Replies to critics

David Estlund

ONLINE
FIRST

⌂ Springer

Springer

Replies to critics

David Estlund[1]

Abstract I offer replies to critical comments on my book, *Utopophobia: On the Limits (If Any) of Political Philosophy*, in four pieces appearing in the same issue of this journal.

I'm grateful to Nic Southwood for organizing this symposium, and to him along with both Geoffs, Zosia, and David for their generous and challenging comments.[1] I have learned much from reading them, and from trying to adequately reply. Though a few topics recur in their comments, each piece pursues its own set of questions and criticisms with little overlap between their points. Except for cross-referencing a few places where the same part of my view is treated, it makes most sense to offer free-standing replies to each.

1 Reply to Brennan and Sayre-McCord

Brennan and Sayre-McCord (hereafter "the authors") describe and defend something they dub "real world" political philosophy. From that vantage point they mount a number of interrelated challenges to my book's arguments, positions, and methodology. I only have space to address some of these challenges. If there is a helpful big

---

[1] Thanks also to Chad Marxen for help with proofreading and editing.

✉ David Estlund
david_estlund@brown.edu

[1] Brown University, Providence, USA

picture to start with it might be this: Among other things, the authors put forward several familiar lines of objection to "ideal theory," or to idealistic normative political philosophy, arguments that I take myself to have anticipated and addressed. But, of course, the authors formulate them in terms that they take still to be compelling even having encountered my arguments, so it is valuable to have a chance to explain, if possible, how my responses are still responsive. What unifies the series of arguments and objections they offer is their roots in this recognizable "real world" political philosophy. By name, that's a hard thing to be against, so in my responses I will hope to show that it isn't really a fair name for any view that I criticize. But even that name does fly the flag for a traditional and familiar stance of anti-idealist thought that many readers will recognize, and it's useful for that reason. So, we might see my replies as a partial defense of my brand of idealistic theory against these remodeled and fortified versions of that traditional critique. As is common in these things, I will be pointing out in a number of cases that the objections may rest on a misunderstanding of my view, which is my fault and not theirs. Not surprisingly, I believe my view escapes the objections even in these refurbished formulations, and so much will hang on my clarifying, if I can, how my view goes with respect to the matters they mention, how it might differ from other idealistic targets, and how those differences make the crucial difference.

1.1 Not merely conditional

It is very common for skeptics about certain kinds of idealistic theorizing about justice, including the authors, to complain that it is the study of principles for people other than, and unlike us. We learn only what justice would require if we were different in certain ways. I begin by explaining how my view avoids the objection, which I'll call the "not for us" objection.

I argue that what I call, "prime justice," has a strong claim to being justice, even if it might be highly idealistic. Prime justice finds the principles of justice in what we all, alone and together, are required comprehensively to do. For example, we might be required to build and comply with certain institutions. This does not imply a requirement to build them, since that probably depends on whether we will in fact comply. By the same token, principles of justice, which are part of what is most comprehensively required of us, also do not require on their own. Again, that might depend on what else people actually do. Rather, they are part of what is required—which is a different kind of requiring force—and this point is at the very center of the book's argument. Contrary to the "not for us" objection, this does not mean that the principles of justice do not apply to us in the real imperfect world. They do, as part of what is required of us, which is surely one way for them to apply. It might seem, and it has often been suggested, that idealized principles of justice, whatever other interest they might have, only apply conditionally: if other idealistic conditions are met. But for these reasons that is not my position on prime justice.[2]

---

[2] The authors misinterpret my view on this score at one point, though their other main points are not deformed by this error. See, p. 4: "But he thinks that, with that point acknowledged, there is still important room in our thinking about such questions for the principles that would be the right principles

The authors have a distinct concern which is not to be conflated with that merely conditional interpretation of the requirements of justice, namely, that even if those principles are part of what applies to us in real conditions, their value or appropriateness remains contingent on the other conditions. That is indeed the case, with a qualification to be added in a moment. If there is a difficulty in this fact, it is not that the requirements don't apply to us, as we have just seen. But it also can't be that while they apply to us, they do so inappropriately since they are fitting only in unrealistic conditions unlike ours. There is no such inappropriateness since they do not have requiring force given that the other conditions will not be met—which we assume they will not be.

So what might the objection be? It might be this: even granting those two points, the resulting principles will offer no practical guidance, since the other ideal conditions will be missing. That may or may not be so, and I will come back to it. That would certainly be a kind of limitation, but a common kind of limitation in requirements of many kinds. Even the requirements of prudence in my financial planning would include requirements that apply only because I have not been fully financially prudent after all. Given that, I should perhaps start saving more than I have been. But there would also be that more primary requirement to save more from the beginning, which is not conditioned on failing to meet any of the others. The limitation is that that more primary principle doesn't itself say what prudentially to do if you will be financially imprudent in some respects. But that seems to cast no doubt on its cogency. The requirements of justice, on my prime justice approach, are part of such a more comprehensive requirement starting without any concessions to moral failure. Then, given such failures, there will also be other "concessive" requirements, and prime justice itself is not addressed to that question. The limitation is no defect. To emphasize: this limitation cannot correctly be framed as its not addressing what we ought to do in the real world. It does. It is part of what we can and ought to do in the real world.

The authors make a further related argument. They question my argument for the claim that prime justice avoids a certain arbitrariness problem facing any alternative—and so any concessive—account of what justice requires. They write,

> [T]o the extent the question is which requirements apply to us, assuming away the conditions we actually face is no less arbitrary than taking them into account. Concerns about arbitrariness do not disappear, but they do face any view that privileges some conditions over others (as ideal and real-world theories alike do).[3]

As I have said, both conditional and prime requirements apply to us. That's not the question that "prime" is meant as an answer to. The question is what conditions set the requirements of justice. The authors seem to suggest that there would be a way

---

Footnote 2 continued

of justice *if* people were free from human failings (such as weakness of will, unjustified self-interest, and moral indifference)." (emphasis added).

[3] P. 15.

to do that in terms of "conditions we actually face." That might mean that all there is to justice is what we are required to do given what else we and others will or not do. No doubt, there is that question, but is that all there is to justice? It's hard to believe, since it would imply that if our basic social institutions are exactly the ones we should have if we were faced with intractable, widespread, vicious racism, then our basic social structure would be, simply, just—or at least our basic institutions meet all that there is to the idea of requirements of a just basic structure.[4] One could accept that jarring view, and maybe the authors do. But why accept it? Perhaps because they believe there is no defensible alternative account of the principles of social justice. But, as I have partly explained, and will further explain below, I do not see any serious difficulty raised by the authors for my alternative account according to which justice is prime justice.

1.2 Not bi-modal

The domain of the theory of justice might include any or all of three sorts of evaluations, to which we can give the following names:

(a) *Partition*: T is just, U is unjust.
(b) *Comparative*: X is more just, or less unjust, than Y
(c) *Distance*: L is nearly just, or extremely unjust.[5]

A theory of justice according to which there is only "partition" evaluation would be crude, as they say.[6] It would have the consequence that amongst unjust societies, none is more unjust than others. I will follow the authors in calling this a *bi-modal* view of justice. They attribute this approach to me,[7] but I reject it in the book,[8] so I put aside their arguments against it.

---

[4] See my reply, in Chapter 1, section 7, to Enoch's paper, "Against Utopianism: Noncompliance and Multiple Agents," for more on this point. *Philosophers' Imprint*, 18:16 September 2018.

[5] Some kinds of distance relations are different from these, such as A is much more unjust than B. Those do not depend on partition information, as do "nearly just," or "extremely unjust," which is the crucial issue in my argument below.

[6] I argue in the book, however, that justice could still turn out, in principle, to be like that. See Chapter 13, section 3, and, for more detailed argument, "Just and Juster," *op. cit.*

[7] For example, p. 4: "One can see immediately why a thoroughly 'non-concessive approach' is going to be hopelessly uninformative for those – all of us – faced with circumstances in which full justice cannot be achieved."; p. 5: "Any theory that fails to have the resources to address such concerns is, we think, unacceptable." Also, p. 5: note 4; p. 6: "A 'non-concessional approach' is equipped only with the ideal requirements in each case."; p. 14: They criticize an approach, presumably attributed to me, where one "can specify an ideal of justice but cannot tell in any situation whether one or another actual situation is more or less just"; p. 12, "…but it requires not stopping there …"; p. 17: "an account that appeals only to the principles appropriate to the limiting case would not constitute an appropriately systematic theory of justice."

[8] In Chapter 13, section 3, "Critique of Pure Comparativism," my argument is that comparisons of "juster" are not enough, and that we need to appeal to our partition-implying judgments in any effort to provide the rankings we need for social choice. However, in that section I also point out that on some egalitarian theories of justice it is not clear that there are any comparisons other than those settled by the bi-modal partition. But I am not endorsing a bi-modal approach, only pointing out that it isn't

Also, however, a theory that went beyond the bi-modal view by including comparisons such as "more unjust," but which did not accept any partition between just and unjust, would face difficulties too. The authors express what might be sympathy for such a view, though it is not made clear.[9] Call that a no-partition, or purely ordinal comparativist approach to justice. One difficulty would simply be that on that view there is no such thing as a just society, or an unjust one, except perhaps in a contextual way—the way in which some objects are, simply, heavy (a 100 lb cat) or light (a 200 lb car); that is, not really, full stop, heavy or light; not really just or unjust.

Another difficulty for the wholly comparative approach is more complex, and is the topic of section 3 of Chapter 13.[10] I sketch it briefly in section 4 of my reply to Enoch in this volume. The authors do not discuss it, so I will let that, and my response to Enoch's points about it, suffice.

1.3 Feasibility and circumstances of justice

An idealizing account such as prime justice is often accused of illicitly moving beyond the very conditions that give justice any point—what Rawls, whose account follows Hume, called the "circumstances of justice."[11] It might seem that a world in which everyone is morally flawless leaves those circumstances behind, but I argue in Chapter 3 that on a proper understanding of circumstances of justice that is not so.

The authors are also concerned on different grounds that I neglect the circumstances of justice. They submit that among the "the circumstances that give justice its point," are facts about, "what is feasible—specifically, considerations concerning what people, generally, might *willingly* do."[12] There is no question that those facts, along with all the others, bear importantly on what ought to be done under whatever circumstances we are in. Of course, that does not make every kind of circumstance a "circumstance of justice," so what is their more specific proposal here? They may mean that, contrary to my own view, the robust facts about what people are unwilling to do are a necessary condition for the circumstances of justice to obtain—without those facts, there would be no question of justice. In reply, for one thing, that may seem to imply—as a conceptual matter—that justice could not be infeasible in that way. That is far from obvious, I think, and I argue throughout the book that it is not correct. I note in this short reply only that it is not a position they offer support for.

