to be so unfeeling. And perhaps such wastefulness is not permissible. We will not explore these options further. This is because we have our doubts about the minimum permissible degrees of caring (see section 5.3). But, if those doubts are misplaced, this provides a promising route to a version of Proportionalism for permissible minimums.

17 We can now also see why the fate of the ancient Egyptians can change a thing's absolute

for Intrinsicism. Intrinsicism, remember, said that how much you are allowed to care about something is fixed by the thing's intrinsic properties. So consider any world that contains some fact, f_1, where f_1 is weighty enough that you are permitted to devote your entire emotional life to it. Now consider a second world in which two duplicates of f_1 exist, f_2 and f_3. Suppose, in this second world, you devote your entire emotional life to f_2. If Limits holds, this means you can devote none of your emotional life to f_3. So your attitudes would be out of proportion. But, if Congruity holds, this ensures you violate some of your obligations. Yet Intrinsicism says you are permitted to devote your entire emotional life to f_2. So it follows that you can permissibly do something that ensures you violate some of your obligations. And this is just to deny Necessary Means. So, Intrinsicism must deny either Limits, Congruity, or Necessary Means. In the next section, we will see how much mileage there is in such denials.

5. OBJECTIONS TO THE MOTIVATION

We now look at ways to reject the claims on which this deeper motivation rests. This is not just an exercise in rebutting objections (although it is that). It is an exercise in seeing how our view would change if one takes these objections seriously. We think that, even if you take these objections seriously, you are left with some version of Proportionalism. We first look at some objections to Congruity. Then, we look at an objection to each of Limits and Necessary Means.

5.1. Too Demanding 1

Congruity may seem too demanding. It seems to imply that you should drain your life of personal concern. You should almost not care at all about your own joys and sorrows. You should care very little about the successes of your projects or the state of your health, or about the lives of your nearest and dearest. This is because (in our universe) a lot of other things matter. If you care proportionally about all these other things, you must care minimally about things of personal concern. But this may seem implausible. So, Congruity—the left–right of Congruity in particular—must be false.

We think this is an interesting objection to Congruity. But we do not think it refutes it. There seems to us two ways to reply to the objection. The first is to dig in one's heels: one could simply maintain that we should all care far less about our personal affairs than we normally do. This position has affinities with certain

---

importance. If the Egyptians lived glorious lives, then you ought to be happier about them than otherwise. This crowds out caring about other things.

demanding, impersonal moral theories. Consider classical utilitarianism. This implies that we should do much less than we usually do to improve our personal affairs. We should act much more for others' benefit. The heel-digging position implies that we should care much less than we usually care about our personal affairs. We should care much more about other people. Both views go against much commonsense thinking. But neither, we think, is totally untenable. We might ordinarily do very poorly in both action and attitude. So, we think that if you find demanding, impersonal moral theories congenial, then you should find this position congenial.

The second response is more concessive. One might simply endorse an agent-relative theory of normative mass. Such a theory says that how weighty a thing is depends on whose perspective we assess it from. From your perspective, your wine collection might be a weighty matter. But from mine it is not so weighty. This view makes sense if you think that agent-relative properties contribute to normative mass. Such properties include those that reference you essentially. To assess how much of your joy and your suffering something has is to assess its agent-relative properties. Your projects usually involve more of your joy and your suffering than do those of anyone else. So, from your perspective, they may make up a larger proportion of the universe's normative mass than do those of anyone else. So, caring copiously about your personal affairs is compatible with Congruity. You just need an agent-relative theory of normative mass. With such a theory, Proportionalism should then be interpreted as itself agent-relative. It defines a different importance ranking for each person.

But maybe neither of these responses works. This would provide grounds to reject Congruity. But it need not be *wholesale* rejected. For it seems we can distinguish between different domains: the impersonal domain and the personal domain. We can understand this distinction as one between different kinds of facts. For a fact to be in your personal domain might just be for it to essentially concern *your* projects, or *your* loved ones, or *your* well-being. But facts in the impersonal domain do not essentially concern you at all. One way to think of the impersonal domain is as the domain of agent-neutral facts. The personal domain is then thought of as a subset of the domain of agent-relative facts. Once we make this distinction, we can say that Congruity applies only in the impersonal domain. It does not apply in the personal domain. On this view, Proportionalism may hold for one class of facts, but it need not hold for them all. We will discuss this view further in section 6.1.

5.2. Too Demanding II

Congruity may seem too demanding in a different way. It may allow us too lit-

tle discretion in what we care about. For example, some of us choose to devote much of our emotional energy to economic injustice in England. But this injustice might involve less of what matters than does the malaria infection worldwide. After all, over two hundred million people suffer from malaria every year. Yet perhaps our attitudes are nonetheless fitting. Perhaps we have some choice over how much, relatively speaking, we care about worthy things. Yet that is inconsistent with Congruity. So, Congruity—the left–right of Congruity in particular—must be false.

One could respond to this by again digging in one's heels. Some people, such as classical utilitarians, think we have little choice over what we may do. Analogously, perhaps we have little choice over what we may care about. But suppose we take the objection more seriously. Even so, there is a reply to it. First, note that normative masses can be incommensurable. Perhaps there is no definite way that economic injustice and malaria weigh up. Rather, the normative mass of economic injustice is indeterminate, but determinately between a quarter and half of that of malaria. This calls for a minor modification of Congruity. Suppose one cares about economic injustice anywhere between a quarter and half as much as one cares about malaria. Then we should say one has fitting attitudes.[18] This already gives us some choice over how much we care about things. But now let us exploit the agent-relative conception of normative mass we previously mentioned. On this conception it is plausible that the fact that you cared a lot about something in the past can make that thing weightier now. It can make it matter more for you today. So suppose, in June, you cared half as much about economic injustice as you did about malaria. This means economic injustice has a little extra normative mass in July.[19] So, in July, perhaps you may care anywhere

18 More formally, we use a *set* of functions $NM = \{NM_1, NM_2, \ldots\}$ to jointly represent the normative mass of facts. We then say that your overall degrees of caring are fitting if and only if there is *some* function NM_i in NM that you satisfy Congruity with respect to. A version of our argument for Proportionalism then goes through.

