

Against Utopianism: Noncompliance and Multiple Agents

David Enoch

The Hebrew University of Jerusalem

© 2018 David Enoch

*This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivatives 3.0 License.
<www.philosophersimprint.org/018016/>*

Political philosophy in recent years has seen a surge of interest in a family of related methodological issues:¹ Questions about the significance, and purported priority, of ideal over non-ideal theory, and about the constraints, if any, that feasibility considerations pose for normative political philosophy (or, viewed in the opposite direction, whether there's any flaw in utopian, or somewhat utopian, theories in political philosophy).

In this paper I emphasize a neglected point in these debates — the fact that political philosophy is essentially about multiple agents. This observation allows room for the distinction between two different questions: how, if at all, an agent's foreseeable violation affects what *that agent* ought to do, and how, if at all, an agent's foreseeable violation affects what *other agents* ought to do. In what follows I place the discussion of this second question in its natural, wider ethical context, and attempt to answer it. I then argue that the distinction between the two questions allows us to grant the defenders of utopianism (most notably, David Estlund) the answer they want for the first question, and still defeat their utopianism in virtue of the very different answer we give to the second question.

The discussion proceeds as follows: In section 1 I present the worry about feasibility and utopianism, and, following Estlund, I note how, understood in one prominent way, the worry is misguided. In section 2 I show how the multiplicity of politically relevant agents allows us to understand the worry in a much more serious way (whose significance, as far as I know, Estlund never fully appreciates). In Section 3 I place the discussion in the context of the wider question about the sensitivity of one agent's duties to the wrong actions of another, and in section 4 I conclude, with observations about how we do and how we ought to do political philosophy. I include a long appendix, in which I argue that ideal and non-ideal theory (in political philosophy) should be seen as intellectually respectable attempts to answer different questions, and against the purported priority of the former over the latter. If you're on board with these claims before we even start, or if, frankly,

1. For helpful recent surveys, see Stemplowska and Swift (2012), Valentini (2012), Erman and Moller (2015), and Vallier and Weber (2017).

you've had enough of the ideal-theory debate, feel free to ignore the appendix: While I think the appendix may be helpful — both in general, in organizing some of the central themes of the now-exploding literature on these topics, and more specifically in placing this paper in its wider context — the main argument of this paper does not, for the most part, depend on it.

1. Utopianism, Feasibility, and Utopophobia

Theories in political philosophy are sometimes criticized for being too utopian, or for not being compatible with human nature, or for not being feasible or some such. Marxists are sometimes accused of failing to take into account the fact that humans are partial towards themselves in all sorts of relevant ways. Plato's requirement that parents submit their children to being raised and educated by strangers (or by the Polis) is sometimes criticized for not being sensitive to the fact that parents won't do it, and indeed, that they can't bring themselves to do it. But it is not entirely clear what to make of such objections.

Perhaps the person most critical of such feasibility requirements in the recent literature is David Estlund (from whom I borrow both examples in the previous paragraph²). And here's his most general line (2008, chapter 14; 2011; 2014), about which it seems to me that he is clearly right.

The relevant parts of political philosophy put forward normative claims. They are claims about, say, what our institutions ought to do, or what we ought to do, or what the state ought to do. Feasibility objections do not start with claims about what we (or our institutions or the state) *cannot* do. Clearly, parents *can* submit their children to being raised by strangers, and people *can* behave impartially. So "ought-implies-can" (which Estlund is happy to assume, at least for the sake of argument³) is irrelevant. Feasibility objections start with claims about what people *will* not do, or what they are *unlikely* to do, or perhaps — at

2. Of course, I do not want to commit myself to the accuracy of the depiction of Marx and Plato implicit in these examples. They are, after all, just examples.
3. E.g. Estlund (2011, 207; 2014, 116).

most — about what they *can't get themselves* to do. And even if *ought* implies *can*, *ought* clearly doesn't imply *will*, or *is likely to*, or any such thing. If I say that you ought to save the drowning child, and you respond with "but I can't!" you may very well have refuted my claim that you ought, and you certainly succeeded in undermining my blaming you for not saving him. But if you respond with "but I am highly unlikely to" or "but I'm not gonna", your response — if it is supposed to be a response at all — is barely intelligible, and it certainly doesn't refute my *ought*-statement or the blame that is likely to follow if you violate it.⁴

And so, if the question we are interested in at the moment is "How ought the state distribute resources?" then the right answer is the one that says the truth about how the state ought to distribute resources, and questions about what the state is likely to do — and in particular, about whether or not the state is likely to act as it ought — are just beside the point. Similarly, if the question is "How are we to live together in a decent society, given reasonable pluralism about conceptions of the good?", then "but we're not gonna" is no response at all to the requirements expressed by the true answer to this question.

What will emerge in the next section, though, is that there are other important questions in the vicinity here, and that this complicates matters in relevant ways. Before getting to that, though, there's a need for an important qualification.⁵

4. "The likelihood that a person will not behave in a certain (entirely possible) way simply does not bear on whether they morally should." (Estlund 2014, 122).
Perhaps there are some weird exceptions to this claim (I thank Talia Fisher for the following one): Some duties are conditional on others also doing their relevant duties. In such a case, if given a requirement to do my duty I say "I'm not gonna", this may make it the case that neither will they, and this may in turn make it the case that my own duty ceases to exist (or is undermined, or some such). The point in the text — and in Estlund — is not meant to apply to such cases.
5. I am sliding over some complications here that will not be necessary for my argument: I remain worried that Estlund's line only works *up to a point*, that perhaps *ought* does entail something like *it's not against one's nature to*. But I don't think this, even if I'm right, will matter here. And there are related

Estlund distinguishes between “aspirational theory” — the normative theory the conclusion of which is an ought statement that is insensitive to how likely it is that the relevant agent will comply — and the setting of practical goals. When it comes to the setting of practical goals, likelihood of success is of course a relevant consideration. And so, Estlund concludes: “Since the likelihood of success is (as I grant) a criterion of appropriate practical goals, ... it is a mistake to suppose that a sound standard of justice must be an appropriate practical goal” (2014, 114). This may seem odd, but it shouldn’t. That it shouldn’t may be better seen if things here are put in the context of a well-known utilitarian maneuver, that of distinguishing between the criterion of correctness for an action, and the decision procedure we are justified in employing in selecting actions.⁶ What makes an action right, according to utilitarianism, is that it maximizes (perhaps expected) utility. This doesn’t mean, though, that utilitarians recommend, as a decision procedure, to always engage in the utility calculus. Their answer to the question which decision procedure to employ in our practical endeavors is surely “that decision procedure, whichever it is, such that your employing it will maximize utility”, and what the procedure that satisfies this description is will vary from case to case, and in fact (it seems plausible to hypothesize) will hardly ever be that of engaging in the utility calculus (because that is a *very* expensive procedure to employ). Nor are utilitarians committed to the thought that we should at least try, to the extent possible, to go through the utility calculus — even such attempts at approximation may be counterproductive in utilitarian terms, compared to some other procedures. And this fact — that often the utilitarianly justified decision procedure is not that of engaging

thoughts here about not what *ought* implies but about what *blameworthy* does (perhaps, for instance, some of the considerations relevant here that don’t suffice to undermine an ought judgment undermine blameworthiness) (see Estlund 2011, 212), and about the relation between *ought to* and *can’t get oneself to* (Estlund 2011, 230 and on). These too, though, won’t matter for my purposes here.

6. See, for instance, Sinnott-Armstrong (2003, section 4), and the references there.

the utility calculus — doesn’t show that the Principle of Utility, as the criterion of correctness, is false.⁷ Now, the distinction between a criterion of correctness and the justified decision procedure is perhaps best known in the context of discussions of utilitarianism, but it is a fully general distinction, and one that any moral theory may need. And really, when Estlund distinguishes between the true aspirational theory of political justice and the setting of practical goals, he is just relying — entirely benignly, it seems to me — on a similar distinction. By putting forward an aspirational theory he is no more committed to the thought that it’s a good idea to try and use it as a decision procedure than utilitarians are committed to using the principle of utility as a decision procedure. So the fact that an aspirational theory may be hopeless,⁸ and that it may therefore not be the best guide to practical goal-setting, doesn’t show that it’s false any more than the fact that utility-counting is often a poor decision procedure (by utilitarian standards) refutes utilitarianism.