Footnote 8 continued

antecedently more obvious that there must be comparisons whether or not there are partitions, than it is that there must be partitions whether or not there are (further) comparisons.

[9] I find such potential sympathy elaborated in their section V.

[10] Also, somewhat more fully, in "Just and Juster," *Oxford Studies in Political Philosophy, Volume 2,* David Sobel, Peter Vallentyne, and Steven Wall, eds., Oxford University Press (2016). I touch on it briefly in my reply to Enoch, sec. 8.

[11] John Rawls, *A Theory of Justice*, (Harvard Belknap, 1971) sec. 22.

[12] P. 15, emphasis in original, but not pertinent here.

1.4 Complacency

Since the authors spend some time on it, let me just briefly respond to their concern that I seem to regard it to be, or to induce, a kind of complacency, to limit the theory of justice to non-ideal or real-world questions. I do argue that one value of highly idealistic theory is that it can inspire and induce higher achievement. They point out that it can also have opposite effects like resignation or demoralization.[13] The dialectical situation is somewhat puzzling. When I present the ways in which highly idealistic theorizing about justice might be valuable (Ch. 13, sec. 2), I begin by admitting those dangers, and then point out that there is also something to be said on the other side: the dangers of not thinking idealistically enough. The authors, in effect, say, "Yes, but there may also be dangers of thinking so idealistically." Yes, but granting that point is exactly what sets up my claims, to add to the mix, about dangers of excessive realism. This is not to say that they do not add anything to the downside point, but in the main, I simply mean to grant it with appropriate nuance and qualifications.

2 Onus of proof, and decisiveness

In Sect. 2, the authors object to what they present as my, a) placing the onus of proof on the anti-idealistic views I oppose, and b) demanding "decisive arguments" from my opponents.[14] To begin to address this, we can look at one of the book's central points in this way: I can neither find nor devise any strong case for the common view that an idealistic theory of justice is, as such, defective. Call this aspect of what is often thought of as "realism," its "anti-idealism." The onus this places on the anti-idealist view is only to make a better case for any of the criticized arguments, not a decisive case, if they rely on them. And it is not as if there is any lesser onus on me, which I do my best to discharge—to defend that critical thesis (which is not my only thesis).

Suppose I were to fully defeat the extant arguments for a position such as this anti-idealism—call that *negative* argument: arguments against arguments for. I admit that is not directly an argument against the position.[15] (I will point out below that I have offered more than a negative argument of that kind.) The authors are not prepared to grant that I have defeated the arguments, of course. They do grant, for the sake of argument I suppose, that I have shown (where it needed showing anyway) that none of the arguments for the view I oppose is "decisive" in its favor, which I understand to mean (something like) sufficient to warrant belief in the view.[16] They point out, however, that a number of non-decisive considerations could indeed add up to at least a strong on-balance case, and they believe this is so

---

[13] P. 9.

[14] P. 3.

[15] Thanks to Nic Southwood for encouraging me to address this point.

[16] P. 3.

for anti-idealism. They might have added that it could, in principle, even add up to a decisive case, so far as any of my negative arguments go. Fair enough. If my critical study should prompt others to lay out such an on-balance case as might remain, and show it to be a strong one, I would be eager to see it and pleased to have helped prompt it. While the authors seem to indicate that they think this could be done, they obviously do not take on such a project here. Advocates of anti-idealism may find it reasonable even without presenting such a case, and I place no onus on them. But many have indicated various reasons they take to support anti-idealism, and I have criticized all the best ones I could find or devise.

Second, however, that negative form of argument is not all I hope (unrealistically, no doubt) to accomplish. I also argue directly against a realist, or what they call a "real-world" view. I have been speaking above of "the view" I mean to oppose, but that has been for simplicity. Realist or anti-idealist views come in a variety of forms, many of which have previously been run together. Also, whether an argument is (what we are calling) negative or direct depends partly on how the background dialectical situation is understood, often implicitly, so it is a matter of judgment which counts as negative and which direct. Finally, I don't claim that my direct arguments are decisive, individually or jointly. The point here is only that these are not merely arguments against arguments for a realistic position, which might (as I granted) leave those arguments for a realistic view as having some weight or merit. These are arguments against the position itself in various forms— directly, whether or not decisively. It may be worth actually listing them, partly because this gives readers a reminder or an introduction to yet more of the book's claims. Here, then, are some candidates for direct arguments of mine against realist or real-world views:

- *Complacent realism:* Justice is however things are or will be.
  No one defends it, but it is helpful to have it here as part of the array.[17]
- *No partition*: There is only "more just" or "less just," but not "just" or "unjust."[18]
  I have sketched earlier in this reply the book's direct argument against that approach, associated especially with Sen, on epistemological grounds.
- *Motivational feasibility:* If some arrangement would not be willingly complied with by realistic people, then justice does not require it.
  While one of my main strategies is to refute the best arguments I can devise in favor of it, I also argue directly against it for its implausible implications. For example, if people were expected to be, in large measure, ineluctably cruel in ways that feasible institutions were forced to accommodate, this view implies that social justice does not require more.[19]
- *Anti-moralism, several varieties*: Political normativity is, in some way or other, not a species of moral normativity.
  I argue directly as well as negatively against this in Chapter 3.

---

[17] I discuss it in *Utop.* at pp. 5ff.

[18] Ch. 13, sec. 3, pp. 261-69.

[19] I argue against it throughout the three chapters that make up Part II of *Utop.*

- *Unknowability*: What ideal justice requires is impossible to know, therefore justice is not ideal.[20]

    No one would put it that way, but it is often the apparent implication of common discussions. It is a *non sequitur*. Moreover, I argue that even if the specific institutions of ideal justice couldn't be known, it doesn't follow that substantial principles of ideal justice could not be known. Is this still only a negative argument: an argument against an argument for? Try it this way: the interlocutor might be expected to grant that many moral principles can be (sufficiently) known. Since (as I argue) principles of justice are best seen as moral principles, we should, barring a strong case against this, accept that they are approximately as knowable. That's a direct argument—not decisive but direct—against the claim that they can't be known.

- *Practicalism:* There is little or no value in studying or understanding anything unless this has practical implications.[21]

    I argue directly (as well as negatively) against this general position at length. Insofar as practicalism is used as a premise in support or anti-idealism, then my arguments are "negative," that is, against that argument.

- *Circumstances of justice:* Justice no longer applies if no one is morally deficient.

    I devote Chapter 3 to arguing directly against this.

Without more space, I must leave unaddressed several other lines of concern raised by Brennan and Sayre-McCord, but I hope I have hit several of the more important ones.

3 Reply to Stemplowska

A central line of argument in my book contends that people are able to do many of the things that idealistic theories of justice are said to require of them, even the ones they can't bring themselves to do, such as being less self-centered, less partial to associates, less lazy, and so on depending on the idealistic theory of justice in question. Thus, I claim, there is no inability present to block the requirement. If someone thinks those are inabilities, their best basis seems to me to be the thought that in those cases even people who set out to do those things often fail to do them, and surely that shows they are unable. I develop the best form of that "conditional account" of ability I can, and try to show that it would not count those as cases of inability after all. Zofia Stemplowska's piece is ambitious, aiming both to critique that account of ability, to improve on it with her own "incentives account" of feasibility, and to argue that the improved account better suits my opponents' purposes.

To anticipate: I will question Stemplowska's charge that the conditional account is defective on account of needing to add a set of exceptions, and also point out that

---

[20] Chapter 10, section 7, Pp. 200ff.

[21] This is the subject of Chapter 16.

her incentives account has the very same form in any case. If her incentive account better accounted for strong intuitive judgments of ability, as she argues, then it might be the better account, but I will resist the suggestion that it does. Even if I'm wrong about that, though, I will argue that, so far as her treatment goes here, the incentive account would appear to work as well for my purposes in the book as the (ostensibly defeated) conditional account I propose. Stemplowska speaks of feasibility, while I speak of ability. It might seem as though we are arguing past each other, but we are not. She tells us that for her purposes ability and feasibility are meant to be two words for the same topic, though she speaks mostly in the terms of feasibility.[22] It can be difficult, but it's crucial for our purposes, not to use the term in a way that invites equivocation between the idea of ability and other things for which feasibility language is at least as commonly used but which are *not about ability*. In its negative form, "infeasible," it sometimes invokes genuine constraints, thereby canceling ability. But infeasibility often refers only to self-imposed criteria, as when someone says that 8 pm dinner won't be feasible since they plan to catch the 9 pm train, or in reference to self-imposed individual or institutional budget limits. When the word "feasible" is stipulated to mean "within the agent's ability," I don't myself see any advantage over the terminology of ability, and there is an important disadvantage—ambiguity. This danger can be avoided if we keep in mind that, by her own stipulation we are free to substitute the terminology of being able, or, synonymously, what a person can do. And, anyway, we rarely, if ever, find people wishing it were "feasible" for them to tell jokes well, or lift more weight, or fix their own plumbing. But many people wish they were *able* to do those things. So, in the following, I will keep before our minds that the question in the individual case is whether we are being given a plausible account of what a person or set of persons is *able* to do—what she or they *can* do—and I think that will make some difference.

3.1 What does it make sense to deliberate about?

Stemplowska endorses the following as indicating the "functional role" of the concept of feasibility:

> *The deliberation criterion:*
> An account of what is feasible ought to align with the acts that it makes sense to deliberate about.[23]

To see whether this criterion divides cases up the way she hopes, consider two examples of mine that she discusses. First:

> *Messy Bill*
> Bill is too lazy to be able to bring himself to take his trash to the curb.[24]

---

[22] p. 2.

[23] Stemplowska loosely borrows this idea from an unpublished paper by Nic Southwood, "Feasibility as Deliberative Jurisdiction." I confine my treatment to Stemplowska's statement and use of the idea.

[24] I introduce the example in "Human Nature and the Limits (If Any) of Political Philosophy," in *Philosophy & Public Affairs,* 39 (3), 2011, and in *Utop* at p. 28.

Perhaps cases like this are among those that Stemplowska wants her account, unlike mine, to count as inabilities. In any case, some do take that view. They would presumably also want to say that it does not make sense for him to deliberate about it. On that latter point, I am happy to agree, it probably doesn't. The reason must be that he knows that he won't, in the end, do it, so why bother deliberating? But next consider my example of,

> *Dancing like a chicken*
> I will never dance like a chicken while giving a talk.