19 Will malaria also have a little extra normative mass? Perhaps not. It might only get weightier when you have cared about it more than you had to, given your other attitudes. It might be voluntary devotion alone that adds mass to a thing. But we need not rely on that. The described mechanism works as long as you may treat economic injustice as gaining proportionately more mass than malaria. Suppose you may consistently treat malaria as getting 9 percent weightier but injustice as getting 11 percent weightier. Then, eventually, the normative mass of the injustice will exceed that of malaria. Why might this be permissible? Well, it is plausibly indeterminate how much mass your having cared about something adds to its existing mass. But your caring might determinately add between 9 and 11 percent to this existing mass. And perhaps you may consistently choose to care in line with the lower end of this range for malaria and the higher end for injustice. So you may consistently treat the injustice as getting proportionately weightier than malaria.

between one and two-thirds as much about it as you do about malaria. Suppose you choose two-thirds. Then, in August, you may care anywhere between half and three-quarters as much about it as you do about malaria. We are off to the races. By iterating this process, you can step up how much you may care about worthy things. And you can do so without ever having unfitting attitudes. We think this gives us ample choice over our relative degrees of caring.

But maybe that is wrong. Maybe we need more choice. Yet it seems implausible that one has complete discretion over what one cares about. Suppose you care a lot about economic injustice. That does not mean that, when someone dies in front of you, you may remain completely cold. This would be to have unfitting attitudes. So we should not reject Congruity wholesale. We should reformulate it. We should say that one's attitudes are fitting when they are not *too far* out of proportion. There are a few ways to capture this. One way is to say that there is a factor $k > 0$ such that, when f is n times the normative mass of g, your attitudes are fitting whenever you care between $k \times n$ and k/n as much about f as about g. So, for example, suppose that malaria has twice as much normative mass as economic injustice. If $k = 4$, then it is permissible to care twice, but at most twice, as much about the injustice as about malaria. But there are other ways to capture the point. Whichever way we do it will get us a version of Proportionalism.[20] If your attitudes cannot get too far out of proportion, then in weightier worlds you may care less about anything in particular.

5.3. *Too Permissive 1*

Congruity may instead seem too permissive. This is because it allows you to care very little about certain things. It says that, if your attitudes are in proportion, then they are fitting. But your attitudes can be in proportion without you caring much about anything. So your attitudes can be fitting when you care about things very little.[21] Yet suppose you are in the midst of a famine. And suppose you care very little about everything. Then this means you are allowed to also

20   The simple suggestion in the text gives us a somewhat complex version of Proportionalism. It implies that the importance of f equals

$$\frac{k \times \mathrm{NM}(f)}{(k-1) \times \mathrm{NM}(f) + \mathrm{NM}(\upsilon)}.$$

21   Indeed, your attitudes are proportionate in the sense of satisfying the right-hand side of Congruity when you care about nothing at all. To see this, imagine f and g are the only facts. Suppose you care to degree 0 about g. But suppose f is twice as weighty as g, so $\mathrm{NM}(f) = 2 \times \mathrm{NM}(g)$. Then you must care to degree 2×0 about f. But you do this by caring to degree 0 about f. So, when you do not care about anything, your attitudes satisfy the right-hand side of Congruity.

care proportionally little about those starving around you. Yet one might object that caring so little about such a famine is pathologically cold. You are obligated to care some minimum amount about someone starving in front of you. So, Congruity—the right–left of Congruity in particular—must be false.

We think this is an interesting objection. But here is a reply to it. This objection, we think, has two sources. First, most of us care a lot about some things. We care a lot about our jobs, our partners, our receding hairlines, and so on. So, for most of us, when we care little about someone starving, our attitudes are out of proportion. So, they are impermissible. But, when this is the source of the objection, it poses little threat to Congruity. Congruity is consistent with this claim. The second source, we think, concerns the link between moral normativity and fitting attitudes. We suggested one such link in section 1. We suggested that one had moral reason to care fittingly about things that matter morally. But perhaps there are other links. In particular, perhaps one does something morally wrong when one cares very little about what matters morally. One's attitudes might not be unfitting, in this case, but they might nonetheless be morally suspect. Yet when this is the source of the objection, it again poses little threat to Congruity. Congruity is a principle about the type of normativity confined to caring, not moral normativity. So, we do not think this objection refutes Congruity after all.

But perhaps that is wrong. Perhaps there are some facts about which you must care to some minimum degree. Nonetheless, this does not really imperil Proportionalism. This is because Proportionalism is a thesis about the maximum degrees you can care about things. It identifies some amount as the maximum you can permissibly care about something. We only used the right–left of Congruity to establish that caring about something this much is permissible. But the objection does not dispute this. It does not suggest that caring up to this maximum amount is impermissible. It says caring below a minimum amount is impermissible. So let us just assume that you are allowed to care up to this amount. But assume you must still keep your attitudes in proportion. Then Proportionalism follows anyway. So, if we maintain this objection, it is not that we must reject Proportionalism. It is just that Proportionalism is not the whole story. When saying how much you can care about things, we must tell a story about minimums as well as maximums. But Proportionalism may be an essential part of the story without being the whole story.[22]

---

22 Note 16, above, contains a brief Proportionalist story about these minimums. So, even if this objection is maintained, we may not need a non-proportionalist story.

5.4. Too Permissive II

Congruity may seem too permissive in a different way. If you live in a universe where nothing much matters, then Congruity permits you to care a lot about tri-fles. For instance, suppose you lived in a world with just you and the James Bond films. According to Congruity, you are then permitted to care enormously about Bond. But one might object that caring so much about Bond is utterly over the top. Even in such a world, there is an upper limit to how much you can care about 007. On this view, you just are not allowed to care that much about certain things. So, Congruity—the right–left of Congruity in particular—must be false.

This objection would imperil Proportionalism. For suppose there are upper limits independent of Proportionalism. Then the degree of caring identified by Proportionalism might exceed the maximum permissible degree. But we are skeptical of the objection. We think its force comes from the fact that, in our world, the Bond films are rather trivial. This is because our world is full of important things. So, in our world, one should not care too much about Bond. Yet suppose you really are in a universe where nothing but Bond matters. Then we see no reason why you cannot devote large parts of your emotional life to Bond. Why must you twiddle your emotional thumbs? Why must you let part of your emotional life go to waste? It seems puritanical to deny us access to the full expression of our emotional capacity. So, we doubt there are independent constraints on caring of this sort.