Estlund (2014, 116) calls the kind of theory that takes into account likelihood of success and that is directly tied to the setting of practical goals *concessive* (2014, 123), and he explains it partly by reference to Jackson and Pargetter’s (1986) Prof. Procrastinate case.⁹ Prof. Procrastinate receives a request to referee a paper, which he presumably should do in a timely manner. However, he also knows that he tends to procrastinate, and if he’s going to agree to referee the paper and then procrastinate, this will harm both author and journal; if he’s going to procrastinate, then, it’s better for the author and the journal if he just declines from the get-go. In such a case, what Prof. Procrastinate ought to do, it seems, is to agree to referee the paper and then do so in a timely fashion. Still, given that he won’t do it in a timely fashion (if he agrees), he should decline. Much of the discussion of the case is an

7. Whether this raises other problems for classical utilitarianism is a matter of some controversy. See, for instance, Markovits (2010, section 4).

8. Estlund (2014, 118).

9. Estlund (2014, 123–5). Estlund (2011, 216; 2014, 120–1) also rightly connects this with the problem of second-best, which I discuss in the appendix.

attempt to reconcile these last two judgments. For us, though, this part of the story is not relevant. What's relevant is that the first ("You ought to accept, and do it on time!") is the moral analogue of the political aspirational theory, and the second ("Given that you're not going to do *that*, you should at least be responsible enough to decline.") is the moral analogue of the political concessive theory.¹⁰

So, the important qualification is that while even hopelessly aspirational theories may be true, and while it is never an objection to an ought judgment that it is not going to be complied with, there is (as Estlund fully acknowledges) more worth doing than just aspirational theory, and when it comes to concessive theories, or to the setting of practical goals, feasibility considerations are going to be of central importance. This concession doesn't show that aspirational theory is not important or worth doing, of course.¹¹ But we should note (with Estlund) that there's more to political philosophy than the aspirational parts to which likelihood of success and facts about human nature are just irrelevant. There's also concessive theory well worth doing.

2. The Multiplicity of Agents and Ideal Theory

In moral philosophy (in its non-political parts) we typically ask about the principles regulating the actions (as well as other things) of individual agents. And in such contexts it's especially clear that the "But I'm not gonna" response is no response at all to an ought judgment directed at the relevant agent. Here too, of course, sometimes the actions of others are relevant (as we're about to see), and here too, even without taking into account the actions of others, there is room for concessive theory (as the case of Prof. Procrastinate clearly shows). 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*

10. It is also more "realistic", in one sense of this word, perhaps the same sense in which some political "realists" accuse aspirational theories of not being realistic.

11. Estlund (2014, 132-4).

about multiple agents. So it is natural to think that what explains the greater temptation "to withdraw a principle on the ground that it is too unlikely to be satisfied" in political philosophy compared to moral philosophy¹² is precisely the centrality of the multiplicity of agents.

In the previous section, I agreed with Estlund that the fact that an agent won't or is unlikely to act as he should is irrelevant to the truth of the judgment that he in fact should. But once there is more than one relevant agent, whether or not *one* agent will (or is likely to) act as *it* ought may be very relevant indeed to what *another* agent ought to do. This is true in non-political cases too, as long as they involve more than one agent. Suppose, for instance, that Prof. Procrastinate has an assistant, and that it's up to the assistant whether or not to accept the refereeing request. The fact that Prof. Procrastinate will not referee the paper in a timely fashion does not, we've been insisting, undermine the fact that he ought to accept and then do it in a timely fashion. But that fact *is* very relevant to whether or not *the assistant* ought to accept the request. Knowing that the professor is highly likely to procrastinate, the assistant ought to decline.

In political cases there is *always* more than one agent involved. Suppose we ask, then, how the state should allocate education resources. If a theory is offered — "The state ought to do so-and-so!" — then that the state is unlikely to do so-and-so is neither here nor there for the truth of this theory. Suppose we ask a different question — how reasonable citizens should go about educating their children, and a theory is offered: "Reasonable citizens ought to do so-and-so!" Then that reasonable citizens are unlikely to do so-and-so is neither here nor there for the truth of this theory. But that reasonable citizens are unlikely to satisfy the latter requirement may very well influence what *the state* ought to do, and the fact that the state is likely to violate the former requirement may very well affect what *the reasonable citizen* ought to do. Estlund is right in insisting that the likelihood of noncompliance *by an agent* is irrelevant to the truth of the ought judgment *about the*

12. Estlund (2014, 123).

same agent (ignoring now concessive theory in Prof.-Procrastinate-like cases). But he doesn't notice (in this context¹³) that often in political philosophy the ought judgment is about one agent, and the noncompliance is that of another. And then, his insistence against utopophobia is just beside the point.

Consider again the example of the objection to Marxism based on the (purported) natural human tendency to be partial towards oneself. If the question we ask is "How ought *we* to behave?", then the likelihood of *us* not complying, perhaps because of our tendency towards partiality, doesn't matter. But if the question we ask is "What ought *the state* to do?", then the fact that *we* — not *the state*, *we* — are partial in this way may very well be relevant. And if our question is something like "What ought the good guys to do?", then the good guys' possible noncompliance is irrelevant, but the others' noncompliance, and indeed, the state's, become potentially relevant circumstances. Indeed, I think that this way of putting things fully captures the intuitive worry that many have about, say, Marxism not being sufficiently realistic or failing to take into account facts about human nature: The problem is that regulating the actions of institutions (one set of agents) by rules that ignore the likely noncompliance of citizens (another set of agents) is unwise.

None of this is strictly speaking inconsistent with what Estlund says,¹⁴ but it nonetheless defeats a very natural and common understanding of his *point* — to the extent, that is, that his anti-utopophobia project is intended to show that worries about feasibility cannot defeat

13. Estlund does, of course, notice similar points in related contexts. Thus, the room he leaves for concessive theory is obviously close, as are his several discussions of how one agent may be required to get another agent to do something. Obviously, then, it's not as if he ignores the multiplicity of agents entirely. But the point in the text here and below stands: As a matter of aspirational theory, and without anything about one agent getting another to do anything else, Estlund's main line against utopophobia is rendered irrelevant by this kind of multiplicity of agents.

14. As Estlund confirmed in correspondence.

a normative theory in political philosophy.¹⁵ For he has been focusing on a small subset of these feasibility worries (and not the best of them either) — those where unlikelihood of compliance by an agent is taken to undermine an ought judgment applying to the same agent. But political feasibility worries are much better seen as primarily about multiple-agent cases. And with regard to these, such worries stand — the infeasibility of compliance on behalf of some agents may very well refute an ought judgment applying to others. By focusing on the fact that "I'm not gonna" is no response to an ought judgment, Estlund has been winning the battle, but he's been losing the war.

Notice that this is not a matter of doing concessive theory. The point above is not about the desirable action of an agent given that *she* won't act in the optimal way.¹⁶ Rather, the question is about how what one agent ought to do is influenced by the fact that *another* agent is unlikely to comply with requirements applying *to her*. Once multiple agents are involved, feasibility concerns take us to the discussion of ideal and non-ideal theory in the Rawlsian sense of full and partial compliance (of which I say more in the appendix).

And so, one thing we could do is straight-up ideal theory (in the sense of full-compliance theory). That is, we could ask about the required actions of one agent, in the counterfactual scenario where all other agents fully comply with all the requirements that apply to them. As I argue in the appendix, this may very well be an interesting philosophical project, even if it has no implication to practical matters in the real world. But — as I also argue in the appendix — asking about the actions required of one agent given realistic assumptions about the level of compliance of others is also a worthwhile project. And there is no

15. But see an important qualification regarding this way of understanding Estlund in section 4, below.

16. Estlund (2017) emphasizes a kind of asymmetry between concessive and non-concessive theory — the requirements of non-concessive theory do not evaporate when you comply with the relevant concessive theory (unlike in the opposite direction). Notice that according to this test too, the discussion of what one ought to do given others' failures is not concessive — no other requirement applies to the relevant agent.

obvious sense in which this project — the one of non-ideal theory — is posterior to or less respectable than the project of ideal theory.

3. The Moral Relevance of Others' Violations

Well, how *should* we take into account such expected violations?¹⁷ Let me put forward a bold answer to this question — the one I call MERELY CIRCUMSTANCES — and briefly motivate it, though a fuller account of it will have to await another occasion.¹⁸

When it comes to the moral requirements (and other moral considerations) applying to an agent, the expected behavior of third parties is always to be taken into account as another piece of the causal circumstances, in principle no different from other circumstances. In particular, this means that whether or not some expected behavior of others is in compliance with the moral requirements applying to them is intrinsically irrelevant to what ought to be done¹⁹ (it may be instrumentally relevant, of course). One way of motivating MERELY CIRCUMSTANCES is to focus on some examples where it seems to generate — in the most natural way — the intuitively right judgment. So, for instance, if you are a decent-but-not-great swimmer, and you are standing near a swimming pool, and a child is drowning, you are presumably required to jump in and save him (I'm assuming, for simplicity, that you can do

17. Valentini (2012, 655–6) asks this question in a related context, and says (following Miller) that the only possible three principled answers (with regard to the central case she considers, that of giving money to famine relief and the like) to the question in the text are: Give more than you would have if others had complied; Give as much as you would have if others had complied; and Give less than you would have if others had complied. Finding counterexamples to all these three, she concludes that no general, principled answer is acceptable here. But she is wrong, because there are many other ways of cutting up the space of possibilities here. In particular, the principle I offer in the text below survives unscathed.

