

Dave Estlund\*

## Replies to Critics

<https://doi.org/10.1515/mopp-2022-0051>

**Abstract:** I reply to several pieces of commentary on my recent book.

**Keywords:** Estlund, utopophobia, ideal theory, non-ideal theory, methodology

I appreciate the opportunity to offer brief responses to the illuminating questions and criticisms offered by the commentators regarding my book, *Utopophobia* (2019).<sup>1</sup> For the most part there is little overlap between the pieces, and so I address each piece mostly on its own. Several of the pieces provide useful overviews of some main themes of *Utopophobia* and so I won't take much space doing that here, but just this: We are told that political philosophy should take care not to be utopian, but there are ways of overdoing it. If someone objects to a theory of what social justice requires by saying, "But you and I both know people will never do that," the right response is (as a starter), "I never said they would. That doesn't mean they aren't required to do so." My thesis is that theories of social justice (unlike practical proposals for taking action) are not shown to have any defect by showing that their standards will never be met, or even that they will have little practical value given how we expect people to behave. Requirements of justice must stay within people's abilities, or so I allow, but unlike proposals to go forward in certain ways they needn't respect probabilities, or even proclivities that are (if any are) owed to human nature. Principles of justice exhibit no vice of utopianism by being unrealistic in those ways, even though practical proposals would. Political philosophy as a field ought certainly to include work supporting practical proposals, but not only that. Even apart from what might be done about it, it matters greatly to many people whether their society is, as they might believe, deeply unjust and in what ways (insofar as it is not all just obvious). What, if anything, is to be done about it is always an urgent question, but it's subsidiary to—because it presupposes—judgment about what social justice, on the best available understanding, requires, and in what ways a society has fallen short. That is a question of political philosophy too, although the nomenclature hardly matters.

---

1 In these replies I take up the pieces alphabetically by author.

---

\*Corresponding author: Dave Estlund, Philosophy Department, Brown University, 45 Prospect St., Providence, RI 02912, USA, E-mail: david\_estlund@brown.edu

Amy Berg (2023) takes on my arguments about motives, ability and requirement. I argue that except, maybe, for a rough category of motives I call “clinical,” such as some phobias and addictions, a person’s motives against doing something do not count that agent as *unable* to do it. If that’s right, then even many highly idealistic conceptions of justice cannot be refuted by their supposedly violating “ought implies can.” Berg begins by observing that some of a person’s cognitive, as distinct from motivational, limitations seem clearly to count against their being able to do certain things. Why, she asks, should we think that motivational features are different from cognitive features in that way? She takes it to be a disadvantage of my view, a kind of disunity, that it treats them differently (239). Whether it is a disadvantage or not, though, her main question is whether they *are* different in the way I claim.

As Berg points out, our abilities to memorize things, for example, are limited and this is a cognitive rather than a motivational limitation. That shows, as she says, that not only external impediments, but sometimes also an agent’s own mental states, can render an agent unable to do something. I haven’t denied that, so the next question is how this supports her main claim that some ordinary (non-“clinical”) motives can also do so.<sup>2</sup>

Berg’s argues that, as the example of memory shows, an agent’s cognitive states can render her unable to do certain things, thus raising a *prima facie* case against the asymmetry in my account according to which our ordinary motives cannot. At one point Berg challenges the asymmetry by saying, “If a cognitive limitation is a genuine inability ... then any motivational limitation that is entirely a result of that cognitive limitation would be a genuine inability too” (245). If correct, that would cast doubt on my claim that ordinary motives are never themselves disabling. Indeed, that is one way that David Wiens’s (2016) story about Claudia, which Berg discusses, could go. Claudia sets out to write a book, and after many good faith stops and starts she abandons the project (Wiens, 8ff.).<sup>3</sup> Suppose her repeatedly stopping is owed to motives that were a direct result of her erroneously believing that, say, any book she writes wouldn’t be any good—a cognitive error. Berg’s formulation would imply that this case of a cognitive variant

---

<sup>2</sup> Berg argues (2023, 246–247) that my use of the idea of “clinical” motives poses a vulnerability for my view, which seems to be this: If I allow these (for the sake of argument) then I’m allowing that not only external impediments can be limitations on abilities. But the memorizing case is enough to show that, and I readily accept it.

<sup>3</sup> Wiens has used the original Claudia case to illustrate “good faith” persistence, arguing that in that case it’s not plausible that she remains able to write a book. I have argued at length against his view (Estlund 2016).

of Claudia (a “Cognitive Claudia,” unlike Wiens’s original Claudia<sup>4</sup>) is unable to write a book.

But I don’t accept Berg’s inference rule that, to repeat: “any motivational limitation that is entirely a result of that cognitive limitation would be a genuine inability too.” Suppose that, because I can’t understand the science, I erroneously believe that my eating more green vegetables, which I don’t enjoy, will not make any difference to my health—two cognitive failures: the failure to understand and the false belief. Consequently, I have little motivation to eat more green vegetables. If I set out to do so at all (maybe because I’ve been exhorted my whole life to do so), I will understandably not persevere. That’s a motivational limitation in the sense Berg and I are considering, and it is “entirely a result of that cognitive limitation,” which is itself “a genuine disability.” If not for the cognitive limitation I’d be motivated to eat more vegetables. But surely this does not count me as *unable* to eat more green vegetables.