I know that, were I to deliberate about dancing like a chicken while lecturing, and set out to do it, I would not, in the end do it. But then consider parallel reasoning in the Messy Bill case, and employing the deliberation criterion. On Stemplowska's view this should show both that it doesn't make sense for me to deliberate about it, and that it is infeasible for me—outside my ability. But dancing like a chicken while giving a talk *is* feasible for me in Stemplowska's sense—I am able to do it. It's easy. I will never prove it, but she takes my word for it.

Suppose, instead, we said that I could, indeed, sensibly deliberate so as to reconsider my standing decision never to so dance. (That's not clearly the same thing as deliberating about whether to so dance, but let that pass.) In that case, the deliberation criterion might travel with ability after all. Still, then Messy Bill could also deliberate about whether to be less lazy. And citizens could deliberate about whether to give less weight to their self-interest, and so on, and those things would turn out to be feasible for them, contrary to Stemplowska's aims as I understand them. So, I don't see that the deliberation criterion would support her goal of counting more of such cases as inabilities than my account does.

In any case, I would add that I do not find the deliberation criterion to be plausible as a constraint on feasibility where we recall that is meant to be the same as ability. Even if, morally speaking, "'ought' implies 'can'," (about which more in my reply to Southwood) Messy Bill is surely not off the hook morally for lazily leaving his trash in the yard. He is still required to deal with his garbage—he wrongs his neighbors otherwise—even if it does not make sense for him to deliberate about it (and so even if it does cancel some non-moral ought, such as a "deliberative ought.")[25] If that is granted to me, then it's not plausible that he is unable to do it (i.e., that it is infeasible for him).

3.2 Sleeping and screaming

Stemplowska offers two examples intended to solidify the intuition that there are genuine motivation-based inabilities[26]:

> A…typical case [of motivational inability] is that of an agent not being able to stay awake once they have been awake for a sufficiently long time. Famous

---

[25] For more on this see my reply to Southwood.

[26] I leave aside her example about blocked synapses. It is not explained how that is an obvious case of motivational inability, and I won't speculate about what fuller story she has in mind.

philosophical examples include Susan Wolf's case of a woman confronted with an attacker who finds herself paralysed and unable to scream for help.[27]

These are pointedly not phobias or addictions, which are among the cases that I allow might be inabilities. These two are also meant, apparently, to be cases that my account, to its detriment, will not count as inabilities.

I don't think either case presents difficulty for my account, for the following reasons. The sleep case does count as a case of inability on my account, so it is not a counterexample. That's because, if I have been awake for sufficiently long, were I to try, and not give up prematurely, to stay awake for another 15 min, I would fail. My account then implies that I am unable to do so, just as she thinks a good account should.

It may be that Stemplowksa would hold that when the time comes, I am overwhelmed by the desire to sleep, and so I give up trying to stay awake. Then my account would imply that I remain able to stay awake, but unwilling. I'm not at all sure about that description of the facts, or the counterfactuals, but my account has little at stake. In the broader argument of the book, I am arguing against the view that such common human motivations as financial self-favoring, partiality toward loved ones, aversion to burdensome work, and so on, block requirements to act otherwise.[28] These are not like the inexorable descent into sleep. Even if the inability to stay awake were a motivational inability, which I do not concede, that would lend no support to the suggestion that those other cases are. I could, for the sake of argument only, allow that it is a motivation-based inability, along with some phobias and addictions. (Stemplowska takes this need to include exceptions to the central criterion as a defect in my conditional account, but we will see that her incentives account has the very same form.)

I think the screaming case (in the context now of ability, not Wolf's context of responsibility) is impossible to judge for sure without a richer description. If the woman tries to scream and doesn't give up prematurely, but still does not scream, then my account will agree with Stemplowska that it would be a case of inability. And if, on a different reading, "paralysed" indicates such things as that certain muscles in the person's throat will not respond—which might be the case in such a terrifying scenario—then it is again inability on the conditional account. However, if, instead, the person is not "paralysed" in that way, it may be that she does not form the intention to scream. She might be frantically considering such a thing, wanting to scream, but also wanting not to incur further danger. Her mind is racing, trying so far as possible to assess dangers, including those of screaming, feeling some impulse to scream, so far holding it back, and, as a result, remaining undecided, for some period of time, whether to scream or not. In that description,

---

[27] Stemplowska at p. 3, Susan Wolf, (1990). *Freedom Within Reason.* Oxford University Press. at p. 99. Stemplowska does suggest that at least in some of the cases she describes, though it isn't perfectly clear which ones, my conditional account differs from the incentives account, and the latter gives more plausible results. I believe my discussion will cover all the cases she might take to have that feature.

[28] As I wrote, "My aim is not to decide precisely where the line falls, as important as that question is." For more, see pp. 91-92 and 100, in *Utop.*

"paralysed" would be an exaggeration, and notably so, since that term unfairly connotes inability all by itself.

On my account that would not be a case of her being unable to scream. Granted, some might intuitively still wish to describe it as a case of inability. But inability to do what? We might well say that she can't, in that hellacious moment, *decide* whether to scream, but that wouldn't mean she can't scream. In ordinary conversation we might sometimes talk as if not being able to decide whether to do something entails that you are unable to do it. But those turns of phrase don't show much philosophically, I think. We also say, "I'm afraid I can't accept that refereeing request." It's not exactly a lie, but it is not literally true: I am not unable. Granted, because of those features of ordinary language, it is, admittedly, not simply *obvious* that this last specification of the screaming case isn't one of inability after all. But the question between us, I take it, is not whether it is simply obvious, but whether my conditional account has a deeply implausible implication in such a case. I don't believe that the sleeping or screaming examples show that.

3.3 The incentives account

Stemplowska presents an alternative to what she calls, and I have been calling, the "conditional account" of ability (which she calls "feasibility"), which she has taxed both with a) a *formal objection*: having to include unexplained exceptions, which is seemingly ad hoc, and b) a *substantive objection:* still counting too many cases as abilities. I develop that account because I take its general idea to lie behind the claim that Messy Bill is not able to take his trash to the curb. For our purposes we can speak of it as my account, to compete with Stemplowska's "incentives account." My conditional account says,

> Agent S is able to do act Φ if, and only if, were S to try, without giving up prematurely, to Φ, S would succeed in Φ-ing.[29]

Stemplowska's incentives account goes basically as follows, to be revised in steps we will see:

> Action Φ is feasible if there is an incentive I such that, given I, [agent] X is likely to do Φ.[30]

Her account is also conditional, so I will refer to "*my* conditional account," to avoid confusion without entirely changing the nomenclature. Let's begin by considering what the advantage is meant to be in the incentives account, and then some problems for it.

Clearly, the two accounts place different sets of actions within the agent's ability. Stemplowska cites the sleep and scream cases as examples that she seems to believe involve obvious inabilities, and counted as such by the incentives account, but not

---

[29] This is a composite from Chapter 5 of *Utop.* to have the best formulation for comparison with Stemplowska's alternative.

[30] Formulation (1) at p. 5.

by my conditional account. I have explained what my account seems to imply about those cases, and these implications are implausible. But consider, now, some cases that the incentive account has difficulty with. This first example is her own. It is only a difficulty for her provisional formulation, which she then revises to avoid the difficulty:

> *The Unwilling Murderer*
> He will, for moral reasons, not murder any innocent person, and no incentive could get him to do so.

As she says, this fully uncompromising case might be rare. But we agree that this person remains able to murder in many circumstances even though there is no possible effective incentive, thus it is a counterexample to a simple version of an incentives account. Her final version avoids the unwilling murderer case by building in a clause specifically to avoid it—in effect: "unless the reason there is no incentive is the agent's moral objection to the action." So (if I'm understanding her), the final version apparently holds that, in my words,

> *The Incentives Account*
> An action is feasible for a person if and only if[31] there is some incentive, given which, at least unless the agent sees the action as wrong, the agent is likely to perform the action.[32]

This could be accused of being *ad hoc*, (for what that's worth, more below) but if this is the one and only proviso needed then perhaps the spirit of the incentive account would be mostly preserved.

But I doubt that this single proviso is enough, due to cases such as the following: I have no moral objection to dancing like a chicken in public, and yet I am dead-set against ever doing such an embarrassing thing no matter what would be gained by it.[33] But this sounds, to my ears, like a *non sequitur*: "I am resolved not do it though the heavens may fall, therefore it is not within my ability to do it." (Martin Luther is said to have testified that he could "do no other," but either he didn't mean he was unable to act otherwise, or it wasn't true.)[34] With that implication, I find even her final incentives account to be implausible.

This brings us to an important point. Suppose Stemplowska altered the account yet again, with a further clause about cases of resolute but non-moral convictions. It

---

[31] Her use of the murderer example shows that she means the provisional simple account to say (despite her initial formulation ("if")) that the possibility of an effective incentive is necessary for the person's counting as able. And so her own formulation using "if," suggests that she evidently means it to be both necessary and sufficient. I will build that in, but I don't think anything here hangs on that interpretive question.

[32] "(2): Action Φ is feasible if there is an incentive I – or had the agent X not seen Φ as wrong there would be I – such that, given I, X is likely to Φ."

[33] You might think this is implausible. I don't see that that would matter here.

[34] I recognize that it is possible to hold that expressions such as these and several others I mention determine the meaning of "able" and "can." My objection to that view rests on the unstated (until now) premise that the following is not self-contradictory: "I can (am able to) do x, but I am strongly, even decisively, disposed—for moral or other reasons or causes—not to do x.".

might begin to feel especially ad hoc at this point, but I am in no position to complain about that. After all, as Stemplowska says, I offer the conditional analysis as only part of the story, while admitting that there may be exceptions including phobias, addictions, and other things. My account, like hers, employs a general criterion, and some exceptions. Our accounts are the same in this way. It is either a serious problem for both, or for neither.

I think it is a serious problem for neither account. What's another case where a generally good analysis or definition is subject to a special class of exceptions? Suppose I said that a compass is by definition a device that points to the magnetic north even as the orientation of the device, with respect to north, is varied—except for certain exceptions: magnetic geological formations, certain metal objects nearby, or certain anomalies when used in the southern hemisphere, etc. If that were thought ad hoc, the way to "fix" the definition would be to more accurately state in general terms what the needle on a compass will tend to point to. But that needn't even mention north, and so it's not describing what it is to be a compass. Both Stemplowska's account and mine take a certain thing to be conceptually central (roughly, success when trying, and success when incentivized), and then admit exceptions. These are roughly like the definition of a compass where pointing north is the central, but not exceptionless thing.