18. For a detailed recent discussion of this question — one that for the most part supports MERELY CIRCUMSTANCES — see Tadros (2016). Tadros does not draw the implications from this discussion to the debate over ideal theory.

19. As Alan Patten emphasized in discussion, the moral status of another's action may yet be relevant to the *attitudes* called for. My point in the text is thus restricted to actions.

that without a serious risk to yourself). If, however, there's a lifeguard on duty, perhaps you're no longer required to jump. But if you know that the lifeguard *will not* jump (thereby violating her duty to jump in), you are again required to jump and save. The fact that the lifeguard won't jump becomes an important piece of the circumstances triggering your duty to jump, of course. But the fact that she *ought* (and so, that her failure to do so is a violation) is just neither here nor there.²⁰

This is just one example, of course.²¹ Here's another, more general way of motivating this claim, then. I call it "the appropriate question test";²² and it's designed to pump intuitions not about simple moral judgments (which action is right, which wrong), but about *comparative* ones, and in particular, about which moral differences make a difference. If someone has to decide on an action — say, whether or not to divert that infamous trolley — and she then asks us "But wait — how much suffering will diverting the trolley cause?", her question seems entirely appropriate, indicating that the amount of suffering caused by an action is a morally relevant factor, that a difference in the amount of suffering caused may very well make a moral difference. If she asks "But wait — are any of the people on the track analytic philosophers?" her question seems paradigmatically inappropriate; unless she has some *very* special story to tell about the (extrinsic) moral relevance of the answer she may get to this question, her very asking it already shows a moral failure, and this seems to indicate that whether or not the people on the track do analytic philosophy is not an (intrinsically) morally significant factor. Getting back to the swimming pool, then: suppose just before deciding whether to jump in (when you know the lifeguard will not jump) you ask "But wait — I know she won't jump; but will this constitute a *violation* of a moral norm applying to her?" I take it this is not an appropriate question to ask (and not just because of considerations of urgency). What is morally significant in the

20. For similar examples and references, see Murphy (2000, 127).

21. For another, see Tadros's (2016, 106) especially convincing Boat case.

22. See my "Intending, Foreseeing, and the State" (2007, 75).

situation is that the child badly needs help, and that he's not going to get it from anyone else. Whether that fact — that no one else will help him — is partly due to noncompliance is just beside the point. And we can generalize, applying the appropriate question test to any such case. Asking how others will behave is often an appropriate question — when others' behavior affects the consequences of one's actions in morally relevant ways. But asking whether others' behavior will constitute noncompliance is just never intrinsically relevant. Holding others' actions constant, the difference between their actions being morally permissible or wrong do not make a moral difference to the actions of others.

Let me note here two complications, for the real world is messier than philosophers' examples. First, then, third parties. I want to restrict MERELY CIRCUMSTANCES so that it only applies to the actions of *third parties*, or *bystanders*, or the *uninvolved*. Perhaps, for instance, what one agent is allowed to do to another in self-defense depends not just on what the other is doing (described in morally neutral terms), but also on whether whatever it is that the other one is doing amounts to a wrong. Perhaps whether the state is allowed to punish someone intrinsically depends also on the moral status of their actions, past or future. MERELY CIRCUMSTANCES doesn't decide such matters. Contrast such cases with the kind of case I've been discussing — that of the lifeguard, for instance. What MERELY CIRCUMSTANCES says about that case is that the way you ought to treat *the drowning child* is (intrinsically) unaffected by the moral status of *the lifeguard's* action (or lack thereof). Generalizing, MERELY CIRCUMSTANCES is a thesis about how A's way of treating B is unaffected (intrinsically) by the moral status of C's actions (but only by their causal role, as a part of the background circumstances). Unfortunately, despite the distinction between second and third parties being reasonably clear, there are complicated borderline cases. How you are allowed to treat B in helping her defend herself against C may depend on the moral status of C's attacking B. And whether a shop-owner is entitled to fire an employee for the reason that the employee's presence deters clients may depend on the *reason* for the

clientele's attitude, and so on its moral status (is it a permissible desire for better service, or a racist desire not to be served by people "of the wrong kind"?).²³ So more needs to be said about the restriction of MERELY CIRCUMSTANCES to third parties. Perhaps, for instance, in some cases — like the ones above — the way A should treat B is parasitic on the way B should treat C, and this is why A's duties towards B are sensitive also to the moral status of C's behavior — because C is the patient of B's action (and not merely a third party regarding it). And so it would be natural to restrict MERELY CIRCUMSTANCES to only cases in which the treatment of the relevant third party is not parasitic on an interaction in which that third party is more directly involved (perhaps as a patient). And there may be other complications, and other kinds of borderline cases as well. But there are also clear cases of third parties, as in the lifeguard case, and for now I am happy to restrict MERELY CIRCUMSTANCES to just those. As we will see, for the political point I am about to make, this will suffice.

The second complication it's important to note here is that in many cases, whether another's expected action constitutes a violation will be instrumentally morally relevant, or perhaps relevant in some other extrinsic way. Perhaps, for instance, we want to incentivize permissible behavior and not to incentivize noncompliance, and perhaps sometimes taking into account another's expected noncompliance as merely background circumstances will incentivize noncompliance, whereas ignoring their noncompliance and acting as if we (descriptively) expect them to comply will incentivize future compliance. If so, these may well make a moral difference (though in the lifeguard case, for instance, they will not easily defeat the reason you have to jump).²⁴

23. I thank Sophia Moreau for this example. For other, related examples of this kind, I thank Erik Zhang, Johan Frick, Larry Temkin, Mike Huemer, Jean-Christophe Bedard Rubin, and an anonymous reader for *Philosophers' Imprint*.

24. There is a common tension between the need not to accept a wrongdoing and the need to take it into account in our action. Think, for instance, about institutional arrangements (say, in tenure procedures, and in particular whether mothers should be treated differently than fathers in them) in the face of the unjust, but common, gender-biased division of domestic labor. In

But then what matters is not intrinsically the fact that the other's action amounts to noncompliance, but the consequences that this has, and so that our possible actions have – and that consequences may matter morally is something we've known for a while.

So much, then, for the (initial) positive case for MERELY CIRCUMSTANCES, the thesis that the expected behavior of others matters just like any other piece of background circumstances, and in particular, that it never intrinsically matters whether their behavior constitutes noncompliance. Before proceeding, though, let me quickly address Liam Murphy's influential rejection of MERELY CIRCUMSTANCES. This discussion is, I believe, of interest in itself, and it will also lead us into the final complication relevant to political philosophy here.

In his *Moral Demands in Nonideal Theory* (2000), Murphy argues that in beneficence cases – such as the central case of giving to famine relief – we are not required to give more than we would be required to give in the hypothetical situation where everyone gives as they are required to give.²⁵ The intuitively compelling thought is that it's unfair for the cost of someone else's noncompliance to fall on our shoulders. And of course, if Murphy is right about this, then MERELY CIRCUMSTANCES – the thesis that it never intrinsically matters whether another's action, which is to be taken into account as a part of the circumstances, amounts to a violation – is false. So something has to be said about Murphy's view.

Without pretending this is a comprehensive discussion of Murphy's ideas, let me make the following points. First, we can insist on the implausibility of his view, for instance by employing the appropriate question test directly to the kind of case Murphy is talking about. You

such cases there's a good reason to take social facts as they are; but there's also good reason not to legitimize these social facts. This conflict is real, I think, and nothing can be said in general about how to resolve it in particular cases. My point in the text is only that the latter of these two reasons is always instrumental (or at the very least extrinsic).

I thank Dani Attas, Liz Harman, Isabella Litke, and Karen Jones for related discussion.

25. See Murphy (2000, 80) for the official statement of the principle.

are considering, then, whether to benefit someone, whether to contribute more to famine relief. You know how much others will give, and this still leaves serious needs unaccommodated. In your deliberation about how much to give, you ask "But wait – the fact that so-and-so will only be giving this-much, does it constitute a violation of the moral requirements applying to him, or is it the right amount he's supposed to give?" To my ears this sounds like an inappropriate question. Surely, this is not one of the factors that serve to determine (or even to indicate) how much you're supposed to give.