But maybe that is wrong. Maybe there are some things you just are not al-lowed to care that much about. Nonetheless, we think that there must be a large class of facts that are free of such constraints. Consider, for example, the Sich-uan earthquake. This was a tragedy. Over eighty thousand people died. We think that, apart from Congruity itself, there are no constraints on how much you can care about such a thing. If it were the only thing that ever happened, you would not be mistaken in devoting all your emotional resources to it. We also doubt that, apart from Congruity, there are any constraints on how much you can care about the death of a loved one. And, to boot, we think you can care as much as you want about Sunset Boulevard. So the above argument will go through with respect to this class of facts. In other words, Proportionalism holds for one class of facts, but not them all.

5.5. Not Limiting Enough

We have just discussed some objections to Congruity. We now discuss an ob-jection to Limits. Limits says that there is a maximum amount you are able to care about things, but you are able to have any attitudes consistent with this

maximum. This objection says that Limits is too *permissive*. It says that there are some collections of attitudes one cannot have, despite their consistency with the limit. Suppose, for instance, that you just cannot bring yourself to care about bird-watching. Maybe your co-author likes bird-watching very much. But it is just not for you. Swifts and swallows (and even sparrows) leave you cold. Bird-watching is something you cannot care much about, and this incapacity has nothing to do with you having exhausted your *overall* emotional resources.

This is another interesting objection. But we do not think it refutes Limits. Rather, it shows that Limits invokes a specific type of modality, and a specific notion of importance. The type of modality just described seems like *psychological* possibility. It is psychologically possible for you to have some attitudes if it is metaphysically possible and consistent with central facts about your psychology. But psychological modality is very variable: what is psychologically possible for someone depends on their peculiar quirks and foibles. So we doubt that the notion of importance explicated in *these* terms is that interesting. It is too contingent, and so lacks generality. Limits characterizes a different type of modality, which we might call Limits-possibility: it is Limits-possible for you to have some attitudes if it is metaphysically possible and consistent with Limits.[23] We think that this gives rise to an interesting notion of importance. This notion is sensitive to the fact that we are finite, but insensitive to any of our further idiosyncrasies.

But perhaps this is wrong. Perhaps the only interesting notion of importance is the one keyed to psychological possibility. Nonetheless, Congruity and Necessary Means still get us a kind of proportional thinking for this notion. To see this, consider the fact f_1 that you are least able to care about, in proportion to its share of the universe's total normative mass. Maybe this is a very weighty fact, but one you can only bring yourself to care about moderately. Or maybe it is a lightweight fact, but one you can hardly bring yourself to care about at all. Congruity and Necessary Means then entail that we must adjust how much you can permissibly care about any other fact, f_k, by how much you can care about this fact. To see how, first identify the fraction of your emotional life you are able to devote to f_1. Then see what fraction of the universe's total normative mass f_1 makes up. Then divide the former by the latter. This new fraction identifies how much you can care about f_1 in proportion to how much you are permitted to care about it. How much of your emotional life you are allowed to devote to any f_k is then this fraction multiplied by the fraction of the universe's total normative mass

---

23 Limits-Possibility might just be *physical* possibility. We doubt anyone is physically incapable of caring about bird-watching. This means we need not think of it as a novel, unfamiliar kind of possibility.

that f_k makes up. So, even if this objection carries the day, you are stuck with a
type of proportional thinking.

5.6. Necessary Means Equivocates

Finally, we will discuss an objection to Necessary Means. Necessary Means says
that you are not permitted to have attitudes that ensure you violate an obliga-
tion. The objection says that Necessary Means rests on an equivocation. It is
a plausible principle for moral permissibility and practical permissibility. But
the notion of permissibility we are using is that which rules attitudes of caring.
And there might be no reason to think Necessary Means holds for this notion.
Having some attitude might ensure you violate some obligation, without that
attitude being impermissible *qua* attitude. So the argument for Proportionalism
fails at the final step.

This objection is also interesting. But we doubt it goes through. We think
that the plausibility of Necessary Means in these other cases is probative. The
normativity that rules attitudes of caring seems enough like these other types
of normativity that we expect it to obey similar structural rules. That Necessary
Means holds in these cases is evidence that it holds for the type of normativity
that governs attitudes of caring. So, we think Necessary Means does hold for the
type of normativity with which we are concerned.

But perhaps this is wrong. Then Proportionalism does not follow for the nor-
mativity specific to attitudes of caring. Nonetheless, we can identify a kind of
normativity for which it does follow. It follows for practical rationality, holding
fixed certain desires. For suppose we hold fixed our desire to have overall fitting
attitudes. The argument in section 4 shows that conforming to Proportionalism
must be a necessary means to having overall fitting attitudes. And practical ratio-
nality does obey a necessary means principle. If you want p, and q is a necessary
means to p, then you ought (practically speaking) to see to it that q. So it follows
that you ought, practically speaking, to conform to Proportionalism.[24] So Pro-
portionalism follows for the practical ought, holding fixed our desire to have
fitting attitudes. This version of Proportionalism has a somewhat pragmatic air.
It identifies what, practically speaking, we need to do if we want fitting attitudes.
But we think most of us do want fitting attitudes. And we think this desire makes
sense: having fitting attitudes is part of a life well lived. If you actually only care

---

24 This necessary-means principle may be wide-scope (see, e.g., Broome, "Wide or Narrow
    Scope?"). So Proportionalism follows for sure only with Necessary Detachment. Necessary
    Detachment says that if p is fixed and you ought to see to it that $(p \to q)$, then you ought to
    see to it that q.

about Bond, then your life is impaired. So, even if you deny Necessary Means, in a practical sense you remain bound by Proportionalism.

That completes our account of importance. We know of no fatal objections to this account, it fits some intuitive data, and it has a nice motivation. So we think the account is not obviously false. It may even be true. We also just identified several fallback positions that count as kinds of Proportionalism. So, if it is not true in letter, it might still be true in spirit. In the rest of the paper, we will explore ways to extend these views. We first look at how one might extend the fallback position identified in section 5.1. We then look at extensions one could make to any of these views, including full-blooded Proportionalism. We think one can take or leave these extensions. But this exploration helps show the implications (and fruitfulness!) of proportional thinking.