3.4 Which button?

As I sketched in the precís, I present, as problematic for my own larger aims (though I try to resolve it in due course), a case in which there is, for some, an intuitive sense that two doctors could and should save a patient, but they do not do so, and yet no one acts wrongly:

> *Slice and Patch*
> These two doctors are needed at noon to perform and close some important surgery on an otherwise dying patient. As it happens, neither would do her part even if the other one were to do so. Each knows their part, and how to perform it, as well as everything else that might be relevant.[35]

I say that it is intuitive, but a difficult thing to vindicate philosophically, that they could and should together save the patient, and that the patient is gravely wronged by their not doing so. What's puzzling is that Slice does the right thing by not cutting, since Patch will not stitch. And Patch does the right thing in not stitching, because Slice will not make an incision. How can the patient be wronged if there is no culprit?

Stemplowska seeks to dispel the intuition that they are even *able* to save the patient, by denying that each agent is able to do her part. The key is how to conceive

---

[35] Utop at p. 33, or op. cit. I paraphrase here.

of their part. She presents what she takes to be a relevantly similar case where, instead, we do not think each is able to do their part. Paraphrasing, consider:

> *Buttons*
> Two strangers are located in separate rooms, with no communication, each with 1000 consecutively numbered buttons. To save the life of a third party, each must press only the button of the same number as the other agent, though neither has any idea which the other will push.[36]

They are not able to push the same numbered buttons as each other and so not able save the third party. If the Slice and Patch case were just like this one, then this should dispel any sense that they were able to save the patient, and thus the sense that they should have.

In reply, the doctor case is crucially different from the button case in that each doctor knows what is needed from her (the proper cutting from one, stitching from the other), and knows how to do it. This is not so in the button case, since neither can possibly know which button to push to do their part. And there is no special button. Even if each person were to try to do her action—what is needed from her alone—almost certainly neither would succeed, and the combined saving action would not be performed. By contrast if each of Slice and Patch tried to do what is needed from her, each would succeed at doing her part. There is a special place on the patient to cut and stitch, and both doctors know what it is. (We could just as well let the example require, fancifully, that each simply presses on that special numbered location, like a special button.) So, on the basis of that central difference, my account agrees with Stemplowska that the button pushers[37] are unable to save the patient, but holds that, by contrast, Slice and Patch are able to do so.

There is a further point: Stemplowska may think that, even if there is no group agent in such a case, there is not even plural action—a case of saving the patient together—unless it is in some sense intentional, and so there is no such action they are able to do. Stemplowska might think that my account cannot meet this criterion, because she thinks, contrary to what I hold, that neither Slice nor Patch knows how to do her part under the circumstances. Let me grant the intention criterion for the sake of argument. In particular, suppose that each individual (or enough of them) must act with, at least, the intention of doing one's part in such a plural action. I think that my account can meet such a criterion, and that while Stemplowska implies a stronger criterion, there is no adequate basis for the stronger one.

The question is whether the individuals can intend their parts in a way that is sufficient for the plural action, should it occur, to intuitively qualify as intentional (even though there is no group agent to do the intending). Can Slice act with

---

[36] I paraphrase, for brevity (Page number not available). Incidentally I think there is a strong chance each would push button number 1, its being by far the most salient possible coordination point. But put that aside.

[37] My response is the same to her example of simultaneous clappers.

sufficient intention? Slice knows what "her action"[38] is: it is to cut in the certain way in the special place. We agree that she knows how to do that, so she can intend it under that description. She also knows what her part is in what they need together to do: exactly the same description. So she can intend her cutting under the description: "my part in what we need to do together."

Stemplowska denies that Slice "knows how" to do her part, but so far as I can see, there is no relevant knowledge deficit on her part, or on Patch's part—neither a lack of "knowing how," nor a lack of "knowing that." Consider a different example. If we are supposed to play a duet, and I know you won't be there to do it with me, there is no relevant knowledge I lack, and so none that can be leveraged into an inability. I know how to play the music. I know that it is my part of a duet. I am in a position to intend it as my part of a duet, if circumstances call for this. If you will not be there, then they do not call for it. But I'm ready: I have the knowledge and the ability.

Stemplowska emphasizes that (translating into this example of mine) nothing I am able to do in those circumstances would count as playing my part in what will be a duet. So, knowing all that, I cannot intend *to play along with you*. Call that unavailable intention, the *rich intention*. I could at best intend to play my part without you. So consider this possible criterion on a plural action, which may be roughly what Stemplowska is assuming:

> *Together intending*
> For the act of playing a duet to occur, each must be in a position to richly intend to do their part, that is to intend it as doing one's part along with others doing theirs.

This involves a controversial conception of intending, since it seems to include within the scope of one person's intention the actions of others, which are not under the agent's control. I presume that one can't intend, on a sunny day, to play in the rain, so how can one intend that we both play our parts of the duet, or our parts in the life-saving-together? But this is unsettled in the literature, so I won't rely on denying it.[39]

Next, though, why think that rich kind of intention is needed here? Suppose that I were to play my musical part alone, but with the slim crazy hope that you will show up and play too. I'm in no position to intend (even if together-intending is a thing) that you show up, not only because it's not in my control but also because I know it's extremely unlikely in any case. I can't plausibly intend my action under the description, "playing along with you." But now, add to the story, that you have misunderstood the venue and are just on the other side of the curtain, ready to go. You, too, can intend to play your part of the duet, but not under the rich description. So far so good for Stemplowska's points. Finally, suppose that I play, but lo and behold, you come in at the first note, though both of us play without the rich intention.

---

[38] See Stemplowska, "While each doctor knows how to perform her action, recall…" (Page number not available). See Zofia Stemplowska, "The incentives account of feasibility," *Philosophical Studies*. https://doi.org/10.1007/s11098-020-01530-y.

[39] For some discussion of this debate by Michael Bratman, David Velleman, and others, see, Roth, Abraham Sesshu, "Shared Agency", *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), Edward N. Zalta (ed.).

(It helps if we suppose that neither notices the other's playing, so imagine they are in a studio with headphones, maybe synchronized by a common metronome.)

Let's allow that if my playing and yours was not in any way owed to some hope of each person's that the other would, implausibly, play, then this could not amount to the resulting set of actions being intentional as a plural action. But, in the example, each does play on the slim hope that the other will play too. Stemplowska's view seems to imply that the duet is not played intentionally as a duet. This is implausible, and I think many cases can be marshaled to support this. Quarterback Eli Manning throws the football with the slim hope that Odell Beckham Jr. will catch it. Beckham runs like crazy into the end-zone on the slim hope that Manning could get the ball to him. But it works—touchdown.[40] It would be strange to suppose that the touchdown is unintentional.

What about Slice and Patch? Suppose each actually does her part strongly expecting that the other won't do hers, but on the off chance that she might. Each would be blameworthy, but they would save the patient together, and while it would be surprising I doubt that we would think that they unintentionally saved him. At least, I can't see what there is to be said for any criterion of intentionality, such as intending *together*, stronger than one that would count the cases of the quarterback and of the doctors. Indeed, if the button pushers happened to find the right combination by way of each trying, that would be a surprising but intentional aversion of a nuclear disaster.

4 Reply to Southwood

4.1 Big picture

Southwood's comments are admirably elaborate, by which I mean that they do not gloss over valuable distinctions or alternative lines of argument that should be considered. My reply, then, mostly follows along. The trees are very interesting to me, but readers might appreciate a quick look at the forest before heading in. I try that in this first section. I begin with a reminder about my own argument, and then we'll see where Southwood proposes to intervene. At the heart of the book is my argument that while justice can't require what we can't do, it could require what we can't bring ourselves to do. Nothing is more important for my purposes than the fate of that line of argument, and this is Southwood's focus.

I argue that many times anti-realist sentiments illegitimately get plausibility from something quite plausible, namely that what can't be done is never morally required. (I will call it "ought implies can" (OIC) for short, but I will be referring to the longer phrase.) That does indeed have a strong pull. It's disputed, and I don't know whether it's true, but let's consider how anti-idealism might try to benefit from it if it were. Some conceptions of social justice would depend on behavior that we might agree

---

[40] For this unlikely scenario in a real game, see:
https://www.youtube.com/watch?v=818_M8gOnqQ.

people can't bring themselves to do, such as great selflessness, or impartiality, or civic engagement. That occurrence of "can't" might seem to let those conceptions—the ones that purport to require those things—be defeated by OIC. I respond by arguing in some detail that what a person is able to do might, and normally does, outstrip what they can "bring themselves to do," as that latter idea is being used in those arguments, and so OIC does not have the alleged application.

Southwood's critique does not rely on OIC in that moral interpretation, and indeed he rests his objection on my willingness to leave it as an open possibility. Instead, he distinguishes several ways in which "ought" can function, and argues that none of them will suit my purposes. He will grant that "ought implies can," understood in what he calls a "deliberative" sense, but those oughts, he thinks are defeated not only by "can't" but "can't bring oneself to." That would, he argues, deprive me of my big distinction in the dialectic above. But, unlike the deliberative interpretation, the moral interpretation of OIC—that is, one is never morally required to do what one cannot do—is implausible, he thinks. So when I grant OIC for the sake of argument I can't reasonably allow the possibility of moral OIC. Hence, I'm stuck with deliberative ought, but that suits the anti-idealist argument I purported to refute.

The linchpin of that objection is his claim that the moral version of OIC is implausible. To see this, suppose we left that open, and so when I allowed for the sake of argument that one is never required to do what one can't do, I could not unreasonably mean morally required. That deprives Southwood of his crucial pivot, the alleged pressure to move to deliberative ought, which in turn is held better to serve the anti-idealist purpose.

So, my reply will be simply that I mean the moral ought when I allow that it may be that one is never morally required to do what one cannot do, and find only a very quick and unconvincing argument against that principle. I will not further revisit the issues about a deliberative (much less "prescriptive") sense of "ought," preferring to cut off their relevance by doubling down on moral ought. I reply to Southwood's argument against OIC (borrowed from Sinnott-Armstrong) below, and stand by my view that the moral version of OIC is one reasonable interpretation of OIC intuitions in relevant cases, with no need to revert to deliberative oughts.

We will revisit these points in more detail, but I want to prepare the way by clarifying and correcting some suggestions about my distinction between proposals and principles, since they will figure centrally in the main issue.