Second, it is also important to note that Murphy is very clear and explicit about the restricted scope of his theory. It's not just that he's mostly interested in beneficence cases and not in other parts of morality. The crucial restriction comes from his *arguments* for the view, which rely both on the relevant moral norms being agent-neutral (as, according to Murphy (2000, 75) beneficence norms are, but other moral norms perhaps aren't), and from the project of beneficence being a collective one – the relevant moral duties, argues Murphy, are in the first instance *ours*, collectively, and only derivatively do you and I have duties as individuals here. This is why if – when we engage in this collective project – you don't do your share, it's unfair if I have to step in (2000, 76).²⁶ So even Murphy doesn't think that your real-world duties in general are exactly what they would have been under ideal theory.²⁷ Still, of course, MERELY CIRCUMSTANCES was put in full generality – and unless restricted, it remains inconsistent with Murphy's view even when his restrictions on its scope are taken into account.

But we can use Murphy's argumentation here also in order to make progress on the relevant questions in political philosophy. As noted, Murphy emphasizes the normative relationship among the members of the collective which bears, if he is right, the primary duty of beneficence. It is in this context that it's highly plausible to say that it's unfair

26. See Tadros's (2016, 110–6) refutation of the fairness point in this context (though he doesn't address the specifics of Murphy's argument here).

27. In fact, for other, non-beneficence, cases he endorses a restricted version of MERELY CIRCUMSTANCES. See Murphy (2000, 96).

if someone's failure to do their share means others have to do more. But this is not the only relevant context. Consider the relationship between the one in desperate need of beneficence and me, a member of the collective. With this normative relationship in mind, Murphy's claim seems much less compelling. Even assuming with Murphy that the beneficence duty is in the first instance the collective's, and even assuming that the fact that I'm a member of the relevant collective is normatively significant, still I am not *only* a member of the collective. I am also an individual agent, and I may have duties under that hat as well. And these are duties that Murphy does not address.²⁸

And so we get not just to the *multiplicity* of agents again, but to the complex relations between them.²⁹ We can ask, to repeat, questions such as "What ought the state to do?", "What ought we to do?", "What ought the good guys to do?", and "What ought I to do?" And we already know that while the fact that an agent is unlikely to comply with an ought judgment applying to her is irrelevant to the truth of that judgment, the fact that one agent is unlikely to comply may very well be relevant (as a part of the relevant circumstances) to the truth value of an ought judgment applying to another agent. But there's a really

28. Perhaps this is because when it comes to beneficence cases, Murphy believes – as he explained in correspondence – that *no duties apply to us as individuals*. If so, then, first, this belief itself is highly implausible, and second, even if we are willing to accept it for a very narrow set of beneficence cases, this just re-emphasizes the highly restricted scope of Murphy's thesis. It would not apply to the general political case we're interested in, for instance.

29. Estlund (2014, 125–7; 2017) does discuss such complicated relations between agents – for instance, regarding how one agent can get another to comply, and about how duties or responsibilities of collectives distribute among members. But he nowhere, as far as I know, notices how the multiplicity of agents challenges his general anti-utopophobia line.

There are hints at how the multiplicity of agents relates to the ideal-non-ideal distinction in Swift (2008, 379 and on), and in Stemplowska and Swift (2012, 388), and especially in Schmidtz (2016, 3, 6); there's some relevant discussion (but whose details at the end of the day I don't accept, for reasons that I can't get into here) in Lawford-Smith (2012). Gilabert's (2009) emphasis on feasibility issues being both dynamic and malleable is somewhat related as well. And Stemplowska (2016) offers a discussion of collective and individual feasibility that addresses some related complexities, but not the ones directly relevant to my point here.

important complication here, having to do with the close relations between these different agents, or between what we may call *entangled* agents. I, for instance, am the agent about whom the question "What ought I to do?" is asked. I am also, however, a good guy, so the question about the good guys is partly relevant to me as well, as is the question "What ought we to do?" (seeing that I'm a part of "we"). So if, for instance, you know that I'm not going to act in the way that "we" ought to act (or to do my part so that we act in the way we ought to act), we're in an important mixed case: *We* and *I* are distinct agents³⁰ – so one is tempted to apply MERELY CIRCUMSTANCES, and to say that whether or not *we* ought to do something may depend on whether or not *I* do as I ought (as a part of the relevant background circumstances). But because of the close relation between the two agents – because they are entangled – we're close to Estlund's observation that "But I'm not gonna" is no response at all. And of course, when it comes to questions about the state, the relations between its agency and that of others (including individuals) become even more complicated (and interesting).

What can we do with such entangled-agency cases? Not much more, I think, than distinguish different questions, and then proceed carefully regarding the relations between them.³¹ I have duties as an individual, but also as a member of numerous groups and collectives that themselves have duties, and perhaps also as a citizen (and so as someone whose agency is connected in interesting ways to that of the state). For each specific question, the general claims defended above stand – that I am unlikely to comply is irrelevant to whether I ought to comply, and that other agents – even ones I am related to – are unlikely to comply is a part of the possibly morally relevant background circumstances. But if I also have a key role to play in what the state

30. If *we* are an agent at all, which I am assuming here. Clearly, as Seth Lazar noted, sometimes a question like "How are we to live together?" is not about the agent *we*, but the agents of which it is composed.

31. This is not a case of "saying more will take me too far afield", or some such. I don't think there *is* more to say here, at least nothing more that is both general and informative. (I thank an anonymous referee for pressing me on this, though I am here rejecting his invitation to say more.)

does, or in what we do, then this too is an aspect of my behavior to be taken into account. At the end of the day, in such cases what I ought to do — all things considered, all-out ought to do — will be a complex and messy function³² of all of these normative considerations applying to me.

The crucial thing to ask, then, in determining whether the likelihood of noncompliance matters to the moral status of an action, is *whose* noncompliance and *whose* action.³³ If the answer to these two questions is the same agent, then Estlund is right, noncompliance is not relevant at all, and in that sense we should reject utopophobia. If the expected noncompliance and the action about whose moral status we're asking are of different agents, then the noncompliance matters as a part of the relevant ordinary, causal circumstances. And often — for instance, but not only, in political cases — more than one question will be relevant, and so messy, complicated answers will be called for.

A corollary follows: If there are parts of normative political philosophy that do not issue directives addressed at any agent, or that are not about the evaluation of any agent's activities, then to those parts of political philosophy *everyone's* expected behavior is relevant (because there's no agent whose "I'm not gonna" response is relevant). And there may be such cases in political philosophy. Perhaps some questions about how things should be organized, or about what our institutions should be like, or perhaps about social design (asked in a way that doesn't assume that the social designer is itself an agent)³⁴ — such questions may be of central importance in political philosophy, but there's no specific agent they are about. So there's no one whose "I'm

32. This may include not just balancing of reasons, but also undercutting, excluding, and so on.

33. Estlund (2011, footnote 24 on 229) rightly rejects the claim that a normative theory that is unlikely to be complied with is not suitably action-guiding. But he doesn't address the obvious complication — the fact that a normative theory *applying to one agent* may fail to be action-guiding *for him or her*, if it falsely assumes the compliance of *other* agents.

34. I thank Philip Pettit for emphasizing to me the importance of this kind of question here.

not gonna" is irrelevant — any foreseen action by any agent enters the set of possibly relevant circumstances.

4. Back to Politics

Get back, then, to the Marxism case, and to the relevance of the purported facts of human nature that make compliance with Marxist requirements unlikely. Is this a problem for the relevant normative political theory?

Well, the answer depends on what exactly it is that the theory *says*. If it's just an answer to the question "How ought we to act?" then the fact that we are unlikely to comply is neither here nor there — here it's the same agent occupying both positions, both that of the agent at whom the requirement is addressed, and the one whose compliance is unlikely. But if the theory offers also answers to other questions — for instance, "What ought the state to do?" — then likely noncompliance (of other agents, like you, or me, or us) is very relevant indeed, just as any other part of the background circumstances may be. And notice that such other-agent-noncompliance is relevant to aspirational theory, not just to concessive theory. For when we take background circumstances into account to determine what an agent's duties are we do not engage in concessive theory, we are not in the business of finding out a second-best taking a violation (by the same agent) as a given.

The multiplicity of relevant agents (and the complex relations between them) thus vindicates (with Estlund) both the genuine normativity of normative theories in political philosophy, and (pace Estlund) the ways in which noncompliance may be very relevant even to aspirational theory, and so a problem for the truth of a normative political theory, not just for its implication.

Above I said that whether likely noncompliance is a problem for the toy-Marxist theory depends on what the theory is, on which questions it answers. But at the end of the day we are not interested in interpreting a theory in political philosophy. So the question becomes — what questions *should* we be interested in when doing political philosophy? If, for instance, there is only one agent about the actions of which

political philosophers should ask (“society”, perhaps?), then much of the discussion above becomes irrelevant.