6.1. Domain-Specific Proportional Thinking

In section 5.1, we suggested that Proportionalism might only apply in the impersonal domain. But we did not say how you should deal out your degrees of caring in other domains. Nor did we say how you should deal out your degrees of caring *between* domains. We think proportional thinking can be extended to both issues. In this section, we will show how this might be done. This means, we think, that such thinking is a much wider phenomenon than that captured by Proportionalism alone. You can adopt it even if you restrict or reject this view.

Let us begin with the personal domain. We previously suggested that each person's personal domain is made up of the facts that referenced them essentially. But we did not say how much you are allowed to care about each fact. So, consider something in your personal domain. Consider, for instance, that fiery love affair you had in college. What is the most you can permissibly care about this? That may depend on what the rest of your life looks like. If the rest of your life is full of only Muzak and potatoes, then this affair is very important indeed. You can devote a lot of your emotional life to it. But if you have lived a more full life, perhaps a life full of such affairs, then the college dalliance is less important. You can devote less of your emotional life to it. So, the importance of any particular fact in your personal domain diminishes as your personal domain gets weightier. The more full your life, the less personally important is any individual event in your life.

Here is how to spell this out. Let $I_p(f)$ be how much concern you can permis-

sibly devote to some fact, f, in your personal domain. Let $NM(P)$ be the norma-
tive mass of the entirety of your personal domain. Then this view says:

$$I_p(f) = \frac{NM(f)}{NM(P)}.$$

This view is formally similar to Proportionalism. In both cases, the importance
of a fact is its normative mass adjusted for how much normative mass there is in
some wider thing. With Proportionalism the wider thing is the universe. With
this view the wider thing is the entirety of your personal domain. So this is an
extension of proportional thinking to the personal domain.

Let us now see how proportional thinking can also extend to the domain of
rights. This domain is made up of essentially rights-involving facts. So imagine
you punch someone. If this is the only rights violation that ever happened, then
it looks quite important. Feelings of penitence may swell your breast. But you
might instead live in a more callous world. You might live in a world in which
rights violations are commonplace. Then perhaps this violation is less import-
ant. It would be self-indulgent to wail and moan over this rights violation, when
rights are being violated all around. We can represent this as before. Let $I_r(f)$ be
how much concern you can permissibly devote to some fact, f, in the domain of
rights. Let $NM(R)$ be the normative mass of the entirety of the domain of rights.
Then this view says:

$$I_r(f) = \frac{NM(f)}{NM(R)}.$$

This gives us two examples of domains—apart from the impersonal—in which
proportional thinking may apply. We see these examples as proofs of concept.
We are unsure how exactly one should demarcate the domains if engaging in do-
main-specific proportional thinking. Perhaps the domain of virtue should have
its place alongside these other domains. Perhaps we should include the domain
of departmental politics. We do not know. The important point is just that, even
if Proportionalism itself is too general, proportional thinking may not be.

So, how should you dole out your degrees of caring between domains? We
will note two views on this. We start with a non-proportionalist view. On this
view, you can devote a fixed proportion of your emotional life to each domain.
The simplest way to implement this is to say you can devote the same amount to
each domain, and together these sum to the entirety of your emotional life. So,
you can devote a third of your emotional life to the impersonal domain, a third
to the personal domain, and a third to the domain of rights. On this view, how
much you can care about the things in some domain never changes. And, on

this particular implementation, you cannot trade-off between domains. You can never devote more than a third of your emotional life to the impersonal domain, even if you do not care at all about the other domains.

But we prefer a second, more proportionalist view. On this second view, how much you can care about each domain can change. It is proportional to—but not the same as—the fraction of the universe's total normative mass that a domain makes up. For instance, suppose that the normative mass of the personal domain is a quarter that of the entire universe. You might nonetheless be allowed to devote half your emotional life to it. But suppose it is an eighth of the mass of the entire universe. Then you might be allowed to devote a quarter of your emotional life to it. In other words, there is some factor, j, such that the proportion of your emotional life you can devote to things in your personal domain is $j \times NM(P)/NM(U)$. If this factor is large enough, you can care a lot about your personal affairs. But the weightier the universe, the less you can care about such matters. We think this view captures much of the virtue of Proportionalism, without its possibly overbearing character. So, we think it is the best way to implement domain-specific proportional thinking.[25]

6.2. Reasons

We have so far focused on the permissibility of attitudes. In this section, we look at how to extend proportional thinking to the strength of certain reasons. The simplest way to do this is to assume that there are links between fitting attitudes and reasons. In particular, it seems that we can think of certain emotions—hope, pleasure, anticipation—as positive. We can think of others—terror, displeasure, trepidation—as negative. Some of these positive emotions may connect to reasons to bring about something. Some of the negative emotions may connect to reasons to prevent that thing. We will call these reasons *consequentialist reasons*. So this claim comes down to:

> *Bridge*: How much consequentialist reason one could have to bring about *f* is proportional to how much of these positive emotions one can (permissibly) direct toward *f*. How much consequentialist reason one could have to prevent *f* is proportional to how much of these negative emotions one can (permissibly) direct toward *f*.

Suppose a principle like this is true. Then a simple extension of Proportional-

25 Every such view is extensionally equivalent to one of the agent-relative versions of Proportionalism mentioned in section 5.1. But they differ in their conceptions of normative mass. Domain-specific views have no need of an agent-relative conception of normative mass. But this is essential to agent-relative versions of Proportionalism.

ism will extend proportional thinking to reasons. This extension holds that how much of these positive or negative emotions one can feel about any fact is proportional to the normative mass of the entire universe. So, how much consequentialist reason one has to bring about (or prevent) some fact is proportional to that fact's share of the universe's entire normative mass. The weightier the universe, the less weighty are consequentialist reasons.