4.2 Principles and proposals

Southwood's discussion helpfully presses for more clarity about my distinction between proposals and principles than I provide in the book. For convenience, I'll use the capitalized forms, "Proposals" and "Principles" to stand for institutional proposals and institutional principles respectively. He tries to give my distinction a more precise meaning, and his interpretations must be hypotheses since they go beyond my insufficiently precise presentation. Before looking at Southwood's critical challenge, I want to correct several errors of interpretation that understandably result. These clarifications (if that's what they are) do not alter my position, I

believe, but they add specificity and explicitness, prompted by this constructive pressure. We should start with these corrections before turning to his substantive critique with these clarifications in mind.

Southwood understands me to hold that Principles and Proposals are, in the following ways,

> …different kinds of ought claims: claims involving conjunctive, plural requirements addressed to societies on the one hand (the ought of justice); and claims involving non-conjunctive, genuine deontic and unconditional ought claims addressed to the state on the other (the ought of public policy). (16)

The deontic/plural distinction is correct,[41] but I think that the other two attributes are not. Proposals and Principles can both be either conjunctive or not; and Principles can apply to the whole society or to subsets; Principles *of justice* in particular do apply to the whole society as such, but so can Proposals. Let me explain.

Southwood suggests that for me, Principles, unlike Proposals, have a "conjunctive" form, as in "Build & Comply," whereas Proposals have an atomic form (let's call it), as in, "Build." He writes that according to me, "institutional principles are…claims to the effect that a society ought to implement and comply with some particular institutional arrangement." I don't say that (and I don't say nearly enough),[42] but he seems to be suggesting that I'm committed to it. I doubt this, though that reading is understandable, because the subset of Principles that I'm especially interested in do have such a conjunctive form. In those examples, one conjunct is about building some institutions, (or at least about some range of things like that) and the other is about complying with the institutions (or something relevantly similar to that)—generally, some appropriate follow-through. Let's call them "follow-through conjunctions," for convenience. I add those parenthetical qualifications because my points about "Build & Comply" would apply as well to cases such as "Impeach & Replace," which is not a case of building and complying. The immediate point is that the examples that are relevant to my arguments are indeed conjunctions of that general "follow-through" kind. And I repeatedly argue that such Principles do not entail requirements to perform the act in either conjunct—say, Build, or Impeach.[43] However, a requirement to Build or a requirement to Impeach could well be a Principle all by itself, and not a Proposal, even though each is atomic rather than conjunctive. That's because there could be a standing requirement to Build or to Impeach under the circumstances. Indeed, if we will Comply then we might well be required, by that fact plus the standing requirement (the Principle), to Build & Comply. A requirement, under the

---

[41] As we saw in the precís, I develop a conception of "plural requirement," which is not a classic deontic requirement, applying as it does to sets of agents rather than to agents themselves.

[42] I mainly discuss the distinction at pp. 115-16.

[43] I am simplifying here. In Chapter 8, on "Concessive Requirement," I consider the controversy about this question, and argue that my own arguments can, with care, be formulated either way. I then proceed to adopt the "actualist" formulation on which O(A&B) does not imply O(A).

circumstances, to Build, is not a speech act, so it is clearly not a Proposal. It takes the form of what I call a Principle.[44] So Principles need not be follow-through conjunctions, or conjunctions at all.

Must what I call Proposals be non-conjunctive—atomic—as Southwood also suggests I hold? Again, I doubt it. Someone might perfectly well propose that we Build & Comply. Doing so would not be appropriate or a "happy" utterance as a proposal (in Austin's phrase)[45] unless certain things were taken as common knowledge in the conversational context—namely that, since the value of Building depends on Complying, the Complying is also predicted to occur. So, as I see it, it is distinctive of (but not sufficient for) a Proposal that this category of speech act "implicates" (in Grice's sense of conversational implicature)[46] that the conditions of that practical path being valuable are met. So, Proposals might be either non-conjunctive as in, "Build," or conjunctive as in, "Build & Comply." In the latter case, the conversational implicature I mentioned allows detachment of the Proposal, "Build." That is, that detached part is implicated, though not stated or logically implied. In the Principle, "Build & Comply," (rather than a Proposal of this same form) there is no speech act, so no such conversational implicature, so no warrant for detaching "Build." However, if there is a Principle, "Build & Comply," the value of Building depends on the Complying, and the Complying will in fact occur, then the Principle, "Build," itself, may also detach and stand alone.[47] But that would not make it a speech act and so it would not be a Proposal. So, in sum, neither Proposals nor Principles are either conjunctive or non-conjunctive by definition.

Next, as Southwood points out, I say that, "Social justice is a moral standard for societies … [R]equirements of social justice morally require things of societies as such."[48] Part of the point of that, as he notes, is that such requirements—Principles of social justice—are not fundamentally addressed to some subsystem such as the state. However, he understands me to hold that Proposals, by contrast, are addressed to the state.[49] I didn't say anything one way or the other about that view in the book, but I don't believe that view is correct.[50] On my conception of Proposals, they might

---

[44] Actually, I also understand Principles as having a further feature that isn't germane here, namely being relatively general, leaving quite open what institutional form might satisfy it. We could just as well let that define a special subset of Principles, General Principles, or some such thing.

[45] See, *How To Do Things With Words*, Harvard University Press; 2 edition (September 1, 1975).

[46] Grice, H.P. (1975). "Logic and Conversation," Syntax and Semantics, vol. 3 edited by P. Cole and J. Morgan, Academic Press.

[47] This is a factual kind of detachment, but "factual detachment" is a name usually given to a disputed principle in deontic logic concerning certain conditionals. This is not a case of that.

[48] P. 6, quoting *Utop.*, p. 126.

[49] Southwood writes that according to me, "Whereas institutional principles involve claims to the effect that a society ought to implement and comply with a particular institutional arrangement …, institutional proposals involve claims to the effect that the state ought to implement a particular institutional arrangement." (NS p. 9) Neither the definition of "institutional proposals" at p. 116, nor any other text that I can find suggests that institutional proposals are addressed to the state.

[50] I do say, "an account including principles of full social justice, so long as it does not presume to fix too much institutional detail, and so long as it is not presented to activists, vanguards, or governments as a practical proposal, is free of the mentioned vices." (p. 9). This does not imply that proposals are always

apply to the whole society as such, some subset of it, a subsystem, or sometimes to the state. Principles could also apply at any of those levels, but Principles of social justice in particular, a subset of Principles, apply to the society as a whole.

As for Proposals, as I say, they can be addressed specifically to the state, as in, "The Attorney General ought to be fired." Or, they can be addressed to the society as a whole, as in, "We ought to constitutionally extend the franchise to some minors." The latter could not be comprehensibly addressed to the subsystem of society that is the state. So Proposals are not necessarily addressed to the state and might be addressed to the society as such.[51]

Having corrected the record in several ways, we can see how several of these misunderstandings get in the way of at least one of Southwood's arguments. He understands Proposals to involve what he calls the "ought of public policy." At p. 18 he writes,

> [Claims involving the ought of public policy] are claims to the effect that the state ought to implement a particular institutional arrangement; and, for all that we are told, it is not wildly unrealistic for the state to implement the institutional arrangement.

His point is that the Proposal to (as he mistakenly assumes) the state is one that the state could indeed implement whether or not it would be complied with. So, he says, it is hardly "wildly unrealistic." But "wildly unrealistic" is his phrase, and here it seems to be suggesting things that are not part of my realism constraint on Proposals. First, as I have indicated, Proposals are not necessarily addressed to the state, as his first clause says. Second, and more importantly, the constraint of realism that I would place on Proposals is, as I said earlier, that they are inappropriate unless the conditions for the value of that practical path are (likely) to be satisfied. The fact that there is nothing stopping the state (or the society) from implementing a policy does not speak to whether doing so has any value when there will not be sufficient compliance. What needs to be realistic about the Proposal to Build an institution is not only that nothing will stop the agents from Building it (as Southwood suggests), but whether the conditions of its being a good idea, under the circumstances, to do so are met—such as sufficient compliance. Short of that, such a Proposal appears to be based on, or at least to implicate, unrealistic suppositions. Having said this, it was hardly obvious from my text that this is how I understand the constraint, and so I elaborate it more fully here. I turn, now, to Southwood's critique.

Footnote 50 continued

presented to subsets, only that they can be. For more on the idea that requirements of justice might be addressed to the state, see Enoch, "Against Utopianism: Noncompliance and Multiple Agents," *Philosophers' Imprint* 18, no. 16 (September 2018): 1–20, Chapter 1, section 7 of *Utop.*, and my reply in *Utop*, and secs. 8 and 9 of my replies to Enoch in this volume.

[51] This is closely related to the issue at the core of David Enoch's paper, "Against Utopianism," *Philosopher's Imprint*, volume 18, no. 16, September 2018. In the book I reply at Chapter 1, section 7, and it surfaces again briefly in his comment and in sections 8 and 9 of my reply in this volume.

4.3 Holding things together

The heart of the critique, which will bring him to the issues about ought and can, is his claim that three positions of mine are not a coherent combination. I'll simplify them and put them in my words.

a. Principles of social justice are requirements, and as such, must be something the society is able to meet.
b. Principles of social justice need not be realistic in the sense of being at all likely to be met.
c. Proposals must be realistic in the sense that any conditions of the value of that practical path are assumed to be (likely to) be met (such as, often, compliance).

Now, it is unfortunate in a way that I give my own formulations, rather than any of the several successive formulations in Southwood's piece. I admit that some of his arguments might not directly apply to my formulations in the way they applied to his own formulations of my positions. But I believe this is fair, since the question must be whether these positions as formulated here (so long as they are not changing the view in the book) are consistent, rather than Southwood's formulations, in case the verdict were to differ. In any case, I believe my reformulations won't prevent us from confronting the full force of Southwood's main arguments.

Southwood's central criticism of that three-pronged view concerns my example of Messy Bill, and an alternative example of Bill*, who I will call, more descriptively, Scheming Bill. Messy Bill can't bring himself (I've alerted us to the importance of this phrase) to take his trash to the curb each week, because he is too lazy and unconcerned. Scheming Bill arranges things—maybe using an interfering robot—so that, should he try, even (unlike Messy Bill) intrepidly, to take his trash to the curb, his prearranged mechanism ensures that he will fail.

Southwood argues as follows: If, as I say, Messy Bill is plausibly required even though he is unlikely and deeply disinclined to do it, it's at least as plausible that Scheming Bill is required even though he is literally unable. (That would violate "ought implies can," of course—I'll come back to that.) In support, he follows the above passage with, "After all, hasn't [Scheming Bill] simply intentionally made it the case that he is unable to do what he plainly ought to do?" The unstated assumption is that this scheming does not exempt him from the requirement to do the thing he made impossible. I'll explain shortly why I don't accept that assumption. But, suppose that it did have to be granted that Scheming Bill is required to take his rubbish to the curb even though "he is unable" to do so. In that case, there would be an unattainable requirement. So, it would appear that unless I appeal to an "attainability constraint," that is, a version of "ought implies can," (OIC), since it is as plausible that Scheming Bill violates a moral requirement as that Messy Bill does, there would be pressure to either say that both are violators or that neither is. If, however, I were to allow OIC, then I could hold that unlike Messy Bill, Scheming Bill does not violate a moral requirement, because he is unable to do it.