But this, it seems to me, would be a highly impoverished picture of political philosophy. Political philosophers should be asking questions about a whole host of agents – society, perhaps; the state, certainly; us; and just as importantly, questions about what the good guys should do (taking into account background circumstances, circumstances that include facts about violations by non-good-guys), and questions about what individuals ought to do – you and I, for instance (taking into account the expected behaviors of others, noncompliance included).

Of course, political philosophers can acknowledge that all these questions – about different agents – are worth asking, but focus in their own work on just one of them. This is one possible understanding of Estlund – he doesn’t have to deny the point about the multiplicity of agents and the relevance of the expected violations of one agent as background circumstances for questions about how another agent ought to act. It’s just that in his own work he focuses on just one agent – society – and insists (correctly) that possible violations *by this agent* do not refute ought judgments *applying to it*.³⁵ It’s just that then, the limitations of such work should be clearly acknowledged – infeasibility or partial compliance considerations are still very relevant to all the multiple-agent cases, and so they may well refute a non-concessive normative theory in political philosophy. After all, as I’ve argued, one agent’s (foreseeable) noncompliance may partly determine another’s duty in an entirely non-concessive way (your duty to jump when you know the lifeguard won’t is not a matter of concessive theory – it can be as hopelessly aspirational as any). And no reason has been given (as far as I know) to focus on or privilege just one-agent cases, and in particular, one-agent cases with *this agent*.

One of the things we should do as political philosophers is non-ideal theory. When we do, often we should answer normative questions about multiple, varied, and entangled agents. A political philosophy

35. I thank David Estlund for clarifying this in correspondence.

that has nothing to offer by way of an answer, say, to “What ought I do to?” questions asked in political contexts, is hardly worth its name.³⁶ Certainly, the parts of political philosophy that are arguably relevant, say, for political science, and for law – and presumably, there are such parts of political philosophy – have to be sensitive to facts of expected noncompliance. In this way, feasibility considerations may very well be relevant to political philosophy, and some healthy aversion to utopianism may be in place.

Appendix: Ideal and Non-Ideal Theory

The terms “ideal theory” and “non-ideal theory” have unhelpfully come to stand for several different things,³⁷ but perhaps the clearest of them, and the one most clearly coming from Rawls,³⁸ is that in terms of full compliance. Suppose we’re trying to determine the appropriate traffic laws and regulations for a given society. One thing that we may ask is what’s the optimal regulation of traffic, assuming all will comply with the regulation (about the content of which we are now asking). Perhaps, for instance, the right tradeoff between safety, convenience,

36. I sometimes hear it said that political philosophy is *philosophy*, so it is concerned with truth, not with practical recommendations. But the suggested dichotomy is illusory – of course political philosophy is and should be concerned with truth. It is concerned with finding the *true* answers to several kinds of questions, including such questions as “What ought I to do?”.

37. Again see Stemplowska and Swift (2012, 374), and Valentini (2012).

For reasons that are not entirely clear to me, Cohen’s (2002; 2008) claims about the fact-insensitivity of basic normative principles became entangled with discussions of ideal theory (see, for instance, Valentini 2012, 657, and the references there). I think that this is just a misunderstanding. Cohen’s thesis is a *metaethical* one, and is consistent with any view in the ideal theory debate understood along the lines discussed in this paper. For this reason, I will not discuss fact-sensitivity in Cohen’s sense in this paper at all.

For the claim that the whole fact-sensitivity debate is just another matter altogether, not closely related to the discussion of ideal and non-ideal theory, see Stemplowska and Swift (2012, 383) and Estlund (2014, 129–30).

38. Though let me emphasize that I make no exegetical claims about how best to read Rawls. For instance, Rawls may be packing much more than just compliance levels into his understanding of ideal theory. (I thank Arthur Ripstein and Seth Lazar for related points.)

environmental considerations, and perhaps other considerations requires that people not drive over 110 kilometers per hour on the highways. If so, assuming full compliance, the speed limit ought to be set at 110 kph. But perhaps we know — on empirical grounds — that many are likely to drive somewhat over the speed limit, perhaps roughly at 10% over the speed limit. In that case, it may be better to set the speed limit at 100 kph.

This is a toy example, of course, but it suffices, I think, to explain the distinction between ideal and non-ideal theory. Ideal theory asks about the appropriate arrangements (traffic regulations, basic social institutions) under the assumption of full compliance with those arrangements. Non-ideal theory takes into account information about noncompliance, and asks about the best arrangements given the empirically plausible partial compliance.

The distinction between ideal and non-ideal theory in terms of full and partial compliance is neither dichotomous nor one-dimensional. For one thing, different levels of compliance may be invoked. And the distinction may be drawn at different levels of generality.³⁹ It can be very specific, as in the case of our toy example, where full compliance is understood as full compliance *with just the speed limit rule*, or perhaps with traffic laws and regulations more generally. It can be general, as when we ask about what justice requires vis-à-vis our basic social institutions, *assuming that all act in accordance with all of justice's requirements*. And it can occupy any number of intermediate positions: to use one of Rawls's own examples,⁴⁰ the theory of criminal punishment is, of course, a part of non-ideal theory in the most general sense, because under the assumption of full compliance with political justice and perhaps with morality as well, no one should ever be punished (because no one ever commits a crime). But we can still usefully and importantly distinguish between ideal and non-ideal theory of criminal punishment in a more specific, local sense — for instance,

39. See on this Stemplowska and Swift (2012, 385).

40. Rawls (1999, 8).

we can ask about the appropriate rules regarding punishment under the assumption that state officials will follow them to the letter, or ask about the appropriate rules assuming state officials too make mistakes, sometimes have ill will, may abuse their power, and so on.⁴¹ The availability of such intermediate positions — idealizing on some but not all compliance — allows us, I think, to see Anderson's diagnostic conception of non-ideal theory as a particular instance of non-ideal theory as understood here. When Anderson insists that we should start with a diagnosis of a social wrong, and see how we can improve things in that specific regard,⁴² she can be seen as insisting on taking much background noncompliance as given, and dealing with more specific noncompliance. Despite the importance of such intermediate positions, in the main text of this paper, I for the most part speak of ideal and non-ideal theory as if the distinction is dichotomous and one-dimensional, allowing the context to determine further details (to the extent that they are needed).

The distinction between ideal and non-ideal theory understood in terms of full or partial compliance is thus a distinction between answers to different questions,⁴³ roughly: What should we do (perhaps in a given domain) assuming that all involved will act as they should; and what should we do (perhaps in a given domain) taking into account that people will act as the evidence indicates they are likely to act, where this includes some violations of these rules.

If the distinction between ideal and non-ideal theory is the distinction between two kinds of question, then the question whether we should be doing ideal or non-ideal theory is the question which kind of question we should be asking. But then it seems clear how we should respond: We should be asking both, and you, in particular,

41. For this example in a closely related context, see Estlund (2014, 132).

42. "Nonideal theory begins with the diagnosis of the problems and complaints of our society and investigates how to overcome these problems" (Anderson, 2010, 6).

43. I think it's safe to say that many now see this distinction in this way. See, for instance, Valentini (2012, 660), Arvan (2014), Nili (manuscript), and Mason's (2016) support of pluralism about different good questions to ask.

should be asking the one you are more interested in. After all, many questions are worth asking, and noticing that one question is worth asking doesn't entail that another is not. We are doing philosophy, not traffic regulation. Perhaps in the case of traffic regulation ideal theory is not worth doing. But in even just slightly more abstract contexts, there are interesting questions both about full compliance scenarios and about partial compliance ones (as I hope the example of criminal punishment theory above shows). And when doing philosophy, that a question is interesting is all that is needed to justify engaging it. What, then, is all the fuss about?

Still, I think more needs to be said. In the remainder of this appendix I quickly go through several possible interpretations of claims about which kind of theory we should be doing, finding none of them satisfactory. This leads up to the discussion of the purported *priority* of ideal over non-ideal theory, which concludes this appendix.

When some insist that we should be doing ideal theory, and some criticize them for being wrong about this, what is the debate about?

Perhaps the thought is that ideal theory, or perhaps non-ideal theory, is not worth doing. Now, perhaps this is sometimes what the *intonation* of such disagreements seems to indicate, but it is hard to see why we should accept this suggestion. Many things are worth doing. There needn't be a competition in this respect between ideal and non-ideal theory.⁴⁴

Perhaps the disagreement is about the role of the political philosopher, or of political philosophy. Perhaps, that is, both questions are worth doing, but only ideal theory is worth doing *by political philosophers* (perhaps relegating non-ideal theory to social scientists, perhaps with the guidance of political philosophers on abstract normative principles⁴⁵). The thought seems to be that the only thing that

44. "From an outsider's perspective, the ideal/non-ideal theory debate looks more like a dispute about what kind of theorizing is worthwhile doing than a competition between genuinely competing projects." (Ismael 2016, 31)

45. That those engaging in social engineering need normative guidance that it's not their professional expertise to investigate is a major point in Swift (2008).

genuinely counts as political philosophy is ideal theory.⁴⁶ But this can't be right — the debate is supposed to be substantive, not terminological.⁴⁷ And more than the non-existent job description of political philosophers must be at stake. Perhaps the point is about relative advantages — perhaps the relative advantages of the philosopher (compared to empirically minded social scientists, say) make her more suitable for ideal theory. This may be so — it won't be a huge surprise to find out that philosophers are not as good with facts as some social scientists are. But still, this doesn't show that political philosophers should never take into account *some* facts, like facts about partial compliance.