In the rest of this section, we will look at two issues this extension generates. The first issue is straightforward. Even if the strength of *consequentialist* reasons varies with the weight of the entire universe, the strength of other kinds of reasons may not. Call a reason that does not vary in this way a *local reason*. It follows that, the weightier the world, the stronger are local reasons relative to nonlocal reasons. Here is a concrete example. Suppose that reasons not to violate rights are such local reasons. The strength of one's reason not to punch people is not diminished by how much joy and sorrow is in Alpha Centauri. This implies that, in weightier worlds, reasons not to violate rights become relatively strong. This is because, in weightier worlds, the strength of consequentialist reasons is diminished. This is an interesting implication. It suggests that the weightier the world is, the fewer are the circumstances in which a rights violation is permissible. In lightweight worlds, consequentialist considerations win out. But, in heavyweight worlds, rights rule the day.

The second issue will take longer to lay out. This is a choice point. If we extend Proportionalism to reasons, we face the following question: What should we do under conditions of uncertainty? The standard view is that one should do whatever has the highest expected value. An action's expected value is the value it would realize in each state of the world, weighted by how likely that state is. The Proportionalist can adopt this view. They can say that, when we do not know what the effects of our actions are, we should do the thing with the highest expected value. The Proportionalist flavor comes in when assessing the importance of such prospects. Their importance, one can say, is their expected value divided by the normative mass of the whole universe. It is not their expected value alone.

But there is a second, more radical option. This involves a different view of how one should act under conditions of uncertainty. This different view says that one should do what one has the highest expected strength of reason to do. The expected strength of a reason is how strong that reason would be in each state of the world, weighted by how likely that state is. To put this formally, let $R_i(A)$ be the strength of reason one has to perform A in the ith state of the world. Then the expected strength of one's reason to A is:

$$ESR(A) = \sum_i R_i(A) \times PR(s_i).$$

This view seems defensible to us. Indeed, expected value maximization may seem right only because, in many cases, it matches expected strength of reason maximization. So, suppose you think that we should do what maximizes expected strength of reason. We then get a radical consequence. There will be cases where you should not maximize expected value. For suppose Proportionalism is extended to reasons. Then maximizing expected strength of reasons will often mean maximizing expected proportional improvements. But expected proportional improvement and expected value do not coincide. They diverge in cases where you can have big proportional changes that are, in absolute terms, relatively small. So, in such cases, you should not maximize expected value. What do such cases look like? Here is an example:

> *Super-technology*: Suppose you can develop the super-technology. If humanity is alone in the universe, then this will make the universe much better. It will allow us to bring life to worlds otherwise empty of value. But suppose the universe is teeming with intelligent life. Then the universe is already overflowing with value. This might seem like reason to rejoice. But, alas, man cannot control his warlike nature. We would inevitably use the super-technology to bring terror to other worlds. This will make the universe much worse. Should you develop the super-technology?

Let us formalize the case. For simplicity, suppose the world where we are alone has a normative mass of 4. Suppose the one where we have company has a normative mass of 40. These assumptions track the fact that more of what matters has happened in the latter than the former. Now let $n \to k$ represent a move from a world of value n to one of value k. Then we can write this decision problem as:

States of the World	$NM(U) = 4$	$NM(U) = 40$
Probabilities	0.5	0.5
Develop	$1 \to 2$	$10 \to 8$
Do Not Develop	$1 \to 1$	$10 \to 10$

So formalized, it is clear that the expected *value* of not developing the super-technology exceeds that of developing it. But the expected strength of reasons to develop it exceeds that of not developing it. The relevant calculations are:

$$ESR(\text{Develop}) = (0.5 \times \tfrac{1}{4}) + (0.5 \times \tfrac{-2}{40}) = 0.1,$$

$$ESR(\text{Do Not Develop}) = (0.5 \times \tfrac{0}{4}) + (0.5 \times \tfrac{0}{40}) = 0.$$

So, if you should maximize expected strength of reasons, you should develop the super-technology. So, you should not do what maximizes expected value. This is why we call this a radical option. It materially diverges from the standard view. We are unsure whether the radical or conservative option is better. But we think this is an important choice point generated by applying Proportionalism to reasons. More generally, we think extending Proportionalism to reasons raises interesting, and *sui generis*, issues.

6.3. Subjective Proportionalism

Let us look at one final extension of Proportionalism. This also has to do with expectations. Namely, people often distinguish between objective and subjective reasons.[26] Suppose you have a glass of clear liquid in front of you. You are thirsty. You think the clear liquid is water. So you have good subjective reason to drink from the glass. But, actually, what is in the glass is poison. So you have got an objective reason not to drink it. Roughly, the distinction is this: objective reasons depend on just the facts. But subjective reasons depend on what you think the facts are, or perhaps what you ought to think the facts are.

This distinction can also be made among fitting attitudes. Proportionalism has an objective character. How much normative mass you *think* the universe has does not matter. All that matters is how much normative mass it *does* have. But there seems to be room for a subjective notion of importance. This would let what you think, or perhaps should think, about normative masses determine how much you can permissibly care about something.

It is straightforward to characterize such a notion. We first help ourselves to the notion of expected normative mass. Something's expected normative mass is the probability-weighted average of the normative masses it might have. For a fact, f, we denote this $ENM(f)$. We then say that the *subjective* importance of a fact just is its expected normative mass divided by the expected normative mass of the entire universe. In other words:

$$I_s(f) = \frac{ENM(f)}{ENM(U)}.$$

We can read the relevant probabilities either as actual credences or as evidential probabilities. On the first reading, this gives you a notion of importance dependent on what you actually think. On the second, it gives you a notion of importance dependent on what you ought (epistemically speaking) to think. Subjective importance and subjective reasons seem interesting on the same grounds.

26   There is a vast literature on this distinction. For an extensive discussion, see, e.g., Zimmermann, *Living with Uncertainty*.

They both satisfy intuitions that our beliefs (or our evidence) should matter to our obligations. And they are perhaps, in a sense, more action-guiding. We could be deeply uncertain what the universe's total normative mass is, but we might still be able to work up an expectation. So we can still establish how much, subjectively speaking, we can permissibly care about something. This notion of subjective importance may be a worthwhile extension to Proportionalism. It is the last extension to Proportionalism we will look at.