In a moment, I will respond to Southwood's challenge to OIC. First, there is a subordinate issue to address. It may seem from what I have said that I need to commit myself to OIC. In the political context, that challenge would continue like this (again, in my own words):

> Estlund says we aren't off the hook of supposed requirements of justice simply by being strongly disposed not to comply. But it is similarly not plausible that we should always get off the hook by being *unable* to do what is required. Contrary to his saying that he grants OIC for the sake of argument, unless Estlund commits to OIC, he is committed to our being required by justice to do what we cannot do. And that commits him to a "manifestly implausible anti-realism."[52]

I don't see how my holding that requirements don't depend on likelihood or dispositions thereby commits me (unless I rely on OIC) to holding that even the impossible can be required. If the impossible can be required, then the impossible can be required, and that isn't borne upon by whether the unlikely and uncharacteristic can be required (as in Messy Bill). Now, I'm happy to allow, at least for the sake of argument, that it would be absurd to hold that justice can require what we cannot do (though some excellent philosophers have held just that, such as G. A. Cohen).[53] But for my purposes I do not need to commit to that, since my point stands either way: we do not get off the hook by being strongly disposed to be more selfish or partial than justice might claim to require. I interpret the relevant opponent to hold that those dispositions amount to inability,[54] and to hold (whether I do or not) OIC. I respond by pointing out that even if OIC is true, those do not, in any case, amount to inability. None of this argument of mine is threatened if it's not the case that OIC doesn't hold. Rejecting OIC might indeed commit a person to what Southwood calls a, "manifestly implausible anti-realism." But that is their business.[55]

As it happens, Southwood indicates his own doubts about OIC. He says that the Scheming Bill case raises such questions.[56] If so, then an especially radical anti-

---

[52] Southwood indicates this would be the cost for me if I didn't impose the constraints he calls "minimal realism," and "attainability," the latter constraint essentially amounting to OIC.

[53] See Cohen, G. A. (2008). *Rescuing Justice and Equality*. Harvard University Press, pp. 250ff. What I attribute to him here is accurate, though there is nuance about whether he rejects "ought implies can." He allows that this might be correct about "ought," but he denies that this would bear on what is "normatively fundamental." Justice might require something impossible, if it is the case that we ought to do it if we can.

[54] Some, like David Wiens, have explicitly embraced a version of this position. See, "Motivational Limitations on the Demands of Justice," *European Journal of Political Theory* 15 (2016) (3):333-352. See also my "Reply to Wiens," in the same issue.

[55] Southwood argues, especially toward the end, that it is also my business, but I am not persuaded. I argue that justice can be at least as idealistic as to even require what we can't bring ourselves to do. I take no stand, for purposes of the book's argument, on whether it might even require what we are unable to do.

[56] See later, referring back to this discussion, "as we have seen, there are legitimate questions to be asked about 'ought' implies 'can,'" (Page number not available). See Nicholas Southwood, "The possibility of wildly unrealistic justice and the principle/proposal distinction," *Philosophical Studies* (2020). https://doi.org/10.1007/s11098-020-01532-w.

realism would seem to follow in its train, as we have seen. But that would be a liability for anyone who does reject OIC in that way, not for me. I neither rely nor take a stand on its being correct, or its being incorrect.

4.4 Ought might imply can

If, in granting that Scheming Bill (who we stipulate is not able to take his trash to the curb) is not required to do it, I must be understood most charitably to have in mind a "deliberative" sense of "ought" or "required," then Southwood hopes to show that, in that sense of "ought", I can no longer insist that Messy Bill is required. Now Southwood's argument that a moral interpretation of the requirement is too implausible derives support from the following very brief statement from Sinnott-Armstrong (about a case similar to Scheming Bill): "We do blame agents for failing to do what they could not do if it is their own fault that they could not do it. For example, we blame drunk drivers for not avoiding wrecks which they could not avoid because they got themselves drunk."[57] That example, then, is tacitly introduced by Southwood to suggest that, at least when an agent intentionally brings it about that she can't do something (like Scheming Bill), the ensuing inability is no (obvious) basis for thinking there is no requirement.[58]

How persuasive is that driving example, as a reason to doubt OIC? One could fairly point out that we might sometimes talk in those terms. But we also talk in ways that should count just as heavily in favor of OIC, even in the same kinds of cases. For example, suppose someone were to say to the driver, "You ought to have steered away from the oncoming car." The driver might very well say, "I couldn't, I was asleep." That's true, of course. But why would he say it? I doubt that it strikes you as a *non sequitur*. And I highly doubt the accuser would say in response, "Yes, I know you couldn't, but you ought to have done it anyway." In fact, I think that would strike us as bizarre. That indicates that the driver believes—and believes that his interlocutors believe—that his inability defeats that particular alleged requirement (even if he is to blame for driving drunk, and so responsible for the crash).[59]

So far, in gathering things we might say, the scale seems to be more or less balanced for vs. against OIC. So, I don't believe that such common responses to the driving example and others like it count seriously against the idea that "ought implies can," any more than the driver's response counts seriously in favor of it. Of course, we do find the drunk driver to have done other than he morally ought, but it's clearly the driver's fault (barring some excuse) that he drank and drove, risking

---

[57] Walter Sinnott-Armstrong, (1984). 'Ought' conversationally implies 'can.' *Philosophical Review* 93 (2):249-261.

[58] Unlike Sinnott-Armstrong, I formulate this issue with "requirement" instead of "ought," here and in the book. Southwood clearly intends the drunk driver example in Sinnott-Armstrong to be an argument against my own view, so formulated.

[59] Suppose it were held that the driver could indeed have avoided crashing. The way to do it would have been not to drive drunk. Fine, but this is no use to Sinnott-Armstrong or Southwood, since they want an example of inability. Suppose it might be natural for the driver to begin his response with, "Yes, but…" Is that conceding the requirement? I don't believe so, since it is just as reasonable to read it as "Yes, if I could have, but…".

his falling asleep and getting into a crash. This is like Scheming Bill's wrongful scheming. It captures the culpability without supposing the impossible is required. So, we certainly morally charge the driver, as with Scheming Bill, for that.

Now, some think that in light of the crash we would blame the driver more severely than we would for only driving drunk. And that would seem to make sense only if the crashing is an extra violation, contrary to "ought implies can." But I can't see on what supposed basis we are *warranted* in judging the driver more severely.[60] Maybe we often do so, but it wouldn't be the first time that certain ordinary moral responses are mistaken, and in conflict with others. Also, in any case, it must be balanced against the fact that a casual ordinary view of these things seems to have no reply if the driver asked, "I know I shouldn't have driven drunk, and that's a serious matter because of the increased risk of crashes. But on what basis am I to be judged more harshly yet for something additional that I was unable to control—my car colliding with another car? What is additionally wrong about that?" The view I am opposing has a name, of course: "moral luck." I don't claim to have advanced the discussion of that topic, but I see no basis for such a thing in cases like these.[61]

Southwood does make clear what positions I need to hold, but he does not try, in addition, to mount much of an argument against them. If I am wrong and there is moral luck of this kind, and also (or therefore) "ought implies can" is false, then perhaps (or at least if we blame Messy Bill) we should blame Scheming Bill for failing to do something he was unable to do, and also count some societies as unjust for failing to do which they are literally unable to do.

Summing up: My view about Messy Bill places no pressure that I can see to hold that Scheming Bill is also violating a moral requirement, and I take no stand on the latter question. But I understand that some others do take that view, and Southwood is right to point this out and pursue what its implications would be. Still, with respect to that dispute about whether "ought implies can," Southwood offers only the ostensibly intuitive pressure from the driving example. For reasons I explained, that exerts very little pressure, on reflection. If one rejects OIC then she may be saddled with a manifestly implausibly anti-realism, or maybe it's a tenable position after all. I'm not myself persuaded by any objections to OIC that I have encountered, but nor do I rely on it for any of my own arguments in the book. Since I have no need to rely on a deliberative sense of ought or requirement, my arguments that justice might require more than people, as we say, can bring themselves to do, is not affected by any features of other senses of those terms.

---

[60] He is, indeed, to blame for the crash. But that only shows, as Zimmerman argues, that he is "culpable for more things," not that he is "more culpable." See Zimmerman, "Moral Luck: A Partial Map," *Canadian Journal of Philosophy*, Volume 36, Number 4, December 2006, pp. 585-608, at p. 598.

[61] Williams, B. A. O. & Nagel, T. (1976). Moral Luck. Aristotelian Society Supplementary Volume 50:115–151. Williams and Nagel are the main original sources on moral luck. My position is similar to that of Judith Jarvis Thomson, and her discussion of two drivers, in, "Morality and Bad Luck", in *Moral Luck*, D. Statman (ed.), Albany: State University of New York Press, 1993.

5 Reply to Enoch[62]

David Enoch begins by promising to disagree with me, difficult though it may be (so he reports). I will argue that, to a large extent he does not deliver. I don't mean that his disagreements with me can be well answered. I rather mean that, despite appearances, in great measure he does not disagree with me after all. We do seem to disagree about whether we disagree, so that is something. And in fairness, there are a few things I agree we (might) disagree about. His comments are organized mostly around a series of objections, each pursued only briefly. I am glad to have the chance to respond to so many concerns, some of which I know others share. So, rather than leave many of them unaddressed by considering a few at length, I respond in kind: with a series of brief responses to his brief (ostensible?) objections. I am not casting his comments as unduly negative. Enoch also indicates quite a lot of agreement, often on matters central to my aims in the book. Nevertheless, he still does think we disagree more than I think we do.

I will treat together the connected points that Enoch names his "first objection" and "second objection," which go as follows. His "First objection:"

> Saying that non-concessive theory enjoys a kind of priority over concessive theory because the former does not—as the latter does—accommodate failures to fully meet the principles of (aspirational) justice, sounds dangerously close to the claim that non-concessive theory enjoys priority because it is, well, non-concessive. So a begging-the-question worry naturally arises.[63]

In brief reply, the passages I quote in footnote 66 are offered both as my actual explanations of that relation in the book, and as showing, I hope, that the ambiguity he alleges in the second objection is not present.