Relatedly, the debate over ideal theory may be a debate about interpreting the canon of political philosophy — how are the texts of the great dead philosophers in political philosophy best interpreted, as engaging questions of ideal or of non-ideal theory? This suggestion is related to the previous one, because it's natural to think that the definition of a sub-discipline is, if it exists, a function of its canonical texts. Let me concede that it's an interesting question to ask about a text, or a thinker, or even a loosely defined body of literature, whether it is best understood as engaging ideal or non-ideal theory. In this respect, then, this suggestion is not without merit. But only in this respect — in others, it is very problematic. After all, there is no reason to suppose that the canon speaks in one voice on this. Furthermore, delineating the canon is going to be neither easy nor uncontroversial, and will likely depend to an extent on one's view regarding the right way to do political philosophy — and is thus likely to be, in the context of the discussion of ideal and non-ideal theory, at least somewhat

46. In the opposite direction, Frazer (2016) argues that it's a part of the job description of the political philosopher to help us navigate the problems we face in actual political life, and so that only doing ideal theory doesn't live up to the moral demands of the vocation of political philosophers. I am *much* more sympathetic to this way of viewing things. See my "Against Public Reason" (2015, 134–7). But I think that — viewed as a way of deciding questions of priority between ideal and non-ideal theory — it is vulnerable to the objections I am about to get to in the text.

47. For a related point, see Estlund (2014, 130–1).

question-begging. And of course, even if it can be shown that “the canon” in political philosophy engages ideal theory to the exclusion of non-ideal theory (or the other way around), still nothing would follow about what we should be doing. Perhaps, indeed, the fact (if it is a fact) that the canon only engages ideal theory is a good reason to reject the canon, and to radically reform political philosophy.

Even if there is no direct competition between ideal and non-ideal theory — for more than one kind of question may be worth asking — still, there may be a competition in terms of the allocation of the limited resources of political philosophy. And so perhaps claims about whether we should be doing ideal or non-ideal theory are best understood as claims about the allocation of those resources — from the point of view of political philosophy central command, as it were, should more philosophers, or more of their time, or more research budget, or more entries on first year intro to political philosophy syllabi, be devoted to ideal theory or to non-ideal theory, or to some mix thereof? And if a mix, what mix exactly? Once again, such questions are not without merit. But this cannot be what the discussion is about, I think, because the discussion is not held from the point of view of central command, or even that of research administrators. And from the point of view of philosophers, talk of such allocation of resources is, at least to an extent, beside the point. Philosophers engage — and to a large extent should engage — the questions that they are interested in. Furthermore, if we do attempt to occupy the point of view of philosophy central command, all sorts of surprising conclusions may arise — perhaps, for instance, both ideal and non-ideal theory in political philosophy should get a larger share of the relevant resources, at the expense, say, of analytic metaphysics. Or perhaps both should get much less, because resources should be diverted to the philosophy of language. Surely this is not the kind of thing the ideal theory debates are about.

So much, then, for my attempts at understanding thoughts about whether we should be doing ideal or non-ideal theory. But Rawlsians — and Rawls — rarely deny that non-ideal theory should also be

engaged (by political philosophers, if need be). Rather, they insist that ideal theory enjoys a kind of priority, that ideal theory is, as it were, where we should start. The Rawlsian thought seems to be that we should do ideal theory — first, or mostly — because non-ideal theory is in an important way parasitic on ideal theory, or because ideal theory is needed in order to guide non-ideal theory, or some such.⁴⁸

But this idea of priority is not transparent. It is unclear and ambiguous. What, then, can be meant by the claim that ideal theory is prior to non-ideal theory?⁴⁹

(i) *Epistemic Priority*

One natural thought is that the only way to come to know truths of non-ideal theory is to first come to know ideal theory. The thought seems to be that noncompliance is a complication, and that the way to get at truths that take it into account is to first get to truths that abstract away from noncompliance, and then add to the theory whatever fix is needed to deal with the complication.

Natural though this line of thought is, it is clearly false. One point here comes from Sen (2006) — it is just not true, to use his famous example, that in order to know which of two given mountains is higher we need to first know which other mountain is the highest one on earth. Analogously, it is at the very least non-obvious whether the only — or even the best, or even a good — way of coming to know about the way to proceed in matters of political justice given realistic levels of (non-)compliance — say, coming to know which of two distributive arrangements is better — is to first come to know ideal theory (including what the best distributive arrangements, under full compliance, is).

48. See, for instance, Brennan and Pettit (2005, e.g. 259, 263), and the references there.

49. Levy (2016) also distinguishes different kinds of priority, but his distinction is somewhat different from mine (partly because he doesn't explicitly distinguish instrumental from other considerations and priorities).

The argument here is not merely one by analogy from a convincing example.⁵⁰ The point is more general, and it is the one highlighted by the general theory of the second-best.⁵¹ We know, of course, that from something being an optimal solution to a choice situation under certain conditions, it doesn't follow that it's the optimal solution under other conditions. We likewise know that if a solution is optimal under a set of conditions, and one of the conditions is not met, it doesn't follow that we should approximate that solution according to some natural proximity metric. Indeed, we know that there is no general content-independent way of determining what the second-best option is, just based on knowledge of the best option, and of the fact that some condition necessary for its attainment is not satisfied. Perhaps the best thing for you to do now is to cross the street over to the other side, but if for some reason you can't get all the way to the other side, it doesn't follow that you should "approximate" crossing the street, getting as close to the other side as you can. Whether this is a good second-best depends on the specific characterization — both descriptive and normative — of the options and situation, not on formal proximity metrics.⁵²

50. Which is a good thing, because the force of the example is limited. Perhaps — as Chaim Gans suggested — a better analogy would be one where knowing the ideal is analogous not to knowing what the highest mountain is, but rather what being high consists in. Or perhaps — relatedly, and as Wayne Sumner suggested — a more suitable analogy here is with comparing two bodies asking which is more spherical, a comparison that does seem to be parasitic, in some way, on the perfect sphere. I think that both these suggestions are closer to thoughts about explanatory priority, which I discuss below.
51. The locus classicus is Lipsey and Lancaster (1956). For discussions in more closely related contexts, see Margalit (1983), Goodin (1995), Brennan and Pettit (2005, 260), and Raikka (2014).
52. We should distinguish between knowing that Option2 is the second-best, and knowing that Option2 is better than Option3. Of course — as an anonymous referee reminded me — In order to know of a specific option — say, Option2 — that it's the second-best option, at least in standard cases we must know what the best option is. (The qualification to standard cases is needed. There are cases — so-called "Jackson-cases" (Jackson 1991, 462) — where I know of two options that one of them is best but I don't know which, and I know of a

What this means — and this is a point often officially acknowledged by proponents of ideal theory, even if they seem to forget it in the heat of discussions of the real world⁵³ — is that nothing follows about the real world of partial compliance from even an infallible ideal theory. Perhaps ideal theory shows us what is, in some way, the best option. But because one of its conditions (full compliance) is not satisfied, we're going to have to go for second-best (at most), and whether ideal theory can be helpful in determining what *is* the second-best option remains to be seen.⁵⁴

Notice that this remains so under the Rawlsian-friendly conception of political justice being grounded in relations of reciprocity (rather than, say, in the promotion of wellbeing). According to such a conception, the level of compliance of others immediately affects what one is required, as a matter of justice, to do. And so, if we are to know what one is required to do, we must first know or be able to find out what others' level of compliance is. And in order to know this, we must first do ideal theory — only given ideal theory, can we determine how far others' behavior falls short.⁵⁵ But in order to know how far others' behavior falls short — in a situation S_1 — from what is required of them, what we need to know is what is required of them *in* S_1 , not in some other situation (for instance, of full compliance). Suppose, for instance, that you and I agreed to come forward and complain to HR about an abusive boss. Perhaps I am only under a duty to do so if you go forward as well (and vice versa). Reciprocity is here central. But what's crucial for reciprocity is that you are also under a duty to

third option that it's the second-best, whichever of the first two is best). But the point in the text here relies on the much weaker point, namely, that a comparison between two less-than-perfect options doesn't require, and often is not helped by, knowledge of the best option.