7. CONCLUSION

We have presented a theory of importance. This theory says that how important something is is determined by the proportional contribution it makes to the total normative mass of the whole universe. We think this fits some intuitive data, has a nice motivation, and can be extended in various interesting ways. But enough of that. Suppose Proportionalism is true. How should you feel about this? Well, it depends. If you are rejoicing at the splendor of your life, proud of your achievements, and excited about your projects, then our view will come as a setback. You must keep it in perspective: in all likelihood, the good in your life is less important than you think. You need to tone down your jubilance. On the other hand, if you are overwhelmed by woe, shattered by loss, trembling with concern, then our view may come as a comfort. You may keep it in perspective: in all likelihood, the evil in your life is less important than you think. You need not be so devastated. You may, we think, cease taking yourself so seriously.[27]

New York University
adam.lovett@nyu.edu

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

REFERENCES

Bartels, Daniel. "Proportion Dominance: The Generality and Variability of Fa-

voring Relative Savings over Absolute Savings." *Organizational Behavior and Human Decision Processes* 100, no. 1 (May 2006): 76–95.

Bostrom, Nick. "Infinite Ethics." *Analysis and Metaphysics* 10 (2011): 9–59.

Broome, John. "Wide or Narrow Scope?" *Mind* 116, no. 462 (April 2007): 359–70.

Bykvist, Krister. "No Good Fit: Why the Fitting Attitude Analysis of Value Fails." *Mind* 118, no. 469 (January 2009): 1–30.

D'Arms, Justin, and Daniel Jacobson. "The Moralistic Fallacy: On the 'Appropriateness' of Emotions." *Philosophy and Phenomenological Research* 61, no. 1 (July 2000): 65–90.

Fetherstonhaugh, David, Paul Slovic, Stephen M. Johnson, and James Friedrich. "Insensitivity to the Value of Human Life: A Study of Psychophysical Numbing." *Journal of Risk and Uncertainty* 14, no. 3 (May/June 1997): 283–300.

Firth, Roderick. "Ethical Absolutism and the Ideal Observer." *Philosophy and Phenomenological Research* 12, no. 3 (March 1952): 317–45.

Glover, Jonathan, and M. J. Scott-Taggart. "It Makes No Difference Whether or Not I Do It." *Aristotelian Society Supplementary Volume* 49 (1975): 171–209.

Loewenstein, George, and Karen E. Jenni. "Explaining the 'Identifiable Victim Effect'." *Journal of Risk and Uncertainty* 14, no. 3 (1997): 235–57.

MacAskill, William. *Doing Good Better: How Effective Altruism Can Help You Make a Difference*. New York: Penguin, 2015.

McMahan, Jeff. "Problems of Population Theory." *Ethics* 92, no. 1 (October 1981): 96–127.

Nagel, Thomas. "The Absurd." *Journal of Philosophy* 68, no. 20 (October 1971): 716–27.

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

Rabinowicz, Wlodek, and Toni Rønnow-Rasmussen. "The Strike of the Demon: On Fitting Pro-Attitudes and Value." *Ethics* 114, no. 3 (April 2004): 391–423.

Railton, Peter. "Moral Realism." *Philosophical Review* 95, no. 2 (April 1986): 163–207.

Singer, Peter. *The Life You Can Save: Acting Now to Stop World Poverty*. New York: Random House, 2009.

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

Unger, Peter. *Living High and Letting Die: Our Illusion of Innocence*. New York: Oxford University Press, 1996.

Zimmermann, Michael. *Living with Uncertainty: The Moral Significance of Ignorance*. Cambridge: Cambridge University Press, 2008.

CAN STREUMER SIMPLY AVOID
SUPERVENIENCE?

Luke Elson

Bart streumer has recently defended an error theory for all normative claims.[1] Streumer's argument is by elimination: he offers original arguments against all competing metaethical views. A major such competitor is non-reductive realism about the normative. This view holds that there are irreducibly normative properties that are not identical to natural or descriptive properties.

To see Streumer's argument against non-reductive realism, first consider Frank Jackson's famous reduction argument against that view. Jackson's argument depends on the supervenience of the normative on the descriptive:

> s: For all possible worlds w and w^*, if the instantiation of descriptive properties in w and w^* is exactly the same, then the instantiation of normative properties in w and w^* is also exactly the same.[2]

Jackson's main idea is this: given a normative predicate such as "is wrong," the truth of supervenience allows us to construct a highly artificial descriptive predicate that is necessarily co-extensive with the normative predicate.[3]

Now, consider the following criterion of property identity:

> n: Two predicates ascribe the same property *iff* they are necessarily co-extensive.[4]

If this criterion is correct, then the normative predicate and its artificial descriptive counterpart ascribe the same property. This implies that non-reductive realism is false, at least assuming that the "shared" property is a descriptive one.

There are obviously many places to object to this argument. For example,

---

1   Streumer, *Unbelievable Errors*.

2   Streumer, *Unbelievable Errors*, 25.

3   Jackson, "From Metaphysics to Ethics," 122–23.

4   Streumer, *Unbelievable Errors*, 30.

one might object to the distinction between normative and descriptive predicates, or to the criterion of property identity N.

One of its (apparently) less controversial aspects is the supervenience claim S. But S has recently been questioned.[5] Streumer offers a new version of the reduction argument that putatively "does not appeal to any claim about supervenience at all."[6]

If Streumer's version of the reduction argument indeed avoids appeal to supervenience, then this is significant even for non-error theorists. Most obviously, it would add to the stock of arguments against non-reductive realism, and reduce their dependence on S. Critics of non-reductive realism (including both error theorists and other brands of realist) should welcome a new argument against that view, especially one that relies on weaker assumptions.

In this article, however, I will argue that Streumer's argument indeed relies on supervenience. The dependence is more opaque than in Jackson's version of the argument, but without S, the argument fails.[7]

1. STREUMER'S SIMPLE MORAL THEORY ARGUMENT

To get the argument going, we first assume that some simple moral theory is correct.[8] By way of example, take hedonistic act utilitarianism:

> *Utilitarianism*: Necessarily, an action is right if and only if it maximizes happiness.[9]

Since "maximizes happiness" is a descriptive predicate, Utilitarianism says that "is right" is co-extensive with a descriptive predicate ("maximizes happiness"), and that this co-extension holds across all possible worlds. In other words, Utilitarianism implies:

---

5   For discussion and references, see Väyrynen, "The Supervenience Challenge to Non-Naturalism."

6   Streumer, *Unbelievable Errors,* 26, 30–35. An earlier version appeared in Streumer, "Why There Really Are No Irreducibly Normative Properties," 315–18.