That brings us to the "second objection:"[64] At one point, I write, about what I call "concessive" justice, that if it should turn out to differ from "non-concessive justice, "It is right only because something is wrong."[65] Enoch worries that this might be misleading since, as he correctly observes, the wrongness of the other "something" is not what makes the concessive thing required. But, in reply, it is indeed the wrongness of the other "something" that makes the concessive thing concessive. The passage in which he notices an ambiguity does not occur in any of the several parts of the book where I explain what is less fundamental about concessive requirements, so those should be factored in. In any case, while that formulation of mine would indeed be misleading read *de dicto*, Enoch admits that it is correct read *de re*. There is no disagreement with me here, then, since I did not

---

[62] Or, "On the Limits (If Any) of Enoch's Disagreement with *Utopophobia*."

[63] P. 3.

[64] P. 4.

[65] *Utop.*, p. 194.

mean it to be read *de dicto*. I believe several passages bear this out, and should also dispel the suspicion that the account is question-begging.[66]

Enoch asks a common question which many take to amount to an objection to my account about what it is to "idealize"[67]:

> Think of all the other standards [other than moral standards] that people may fail to fully comply with. Think, for instance, about epistemic standards, or prudential ones. Why stop with prime justice, that idealizes away moral violations, and not go for, say, super-prime justice, that idealizes away epistemic and prudential violations as well? Perhaps aesthetic ones too?[68]

That question only arises if someone is trying to figure out what "idealizing" is or what is the right way to do it. I don't consider that question in the book. My questions are specific, and make no essential use of the generic idea of "idealizing." For example, a central question arises within the sphere of moral requirement (broadly conceived): If requirements of justice are, as I argue they are, moral broadly speaking, then why should they take as given certain other moral violations? At a concessive level, yes, obviously, but that is not fundamental as we have discussed. The way in which the non-concessive has primacy is central to this argument of mine, though it goes along with an argument that any view that leaves justice wholly at the concessive level faces a fatal indeterminacy.[69] The details of that argument are not necessary here. I hope that gloss indicates how Enoch's question about what idealizing amounts to does not need to arise (and so we are not disagreeing about it).

---

[66] I won't go into detail about what the primacy relation amounts to here, except to quote without comment several places where I say how I understand it, noting that the *de dicto/de re* ambiguity does not seem to me to be present, and I do not believe these suggest anything question-begging:

"That is all I mean by a "concessive" principle or requirement: it is a requirement that is in place owing to our conceding certain violations of other requirements. Some requirements are in no way conditioned by violations in that way." (*Utop*. 6)

"I use the term "concessive theory" for questions regarding what institutions society ought to build or maintain, given that it will not comply with what justice requires." (*Utop*. 22)

"there is a certain primacy of the nonconcessive requirement: [Professor Procrastinate's] requirement not to accept arises only if the nonconcessive requirement—to both accept and write—is violated by his not writing. It evaporates if that requirement is met. The reverse is not the case: the nonconcessive requirement to accept and write does not appear or disappear depending on whether he accepts or writes. Call this the primacy of the nonconcessive: concessive requirements are subordinate, arising only because of violations of nonconcessive requirements." (*Utop* p. 30)

"To concessive requirements arise only from failure to meet the nonconcessive. The nonconcessive requirement is not contingent in this way: even if society is not just, or people are not morally good, that is all (together) still morally required." (p. 31).

[67] P. 6.

[68] P. 7.

[69] The comment by Brennan and Sayre-McCord also remarks on the alleged indeterminacy. See, "Estlund identifies two interesting grounds for thinking that ideal theory is the place to start. …The other is that on the fully concessive view 'there is no single salient standard of social justice at all…'" (Page number not available). See David Estlund, "Replies to critics," *Philosophical Studies* (2020). https://doi.org/10.1007/s11098-020-01534-8.

Enoch criticizes my account (in Chapter 17) of the value of understanding justice in terms of its contribution to being a morally well-constituted person, as arguing that only philosophers can be fully virtuous.[70] We agree, of course, that a person does not need to be a philosopher to be rightly oriented toward justice. Some people might have the relevant understanding and orientation without having done the philosophy. But it may be that others gain it only by doing, or learning by way of others doing, what amounts to some philosophy. Or, maybe even no one could have the right orientation if no one had done the (not necessarily professional) philosophy. I don't choose between those, but they should suffice to reassure us that the account does not make virtue depend on being a philosopher. We are in agreement that such a thing would be absurd.

I argue, as against a well-known argument of Sen's,[71] that searching for coherence among our convictions (and adjusting them as well) is indispensable in arriving at new moral judgments. Practical social choice needs only a pairwise ranking of alternatives. But it is unclear how to arrive at a fulsome set of rankings, rather than only the few obvious cases mentioned by Sen, unless we may make use of the many common convictions about justice that imply that there is such a thing as full justice, rather than merely "juster."[72]

Enoch expresses agreement with the importance of such "coherence considerations,"[73] though he adds that he doesn't think "the line of thought in the text shows that thinking about the ideal is necessary for theory construction, or that the only way of thinking about comparative judgments (juster than) without thinking about the ideal is 'the eyeball method.'"[74] I agree, and my text is not meant show those things. But we apparently do disagree about whether they are correct, though neither of us attempts to show it.[75] Still, we agree on my main claim, that depriving ourselves of our threshold-implying convictions might well hamper our moral thinking. I don't know whether he agrees with my stronger implication, namely, that so depriving is not only conceivably handicapping, but very likely so. I'm not sure whether there is any disagreement here, but not very much in any case.

Enoch argues that my use of what I call "countervailing deviation" to establish a certain kind of practical significance for a theory of full and complete justice "fails."[76] His argument suggests to me that there is no disagreement between us, but let me first sketch the countervailing deviation idea as economically as possible, though unfortunately a bit cryptically: Suppose you have a sound account of the several principles that a society must meet in order to be fully just. However, in real life no society, suppose, will meet all of them. But then you cannot assume that

---

[70] Footnote 12.

[71] Sen, Amartya. "What Do We Want from a Theory of Justice?" *The Journal of Philosophy*, vol. 103, no. 5, 2006, pp. 215–238. *JSTOR*, www.jstor.org/stable/20619936. Accessed 5 May 2020.

[72] See *Utop* pp. Chapter 13, section 3, pp. 261-70.

[73] P. 8.

[74] Note 13. The "eyeball" reference to *Utop* is from p. 266.

[75] I don't say it is "necessary," so I won't endorse that stronger claim here.

[76] P. 9.

meeting any of the other principles has any value at all, since that might depend on the meeting of all of them. (This is to avoid what I call the "fallacy of approximation.") Nevertheless, it will sometimes be possible to highlight the specific deviation of a society from full justice by comparing the reality with that sound standard. The practical value of doing that might be that even if there is no way to correct that deviation, there might be certain changes that are available that can, with some thought, be seen to correct the loss of value, and to do so even though it is yet a further deviation from the full standard. (The fuller account makes up Chapter Fifteen.)

Here is Enoch's critique:

> This attempt at defending ideal theory's practical significance fails, I think. To see this, consider first the following quote: 'By understanding the epistemic value of tolerance in its ideal setting, Marcuse is able to conclude that this is a value that tolerance will not have without that fuller setting and its other elements.' (*Utop* 289). As stated, this is clearly a fallacy. Based on the observation about what makes tolerance of (epistemic) value in the ideal, together with the observation that that feature is missing in the real world, Marcuse is only entitled to the conclusion that tolerance is not of value in the real world for the same reason it is in the ideal."[77]

In response, it would indeed be a fallacy. However, I didn't say it follows (and Enoch says, "I am not accusing Estlund of [that] fallacy") but only that Marcuse's understanding of that value allows him to come to that conclusion.[78] Also, though, it is clear, I think, from what I say about the fallacy of approximation that I grant and emphasize that the element that is valuable when other elements are present might also be just as valuable and even valuable in the same way even if they are not present. But the point is that this is not guaranteed. And with some thought we can sometimes determine that this is not so. Marcuse gives reasons to believe that a certain kind of freedom of speech, while a contributing part of a highly valuable combination of conditions (such as a certain kind of social equality of power, among other things), lacks that value where some of them are missing, and—a further finding—there is no evident alternative kind of value that it does have in that case. So, I don't believe there is any real disagreement here, though it takes us directly to the next point.

Enoch argues that the point about countervailing deviation—which, in the non-deductive form in which I present it, he does not criticize—shows no practical value for ideal theory, "compared to just starting off with non-ideal theory."[79] That is, he grants that if our strong attachment to free speech derives from our attachment to a fuller package of rights, it could have practical value to realize that strong free speech might not have that value without the rest of those conditions. Enoch's point

---

[77] P. 9.

[78] Herbert Marcuse, "Repressive Tolerance," in Robert Paul Wolff, Barrington Moore, Jr., and Herbert Marcuse, *A Critique of Pure Tolerance* (Boston: Beacon Press, 1969), pp. 95-137.

[79] Pp. 10-11.

here is this: But in that case we only got into trouble by asking what would be a good or even fully just cluster of conditions—roughly a kind of ideal theory. Had we never done that, we would not have made that mistake. That's correct, and that is all my argument is meant to show. If (as is indeed the case, I believe) many people do often value things like certain liberal rights in a way that is owed to their (or those they have learned from) erroneously inferring it from the value of a set or system of rights, then we agree, as far as I can tell, that doing some ideal theory could help to expose the mistake. The casual suggestion that theorizing about the fuller set of rights might have better been skipped altogether I let pass here. Whether and in what way understanding the principles of full justice would be valuable is a separate issue. The point here is that it does happen, and encourages that kind of mistake.

Enoch writes,

> Estlund characterizes "the guiding question" of the book as "whether people might … be robustly politically deficient. Political 'realists' and others often say, in effect, 'no.'" (*Utop* 3) But while he may be right about some "political realists," this is not, I think, the best understanding of anti-utopian concerns.[80]

Notice that I'm not saying whether it is the best understanding or not (no disagreement so far), only that (I argue for this, and I doubt Enoch disagrees with it) this is one characteristic thing that realist writers are often committed to. I argue for this, and I doubt Enoch disagrees with it. That is, I'm not, in that one passage, trying to interpret political realism, which of course comes in many varieties, many of which I discuss throughout the book. Rather, I'm pointing to an implication of many realist views, probably unwelcome to them, by virtue of their objecting to standards in politics that significantly outstrip what humans are ever likely to meet.