53. See my (2015, 125), and the references there.
54. In unpublished work, David Estlund refers to the failure to notice the point in the text "the approximation fallacy".
55. I thank an anonymous referee for this suggestion.

come forward *in the very nonideal situation we're in*, not what duties you would have had under full compliance (with a non-abusive boss, for instance).⁵⁶ Even assuming a central role for reciprocity, then (an assumption that is, of course, neither obvious nor uncontroversial) it is not clear how ideal theory is supposed to be epistemically prior to nonideal theory.

I don't want to overstate the point. Nothing that's been said rules out the possibility that in some cases ideal theory may be epistemically helpful in doing non-ideal theory. It's just that there is no guarantee that it will be. In general, ideal theory is neither necessary nor sufficient for doing non-ideal theory. Whether ideal theory is helpful in a specific area will have to be established in a way that's peculiar to the contents of the relevant ideal and non-ideal theory. If you want to argue that in the theory of political justice ideal theory has this kind of epistemic priority over non-ideal theory, you just have to *show* how epistemically poorly non-ideal theory is doing without the guidance from an ideal theory, and how much such guidance can help.⁵⁷ I don't know of anyone who has shown this, and I have no reason to believe that this is true.⁵⁸

56. See also the point above about the distinction between ideal and nonideal theory being non-dichotomous.

57. Levy (2016) makes a similar point.

58. I don't know of any suggestions along these lines, but because they're there in logical space, let me mention two more possibilities here.

One other kind of priority is *metaphysical*. It may be argued, then, that ideal theory is metaphysically prior to non-ideal theory, perhaps that ideal theory *grounds* non-ideal theory, that truths of non-ideal theory are true partly *in virtue of* truths of ideal theory. This may be an interesting line to pursue, perhaps especially given the current explosion of writing on grounding. But I don't know how to develop it further, so I'll leave it at that.

Another kind of priority is *moral*. One may think, perhaps in a Kantian-inspired way (I thank Dani Attas, Larissa Katz, and an anonymous referee for suggestions along these lines) that we morally owe it to our fellow citizens not to believe ill of them, and so not to assume less than full levels of compliance with the demands of justice. But I don't think this will work, first, because for many purposes we most certainly should believe ill of our fellow citizens (when the evidence so indicates), and so what the defender of ideal theory needs is not an argument against considering failures, but rather for

(ii) *Instrumental Priority*

Thoughts about the priority of ideal theory may be instrumental: they may be, that is, not necessary claims about some justificatory relations between ideal and non-ideal theory, but contingent claims about how we are likely to succeed better in doing non-ideal theory — namely, by already having ideal theory at our disposal.

But such a suggestion seems *too* contingent, and too empirical as well.⁵⁹ On this suggestion, we would simply need to check — if interested primarily in non-ideal theory, how instrumentally useful is it to do ideal theory first? How does it compare to other ways of investing our philosophical resources? Perhaps, for instance, in some cases doing ideal theory is actually counter-productive, drawing attention away from what really matters in the world, things like the need to fight such gross noncompliance as racism and oppression? Perhaps doing ideal theory can actually help in doing non-ideal theory in some domain or context, but doing sociology can help more? And perhaps in some cases really the best instrument to non-ideal theory (perhaps alongside others) is ideal theory. Or perhaps doing ideal theory has other instrumental advantages as well. There is no general, a priori way to tell. The champions of the priority of ideal theory seem to want much more.⁶⁰

a principled distinction between when we should and when we shouldn't; and second, because the idea that we owe it to someone to believe against the evidence is deeply puzzling. I discuss it in the context of a possible understanding of paternalism in "What's Wrong with Paternalism" (2016), and in the context of a discussion of the public reason tradition in my "Political Philosophy and Epistemology: The Case of Public Reason" (2017).

59. Perhaps this is what Schmidtz (2016, 10) has in mind when he refers to one of Rawls's claims about the priority of ideal theory as a "testable factual claim".

60. Simmons (2010, 15) designates, perhaps following Rawls, for ideal theory a privileged instrumental role, according to which in non-ideal theory violations of liberty are only permissible if "part of a process aimed at achieving the fullest possible societal justice." But though it is natural to think of such instrumental considerations as significant — that one non-ideal arrangement is more likely than another to lead to the fully just society surely counts in

(iii) Methodological Priority

At times it seems like the purported priority of ideal theory is neither metaphysical or epistemological nor instrumental but rather *methodological*. But what could this mean, if it is not another way of specifying either epistemic or instrumental (or perhaps explanatory) priority?

I have to confess that it is not clear to me what such methodological priority can come to. Perhaps the thought is *educational*. Think of the way youngsters are introduced to Newtonian mechanics by assuming frictionless surfaces. This seems like good practice, even though — of course — no such surfaces exist, and if you want to do Newtonian mechanics for the real world you're going to have to take friction into account. Presumably, this is nonetheless good practice at least partly because frictionless mechanics is a good heuristic device on the way to doing the much more complicated real-world mechanics. Perhaps, then, something analogous can be said about ideal theory's priority in political philosophy?

But first, it's not clear that the two cases are sufficiently similar, and second, it's not clear what the priority in the case of frictionless mechanics comes to. To start with the latter: it seems that the methodological priority of frictionless mechanics is to be understood at least partly as a combination of instrumental considerations (it's really hard to do physics, and a good, gradual way of introducing students to the topic and helping them develop the required skills is to start with frictionless mechanics), with perhaps epistemic and metaphysical ones as well (perhaps, for instance, truths of mechanics are grounded in truths of frictionless mechanics). It's not clear that more is involved in this kind of methodological priority, then, on top of the kinds of priority discussed elsewhere in this section (though I qualify this claim below). And, to return to the adequacy of the frictionless mechanics

favor of that arrangement — there is absolutely no reason to privilege this instrumental consideration compared to many others.

Against Utopianism: Noncompliance and Multiple Agents

analogy — while it is clear that doing frictionless mechanics is very useful — perhaps necessary — en route to doing more complicated physics, it is not at all clear that doing ideal theory has a similar status vis-à-vis non-ideal theory (as was argued in previous sections). So it's not clear that analogy holds.

(iv) Explanatory Priority

Perhaps the most interesting suggestion is that ideal theory is *explanatorily* prior to non-ideal theory. Perhaps, in other words, non-ideal theory is parasitic on ideal theory — perhaps in order to understand the requirements of justice in conditions of partial compliance it is necessary to first understand ideal theory.

Of course, the relevant notion of explanatory priority is not entirely clear (for one thing, it may take us back to something like Sen's mountains example and the objection that it grounds). I think that points about *conceptual* priority are also along these lines — that non-ideal theory is somehow parasitic on ideal theory, that ideal theory is the basic case, that perhaps the very concept of non-ideal theory requires the concept of ideal theory (as the concept of an attempt requires that of an action) but not the other way around. Despite some unclarity, then, there does seem to be something interesting and potentially important in the vicinity here. But it doesn't seem to be what proponents of the priority of ideal theory have in mind. For the explanatory priority claim to be plausible what is needed is also, somewhat vaguely, the sense that the problem we're after is at least in outline the same problem. This, it seems to me, is plausible in the case of frictionless mechanics. When you do frictionless mechanics, you often get an "aha" feeling of understanding all sorts of real-world phenomena as well. The no-friction idealization seems to let us see better other, deeper, regularities and explanations, rather than to obstruct our view.⁶¹ Something similar must hold for ideal theory and non-ideal

61. See Ismael (2016) for an extended discussion of such cases (like Newton's ideal pendulum) and the analogy between them and ideal theory in political

theory if the explanatory priority claim is to stick. And I'm just not sure this is the case — at least not in general. Sometimes, the problem we are interested in is *entirely due to* the noncompliance ideal theory assumes away.

Think, for instance, of just war theory⁶². We could embark on ideal just war theory, but this would be a very short project. Obviously, some fairly strong form of pacifism is the true *ideal* just war theory — such a theory, under full compliance, would be hard to reject. But this just shows that just war theory — the kind of stuff that you know from the literature — is *essentially* non-ideal, that the very problems just war theory attempts to answer have to do with noncompliance (it should be obvious, at this point in the paper, that they are also essentially multiple-agency cases)⁶³. Similarly, presumably, for a theory of criminal punishment. So it doesn't seem plausible that ideal just war theory has the kind of explanatory priority over non-ideal just war theory, for the latter's main concern is precisely to answer the question that the former assumes away.

If so, what we must now ask is whether when it comes, say, to basic social institutions, or to political legitimacy, or to distributive justice, the very problems addressed presuppose noncompliance, or whether, when we assume full compliance, the problems that remain still seem like essentially the same problems that are there under partial compliance, or their close relatives.

philosophy. Ismael is clearly interested in *a* sense in which ideal theory (in both domains) has priority over non-ideal theory, but she's not very clear about which kind it is. I think that at the end of the day, her view is close to the one sketched in the text here.