7   Dunaway also discusses reduction without supervenience ("Supervenience Arguments and Normative Non-Naturalism"). I will not consider his view here, because I am evaluating whether Streumer's argument is a *distinctive* supervenience-avoiding refutation of non-reductive realism.

8   The normative property in question throughout is that of moral rightness, but the argument is intended to extend to all normative properties.

9   Streumer, *Unbelievable Errors*, 30.

> *Right-Description Necessity*: The predicate "is right" is necessarily co-extensive with a descriptive predicate.[10]

Now, applying N, Right-Description Necessity in turn implies that rightness and a descriptive property (in this case, maximizing happiness) are the same property. So non-reductive realism is false.

We reached this conclusion by assuming that the correct moral theory is simple, and that it is Utilitarianism. But the conclusion would not be very interesting if it relied on these assumptions. So Streumer appeals to the following principle:

> w: Whether normative properties are identical to descriptive properties cannot depend on which first-order normative view is correct.[11]

If w is true, then nothing hung on the truth of Utilitarianism, or more generally on the truth of a *simple* first-order moral view. So the assumptions can be dropped, but the conclusion—that non-reductive realism is false—stands.

The supervenience claim s did not explicitly feature in this argument. But, I will argue, the argument nevertheless depends on s: its conclusion applies only to moral theories that imply the truth of supervenience. In particular, the modal operator "necessarily" in Utilitarianism smuggles in the supervenience claim.

2. REDUCTION-FRIENDLY THEORIES

The heart of my criticism is this: Right-Description Necessity amounts to supervenience, and the argument succeeds only for that restricted set of first-order moral theories that entails Right-Description Necessity. Therefore, the argument covertly relies on s.

Utilitarianism has the following structure: necessarily, an act x is right *iff* $\phi(x)$, where ϕ is a descriptive predicate. In other words, Utilitarianism ascribes a necessary and sufficient condition ϕ for rightness, where ϕ is *descriptive*, and says that this biconditional holds necessarily. Call first-order moral theories with this structure "reduction friendly."

Clearly, any reduction-friendly moral theory implies Right-Description Necessity. But Right-Description Necessity entails supervenience. To see this, suppose that two worlds w and w^* are such that all descriptive properties are the same. ϕ is a descriptive property, so the same acts are ϕ in w and w^*. Therefore,

---

10 As Streumer puts it: "If [Utilitarianism] is correct, the predicate 'is right' is necessarily co-extensive with the descriptive predicate 'maximizes happiness'" (*Unbelievable Errors*, 31). I take this—and the other necessity claims in play—to involve a kind of metaphysical necessity: moral theories that violate it are not logically incoherent.

11 Streumer, *Unbelievable Errors*, 31.

and applying our reduction-friendly theory of the normative property in question, the same acts have that normative property in w and w^*. But this is just a statement of s.

So Streumer's chosen example happens to be one that implies supervenience. But could his argument instead work with some *other* simple moral theory—one that does not imply s?

To see why not, consider some other putative moral theories. These examples show that a moral theory that is not reduction friendly either fails to entail Right-Description Necessity (and is silent on the question of supervenience), or entails the falsity of Right-Description Necessity (and the falsity of supervenience).

First:

> *Simpler Utilitarianism*: For all actions x in the actual world, x is right if and only if x maximizes happiness.

Simpler Utilitarianism is a universally quantified biconditional with no modal content. It says that an act in our world is right *iff* that act has a descriptive property (maximizing happiness). But Simpler Utilitarianism does not say anything about rightness in other possible worlds. For this reason, Simpler Utilitarianism implies neither Right-Description Necessity nor supervenience. It is compatible with Simpler Utilitarianism that in some other possible world, actions are right *iff* they have some other descriptive property (such as minimizing utility or being done on a Tuesday), or that there is *no* descriptive predicate that is co-extensive with rightness in that world. Because Simpler Utilitarianism is not reduction friendly, it does not imply Right-Description Necessity, and Streumer's argument does not go through.[12]

Second, a more extreme example:

> *Completeness*: Every logically possible distribution of rightness over descriptive properties is realized in some possible world.

Completeness implies the falsity of Right-Description Necessity. Here is why. It is logically possible that the normative property of being a right action is co-extensive with the descriptive property of being an act done on a Tuesday; it is logically possible that the normative property of being a right action is co-extensive with the descriptive property of being an act done on a Wednesday.

Now, consider two possible worlds that are descriptively identical, and in which Bart Streumer buys a cup of coffee on some Wednesday. According to

---

12 But, as I will argue below, Simpler Utilitarianism is the kind of theory that makes w look plausible.

Completeness, Streumer's act of buying a coffee is wrong in one world (because done on Wednesday, not Tuesday), and right in the other world (because done on Wednesday).

So if Completeness is the correct moral theory, then Streumer's argument fails, because the correct moral theory does not imply Right-Description Necessity. Completeness also implies that s is false, because the two possible worlds are descriptively identical, but differ normatively.

These examples show how Streumer's argument fails more generally. The structure of the argument is that if we assume a simple moral theory such as Utilitarianism, then we see that Right-Description Necessity is true, and therefore by appeal to N that non-reductive realism is false. Then w tells us that the choice of Utilitarianism played no logical role, and thus that we can drop the assumption of that particular moral theory, yet keep the conclusion that non-reductive realism is false.

But if I am right, then the work in the argument is done not by the simplicity of Utilitarianism, but by its reduction friendliness. And this reduction friendliness *does* play a logical role in the argument: if the true moral theory is not reduction friendly, then non-reductive realism is not refuted. So the assumption that the true moral theory is reduction friendly cannot be dropped—and reduction friendliness implies s.

An anonymous reviewer has suggested the following ingenious response to my criticism: What if there is some constraint on what counts as a theory—or more specifically, on what counts as a moral theory? We might think that theories in general, or moral theories in particular, must be able to ground counterfactuals, or respect universalizability, for example. If there are such constraints, might they rule Simpler Utilitarianism and Completeness ineligible as theories, let alone as moral theories, let alone as possibly correct moral theories?

Though I agree that this is the best line of response for Streumer, here is a general argument that it cannot work. The response faces a dilemma.