Enoch then presses two interpretations of anti-Utopian concerns that he says I "misdiagnose." The first one is about "defects" in a theory. Enoch makes an observation about the following quotation from my book:

> *My Claim (against Utopophobia)*
> It is no defect in a theory or conception of social justice if it sets such a high standard that there is little or no chance of its being met, by any society, ever. Such a theory could nevertheless be true.[81]

He points out that, "the two sentences here are not equivalent," because there might be defects in a theory other than it's not being true.[82] Indeed there could, we don't disagree about that. But as I would have read that passage, the second sentence

---

[80] P. 11.

[81] *Utop.* p. 126.

[82] P. 12.

appears to settle that this is the particular kind of defect I'm referring to.[83] "My Claim" does not take a position, I think, on whether a theory like that might be defective in some other way even though the theory might be true. I believe that's the best interpretation of my text, in which case there's nothing here we disagree about. Nevertheless, I hesitate just slightly, since I do at one point say this, which might be interpreted more broadly (I add emphasis in two places for our purposes):

> It is hard to resist the sense that a hopeful theory is a better kind of theory. Still, I think this is an important mistake. There is *no defect* in a hopeless normative theory, and so none that hopeful theories avoid to their advantage. Things are better in one way, of course, if the best theory turns out to be hopeful rather than hopeless: it is unfortunate if people and societies will not live up to sound requirements, and fortunate if they will. But this consideration is patently *no support* for a less hopeless theory. That would be to believe in different, more easily satisfied moral standards for the reason that they are more likely to be satisfied. This is not moral or normative reasoning at all, it seems to me.[84]

"No defect" might suggest a broader claim that it doesn't refute the theory. On the other hand, I do speak even there about what counts as "support" for a theory, which sounds to me like the truth question again. But, in any case, even though a theory can have defects other than being false, and a theory of justice's being "hopeless" might be held to be some other kind of defect, nevertheless, "still it would be nice to have arguments for this view. Why should we think that a theory's hopelessness counts against it, even just a tiny little bit?"[85] This complaint is from Enoch, taking my side in this respect.

The second way of understanding some Utopophobic thought, which he complains that I neglect is this:

> many of the…worries about utopianism should be understood as worries about multiple-agent cases: Worries, for instance, about whether society should put in place institutions and arrangements that citizens are unlikely to comply with, or concerns about what *me and my fellow good-guys* should do on the political scene, given the violations by the *bad guy*. If this is a good reading of much of the anti-utopian sentiment, then Estlund's discussion does nothing to weaken it.[86]

Certainly, some political philosophers are concerned that other political philosophers propose to put in place institutions and arrangements that, so these critics

---

[83] There is some corroboration at p. 28, when I describe a "human nature constraint" that I will reject, as saying, "A normative political theory is defective and so false if…" and at p. 84, "So far, I contend, the theory has no defect. It might be a false theory if it claimed that the standards would someday be met, but it does not say that." And at p. 118, "they nevertheless reveal no defect in the hopeless theory, which might be perfectly correct."

[84] *Utop.* p. 199.

[85] P. 13.

[86] P. 14.

believe, citizens are unlikely to comply with, etc. But it does not strike me as a very interesting question whether such proposals would be ill-founded—of course they would. Any significant dispute would presumably be about their respective predictions. And obviously Enoch and his fellow good guys ought to condition their political action on all the relevant facts, which often include "violations by the bad guy." We agree about all this, but I expect we also agree that this is obvious. So, I have not tried to do anything to weaken such concerns—they are perfectly legitimate concerns and obviously so.

Following on that point, if Enoch believes only that political philosophy is partly, and importantly, about what some members of society ought to do taking the behavior of the others as given, then I agree, as he acknowledges.[87] As he also says, we agree that "that not all cases in political philosophy are multiple-agent cases." So we agree so far: those aren't all there is to political philosophy, but they are an important part of it. However, does Enoch believe that political philosophy is really, or, or most centrally, or mainly about such questions? If so, I do not agree for reasons I give in that section of the book. And, unless Enoch means that the multi-agent questions are in some way, to put it broadly, privileged over single-agent cases (which he takes to include such questions as "what should we, this society, do?") in a proper understanding of the field of political philosophy, then, I cannot see what it is that he thinks I have neglected or that we disagree about.

Enoch says in his comments, "that political philosophy is essentially about multiple-agent cases," but a few sentences later it appears that he may mean only that there are "many" such cases. As he says, I had understood him in his article, "Against Utopianism," to somehow privilege the multi-agent cases and I argue directly against that position, in the book.[88] So, consider this passage from that article, which leaves me thinking that perhaps he does mean something more privileging, so to speak, by saying, "political philosophy is *essentially* about multiple-agent cases."[89] He writes there,

> In moral philosophy…we typically ask about the principles regulating the actions…of individual agents.… Here…of course, sometimes the actions of others are relevant.… But these are, when we're doing moral philosophy, complications, perhaps atypical ones. In political philosophy, though, the multiplicity of agents is a crucial part of the problem. Political philosophy is *essentially* about multiple agents.[90]

He doesn't say the weaker thing that questions about multiple agents are a crucial part of the subject matter of political philosophy, but a crucial part of "the" problem. This seems to mean (what else could it mean?) the problem of political philosophy. And, he apparently means, such multi-agent questions are crucial to that problem in a way that (single-agent) questions such as "what ought we to do as a

---

[87] "about these judgments Estlund *agrees* that they are essentially a part of concessive theory." (14).

[88] Chapter 1, section 7.

[89] The italics are his in the earlier article, (notably?) dropped in his comments here.

[90] Enoch, "Against Utopianism" Philosopher's Imprint, volume 18, no. 16, September 2018 p. 4.

society?" are not a crucial part of the problem of political philosophy. If I have him roughly right, we (finally) disagree.

Being unsure, I didn't quite attribute that view to him in that article. But, in case anyone might be tempted by it, I state such a view and criticize it in the book.[91] Now, if Enoch does not mean to take any such view, then I'm afraid I do not know what he means, and then for all I know we don't disagree about this either.

Perhaps the leading traditional objection to Utopian theories is that they betray a dangerous naivete, a blind, blithe optimism. Naturally, then, in writing the book I took pains to be clear that nothing about my view is motivated by, nor does it enlist support from, any particular optimism. It is not for nothing that I dub the approach I defend, "hopeless." And, with an eye to charges of undue optimism, the book's epigraph from James Baldwin draws a picture of impending doom, which, as I say in the Preface, "is not blinkered by optimism or even by any clear hope," despite its evocation of the highest standards of love and fellowship.

Still, for Enoch, "it's hard to shake the feeling that [his] hope of compliance… is important for Estlund, despite the fact that his arguments do not officially depend on it."[92] With my space coming to an end, I must leave it mostly to the reader to judge why this question is raised if it is admitted not to bear on any of the arguments. It is not just idle speculative biography, but is meant to be part of his critique. I suppose, (and we are apparently permitted to psychoanalyze) it is meant to enlist the traditional naivete trope in order to be discrediting, if only a little bit, in some way that is additional to whatever success his critical arguments have on their own. If he were to show that I am unreasonably optimistic, how would that bear on whether the book's arguments are successful? Enoch doesn't say. Again,

> Officially, nothing in Estlund's account depends on optimism about us humans… But I think that deep down, at the level of motivation if not at the level of explicit argumentation, Estlund is optimistic, indeed optimistic beyond what is plausibly consistent with the evidence.[93]

Having worked in this area for about a decade, I can tell you that the charge would have illegitimate *ad hominem* success with some readers, which is why I studiously tried to preempt it. But what matters is that it would have no probative force at all.

In any case, it remains unclear to me what gives Enoch this impression of my unreasonable optimism. I can only briefly touch on two pieces of supposed evidence, though he offers a few more: First, in a couple of places I counsel against irrational, premature pessimism.[94] Given the naivete trope, it could easily be misleading for Enoch to cast this as a sign of optimism, as if counseling against driving too slow is a sign of a thrill seeker. Since Enoch suspects that, however

---

[91] Chapter 1, sec. 7.

[92] P. 16.

[93] P. 15.

[94] One is the pastoral metaphor at p. 10, the other my remarks around "unbelievable moral progress," at p. 259ff.

pessimistic I might or might not be, he is even more pessimistic, I won't try to compete, but nothing hangs on that question.[95]

Second, and finally, while my main arguments wouldn't depend on it, I suggest (in Chapter 13) that good careful idealistic political philosophy can contribute to a salutary cultural openness to better human possibilities. Too much emphasis on what strikes us as sufficiently realistic would, arguably, have slowed or prevented such things as the abolition of slavery, or the election of a Black president, developments which were simply implausible, and launched by extreme idealists. The fact that there are, of course, also dangers of too much idealism (to put it roughly) is compatible with there being a danger of too little. This may seem to have me assuming, naively, that idealistic political philosophy makes any difference to history at all. For better and worse (and so this is not optimism), I believe that it does—at least as a plausible contributor among very many contributors. No one agrees with me more about this than legions of anti-Utopians who share Jerry Gaus's unusually explicit fear that ideal theory risks repeating the disasters of Hitler, Stalin, and Pol Pot.[96] Given Enoch's doubts about philosophy's effects expressed here, I would expect that he would not align himself with any such qualms about ideal theory. But he does write, in historically familiar tones, "It seems to me that suggestions for small-scale improvements from where we find ourselves are going to be much more promising causally compared to fantastic utopias."[97] Part of this is apparently careless, since he presumably knows that where my own views are under discussion reference to "fantastic utopias" is a straw man.[98] But the rest of that remark indicates his belief that some ways of doing political philosophy are "much more promising causally" than others. There is a somewhat inchoate disagreement lurking here, but on the surface, on the question whether political philosophy has any significant affect on the world, it would appear that we might not disagree at all.

I am sure that there remain important disagreements between Enoch and me about idealistic political philosophy. I hope this "lightning round" of charges and responses will help us, going forward, to identify those more clearly. We agree about a lot, even according to him. And according to me, we appear to agree about even more than that.

Publisher's Note   Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

---

[95] As I was writing the book, occasionally explaining the ideas to non-philosophers, I sometimes joked that I might choose as a subtitle: "We Might Suck." Enoch might think that there's no question about it. Here, yet again, our positions would not be incompatible.

[96] Gerald Gaus, *The Tyranny of the Ideal: Justice in a Diverse Society*. Princeton University Press, (pp. 88-89), but also many others. More mildly, Brennan and Sayre-McCord speak in their comment for this volume of, "the moral recklessness ideal theory might …encourage." (Page numbers unavailable). https://doi.org/10.1007/s11098-020-01531-x.

[97] P. 16.

[98] I distinguish my view from views that imagine and promote vivid utopian worlds at pp. 8-10.