62. Again, an example that Rawls himself cites in this context. See Rawls (1999, 8).

63. This is consistent with interesting versions of just war theory that are *partly* ideal — say, that take the occurrence of wars as a (non-ideal) given, but that idealize away abuses of the rules of engagement, etc., analogously to the point about different intermediate positions with respect to criminal justice, earlier in the text.

I think that the answer is somewhere in between these two options. Unlike with just war theory, *some* relevant problems remain even under full compliance. Ideal just war theory is obvious and boring, but questions about, say, collective decision-making or resource allocation remain interesting even under the assumption of full compliance. Furthermore, to an extent they do feel like close relatives of the questions we ask under partial compliance. To an extent, then, there is some plausibility to the thought that ideal theory has explanatory priority over non-ideal theory. But only to an extent, because it seems to me clear that in many quite central issues in political philosophy — perhaps even regarding basic institutions — noncompliance is partly definitive of the questions we try to answer: Think, for instance, about democratic theory. And this means that perhaps here too we should proceed on a case-by-case basis, checking each specific instance of an ideal theory to see whether it has abstracted away from the very problems we are interested in solving. For each such case we must ask, in other words, what question exactly we are asking, and how it is influenced by idealizing away noncompliance.

None of these kinds of priority can give plausible content to the thought that ideal theory is prior to non-ideal theory in a way that will give the supporters of this claim what they seem to want.

You can do, then, ideal theory, as it may be interesting in its own right. But then you shouldn't pretend you've been saying anything about the real world and its politics. And you also can — and sometimes should — do non-ideal theory, which is in no obvious way less respectable than ideal theory, or indeed posterior to it. When you do that, and when you engage in multiple-agent cases (as you almost always do, when you're doing political philosophy), the noncompliance of some may affect the duties of others. And it may do so in a way that defeats utopianism.⁶⁴

64. For helpful comments and discussion I thank Marcus Arvan, Dani Attas, Jean-Christophe Bedard Rubin, Hanoch Dagan, David Estlund, Talia Fisher, Chaim

References

- Elizabeth Anderson (2010), *The Imperative of Integration* (Princeton: Princeton University Press).
- Marcus Arvan (2014), "First Steps Toward a Nonideal Theory of Justice", *Ethics and Global Politics* 7, 95–117.
- Geoffrey Brennan and Philip Pettit (2005), "The Feasibility Issue", in Frank Jackson & Michael Smith (eds.), *The Oxford Handbook of Contemporary Philosophy*. Oxford University Press 258–279.
- Gerald Allan Cohen (2002), "Facts and Principles", *Philosophy and Public Affairs* 31, 211–245.
- (2008), *Rescuing Justice and Equality* (Harvard University Press).
- Eva Erman and Niklas Moller (2015), "Practices and Principles: On the Methodological Turn in Political Philosophy", *Philosophy Compass* 10/8, 533–546.
- David Enoch (2007), "Intending, Foreseeing, and the State", *Legal Theory* 13, 69–99.
- (2015), "Against Public Reason," *Oxford Studies in Political Philosophy* (Volume 1), D. Sobel, P. Vallentyne, and S. Wall (eds.), Oxford: Oxford University Press, pp. 112–142.
- (2016), "What's Wrong with Paternalism: Autonomy, Belief and Action", *Proceedings of the Aristotelian Society* 116, 21–48.
- (2017), "Political Philosophy and Epistemology: The Case of Public Reason," *Oxford Studies in Political Philosophy* (Volume 3), D. Sobel, P.

Gans, Alex Guerrero, Mike Huemer, Pini Ifergan, Seth Lazar, Liam Murphy, Philip Pettit, David Plunkett, Ariel Porat, Doug Portmore, Arthur Ripstein, Hamish Russell, Reem Segev, Amy Shuster, Wayne Sumner, Larry Temkin, three readers for *Philosopher's Imprint*, and audiences at Princeton, Jerusalem, Bar Ilan, Tel Aviv, Toronto, McMaster, and Boulder. I want to thank especially – for more than just their comments – Amy Shuster (for making her entire Dropbox folder on ideal and non-ideal theory available to me when I was starting to do the reading relevant for this paper), David Plunkett (for a lot of encouragement regarding this project), and Dave Estlund and Liam Murphy (for being so generous with their time with someone who was out to reject their views).

My research was supported by the Israel Science Foundation, grant 439/15.

Against Utopianism: Noncompliance and Multiple Agents

- Vallentyne, and S. Wall (eds.), Oxford: Oxford University Press, pp. 132–165.
- David Estlund (2009), *Democratic Authority: A Philosophical Framework* (Princeton University Press).
- (2011), "Human Nature and the Limits (If Any) of Political Philosophy", *Philosophy and Public Affairs* 39 (2011), 207–237.
- (2014), "Utopophobia", *Philosophy and Public Affairs* 42, 113–134.
- (2017), "Prime Justice", in Weber and Vallier (2017), 35–56.
- Michael Frazer (2016), "Utopophobia as a Vocation: The Professional Ethics of Ideal and Nonideal Political Theory", *Social Philosophy and Policy* 33, 175–192.
- Pablo Gilibert (2009), "The Feasibility of Basic Socioeconomic Human Rights: A Conceptual Exploration", *The Philosophical Quarterly* 59, 659–681.
- Robert E. Goodin (1995), "Political Ideals and Political Practice", *British Journal of Political Science* 25, 37–56.
- Jenann Ismael (2016), "A Philosopher of Science Looks at Idealization in Political Theory", *Social Philosophy and Policy* 33, 11–31.
- Frank Jackson (1991), "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection", *Ethics*, 101(3), 461–482.
- Frank Jackson and Robert Pargetter (1986), "Oughts, Options, and Actualism", the *Philosophical Review* 95, 233–255.
- Holly Lawford-Smith (2012), "The Feasibility of Collectives' Actions", *Australasian Journal of Philosophy* 90, 453–467.
- Jacob Levy (2016), "There's No Such Thing as Ideal Theory", *Social Philosophy and Policy* 33, 312–333.
- R. G. Lipsey and Kevin Lancaster (1956), "The General Theory of the Second Best", *The Review of Economic Studies* 24, 11–32.
- Avishai Margalit (1983), "Ideals and Second-Bests", in Fox, Seymour (ed.) *Philosophy for Education Jerusalem* 1983, 77–89.
- Julia Markovits (2010), "Acting for the Right Reasons", *Philosophical Review* 119, 201–242.
- Andrew Mason (2016), "Justice, Feasibility, and Ideal Theory: A Pluralist Approach", *Social Philosophy and Policy* 33, 32–54.

- Liam Murphy (2000), *Moral Demands in Nonideal Theory* (Oxford: Oxford University Press).
- Shmulik Nili (manuscript), *A Theory of Global Injustice*.
- J. Raikka, (2014), "How to Find the Second-Best Option?", *Social Justice and Practice* 14, 33–48.
- John Rawls (1999), *A Theory of Justice: Revised Edition* (Harvard University Press).
- Amartya Sen (2006), "What Do We Want from a Theory of Justice?" *Journal of Philosophy* 103, 215–238.
- David Schmidtz (2016), "A Realistic Political Ideal", *Social Philosophy and Policy* 33, 1–10.
- A. John Simmons (2010) "Ideal and Non-Ideal Theory", *Philosophy and Public Affairs* 38, 5–36.
- Walter Sinnott-Armstrong (2003, revised in 2015), "Consequentialism", in *The Stanford Encyclopedia of Philosophy*, available here: <http://plato.stanford.edu/entries/consequentialism/>
- Zofia Stemplowska (2016), "Feasibility: Individual and Collective", *Social Philosophy and Policy* 33, 273–291.
- Zofia Stemplowska and Adam Swift (2012), "Ideal and Non-Ideal Theory", in David Estlund (ed.) *The Oxford Handbook of Political Philosophy* (OUP), 373–388.
- Adam Swift (2008), "The Value of Philosophy in Nonideal Circumstances", *Social Theory and Practice* 34 (3):363–387.
- Victor Tadros (2016), "Permissibility in a World of Wrongdoing", *Philosophy and Public Affairs* 44, 101–132.
- Laura Valentini (2012), "Ideal vs. Non-Ideal Theory: A Conceptual Map", *Philosophy Compass* 7/9, 654–664.
- Kevin Vallier and Michael Weber (2017), "Introduction", in Weber and Vallier (2017), 1–10.
- Michael Weber and Kevin Vallier (eds.) (2017) *Political Utopias: Contemporary Debates* (New York: Oxford University Press).