First, suppose that there is such a constraint—for the sake of argument, a constraint on what counts as a moral theory. If the constraint restricts moral theories to the reduction-friendly (and so supervenience-implying) ones only, then Streumer's argument is not independent of supervenience after all. The argument relies on a truth that all moral theories imply supervenience. It covertly assumes s by relying on constraints that entail s.

On the other hand, if the constraint does not restrict moral theories to the reduction friendly—if some non-reduction-friendly moral theory meets the constraint—then my earlier criticism stands. Whatever the eligible but non-reduction-friendly moral theory in question, substitute it for Completeness in the

counterexample above. Because the theory is not reduction friendly, it does not imply Right-Description Necessity, and Streumer's argument does not get underway.

So restricting which (moral) theories are in play does not seem to be a promising line of response.

But Streumer could say: if the *correct* moral theory is simple and reduction friendly, then it need not matter that there are other possible moral theories that are neither simple nor reduction friendly. We just need to assume that a simple and reduction-friendly theory is correct, show that Right-Description Necessity follows, and then use w to generalize the conclusion (that non-reductive realism is false) to the other moral theories, whether they are reduction friendly or not. To close off this line of response, I will criticize w.

3. AN ERROR THEORY FOR W

Streumer's argument relies on w, which is:

> w: Whether normative properties are identical to descriptive properties cannot depend on which first-order normative view is correct.[13]

I will remain neutral on what kind of dependence is at stake in w. Implicitly, I have been arguing that w is false. Given some assumptions, including N, I showed that some first-order normative views (the reduction-friendly ones) imply that normative properties are identical to descriptive properties, but some other first-order normative views (the non-reduction-friendly ones) either imply that normative properties are *not* identical to descriptive properties or are silent about the matter.

In other words, I have argued against Streumer's claim that neither non-reductive realism nor its denial "contradicts any first-order normative view at all."[14] His master argument fails because it applies only to a restricted range of first-order theories.

In this section, I will show more explicitly that w is false, and explain why it seems so plausible. Let us again focus on the moral. What is a first-order moral theory? Roughly, it will tell us

(i) which objects have which moral properties, and which descriptive properties these objects have.[15]

13   Streumer, *Unbelievable Errors*, 31.
14   Streumer, *Unbelievable Errors*, 33.
15   Compare Streumer, *Unbelievable Errors*, 31.

For Streumer, answering (i) simply is—conceptually—what a moral theory does. He defends w by claiming that whether non-reductive realism is true instead

  (ii) "seems to depend on the nature of [moral] properties."[16]

So the argument for w is that (i) and (ii) are distinct questions; indeed, w arguably simply states that they *are* distinct. Clearly, the plausibility of w depends on keeping (i) and (ii) distinct.

  But if the correct criterion of property identity is n, then the distinction between them collapses. This is because (i)-type facts about whether all objects with moral rightness also, across all worlds, have some descriptive property can imply—according to n—a (ii)-type fact about the nature of moral rightness: whether moral rightness is identical to that descriptive property.

  If the (i)-type facts in question are modally strong, and tell us about how moral and descriptive properties are distributed across all possible worlds, then they can satisfy the antecedent of n. This is just what reduction-friendly first-order moral theories do, and why it is the reduction friendliness (and not the simplicity) of those theories that does the work in that part of Streumer's argument. w cannot be used to hold fixed (ii)-type facts about non-reductive realism while ranging across both reduction-friendly and non-reduction-friendly answers to (i). In slogan, w says that first-order theories and metaethical theories answer distinctive questions. But given n and modally strong first-order moral theories, the distinction breaks down.

  There are two main ways that Streumer could rescue w. First, he could say that n is false: even if our first-order moral theory says that rightness is necessarily co-extensive with a descriptive predicate, this does not imply anything about property identity. But of course then the conclusion of his argument—the falsity of non-reductive realism—would not follow, because n is a crucial premise in that argument.

  Second, he could restrict w in a way that is compatible with n:

Weak w: Whether some normative property is identical to a descriptive property cannot depend only on what things in this possible world have that normative property.

This claim is indeed plausible, and compatible with n. It says that, for example, whether non-reductive realism is true cannot depend only on whether Simpler Utilitarianism is correct. Weak w is compatible with n because the latter says that to imply facts about property identity, a moral theory must say something

16 Streumer, *Unbelievable Errors*, 31.

about the moral property across all possible worlds, not just this one or just a few of them.

But as we saw with that example, Simpler Utilitarianism is not modally strong enough to entail Right-Description Necessity. A first-order moral theory that only tells us about wrongness in this possible world is not sufficient for Streumer's argument to proceed.

But why does w seem plausible, even in the presence of N? Here is my diagnosis. First, Weak w is very plausible, and because of that we have failed to notice how implausible w is, at least in the presence of N and of first-order moral theories that include claims about all possible worlds. Second, w itself is very plausible if we are not attached to N—both non-reductive realists and reductive realists can accept w, but not at the same time as N.[17]

4. CONCLUSION

I have argued that Streumer's attempt at reduction without s fails. The simple moral theory argument might not explicitly depend on supervenience, but it succeeds only for reduction-friendly moral theories—and those theories entail the truth of supervenience.[18]

University of Reading
luke.elson@reading.ac.uk

REFERENCES

Dunaway, Billy. "Supervenience Arguments and Normative Non-Naturalism." *Philosophy and Phenomenological Research* 91, no. 3 (November 2015): 627–55.
Jackson, Frank. *From Metaphysics to Ethics.* Oxford: Oxford University Press, 1998.
Streumer, Bart. *Unbelievable Errors.* Oxford: Oxford University Press, 2017.
———. "Why There Really Are No Irreducibly Normative Properties." In *Thinking About Reasons: Themes from the Philosophy of Jonathan Dancy*, edited by

17 At least, as we have seen, not without constraints on what counts as a moral theory that restricts w to either the reduction friendly or to the non-reduction friendly.

David Bakhurst, Brad Hooker, and Margaret Little, 310–36. Oxford: Oxford University Press, 2013.

Väyrynen, Pekka. "The Supervenience Challenge to Non-Naturalism." In *The Routledge Handbook of Metaethics*, edited by Tristram McPherson and David Plunkett, 170–84. New York: Routledge, 2017.