Berg doesn’t explicitly use that dubious inference rule in conjunction with either of the Claudia examples, but she does say this: “The parallels between original and Cognitive Claudia show us that even everyday motives can constitute true inabilities—in sum, everyday motives can block the requirements of justice” (246). Unless that means to employ the dubious inference rule in the way I sketched, then I don’t see which “parallels” are meant. Both Claudias involve an agent’s ability being limited by her own mental states, but as noted earlier that has never been at issue. The issue is about ordinary motives in particular.

As for my treating cognitive and motivational cases differently, surely there does remain this important difference between them: In the case of a cognitive inability to memorize a book—and here is my conditional criterion for being unable, and by appropriate adjustment, being able—even if the agent continued to try without giving up she would not memorize the book. That is not so for the motivational case of the original Claudia, the aspiring author, as Berg appears to agree. Claudia might well succeed if only she didn’t give up, which she wouldn’t do if her motives were different. This is a difference that makes all the difference, according to my conditional analysis of being able or unable to do something.

Neither of the Claudia examples works as a counterexample to that analysis. If it were obvious that Motivational Claudia, who gives up even though she would otherwise succeed, must count as unable to write the book, that would be a counterexample. But failing because you stop trying is no sign of inability (see Estlund 2016). And Berg’s own Cognitive Claudia, who tries to memorize rather

---

<sup>4</sup> Wiens doesn’t use the example in this way to emphasize cognitive limitation. Berg’s Cognitive Claudia example is about the inability to memorize a book, but its point is cognitive inability more generally (244).

than to write a book, is not, and is not meant as, a counterexample either, because unlike the original non-persevering Claudia, even if she kept trying she would not succeed. My criterion then gives us the intuitive result that she is indeed unable to memorize the book.

Berg doesn't dispute either that criterion for ability or my claims about what it implies in the two Claudia cases. As I say, she might have in mind a third Claudia, a second cognitive case, in which her giving up on the book project is owed to a cognitive mistake such as false or irrational belief that she's a terrible writer. That example would only be trouble for my criterion if such a case must count as an inability to write a book. But the only reason we're given to think so, I think, is the dubious inference rule, which is apparently refuted by the example of green vegetables.

I'm not persuaded, then, that any parallels between cognitive and motivational causes for an agent's not doing something show that they must both or neither be counted as cases of inability. There is no evident inconsistency in holding, for the reasons I've given, that some internal states or capacities—such as limited memory—render us unable to do certain things, others—such as ordinary motivations—do not.

*Stefano Bertea* (2023) calls attention to the important fact that proposing, or even partially pursuing, an entirely unrealistic political project can sometimes have good effects, and those effects bear on whether it is wise or justified to make the proposal. If what is proposed is utterly unrealistic—for example, because people will probably not do the things that the project's success or value depend on their doing—that counts against offering that proposal, but other consequences must also be factored in and might outweigh that. For example, it might be both that making the proposal has very little chance of inducing its disastrous pursuit, and also that offering it will get people to think about it and take it seriously further down the road in times when that kind of compliance may be much more realistic. No blanket condemnation of publicly making highly unrealistic proposals is plausible. The point is important, and it's also important that my own arguments not be misunderstood as suggesting otherwise, as Bertea suggests.

The term “proposal” has an ambiguity between an object and an action. When something like universal basic income is proposed, there is the proposing, the act, and the thing proposed, the object. Either might be called the proposal (For example, respectively, “The proposal at the meeting was interrupted by a fire drill,” and “The proposal is expensive”) but they are not the same thing.

For this reason, it's helpful to have some terms that avoid that act-object ambiguity. Let's distinguish between a proposing, a project and a pursuit as

follows:<sup>5</sup> A *proposing* is the act of calling for pursuit of a project—pursuing it meaning to set out to carry it through.<sup>6</sup> A *project* is something that is or could be proposed, which means to put it up for practical consideration—that is, for possible pursuit.<sup>7</sup> A *pursuit* is a project pursued. A *bad project* is where comprehensive pursuit—pursuit of it toward completion—would be bad. It might be bad because it is doomed to fail, or in a variety of other ways. On one meaning, a bad proposal is a bad project proposed, which leaves open whether the proposing is bad—a bad proposal on another meaning.<sup>8</sup>

As Berteau points out (now putting it in these terms) proposing a bad project might still be good as an act of proposing for various reasons, such as opening minds, stimulating constructive debate, enlarging possibilities, etc. The merits of a proposing, then, are not settled by the merits of the project that would be proposed.<sup>9</sup> Proposing the full project, even if it's a bad one, might have the value of inducing its partial pursuit—pursuit of a subproject that is good. In that case the proposal is, in one way, a bad one, but proposing it might be good because it induces pursuit, along the way, of a good subproject. That said, the value of the proposing is not, as such, any defense of the merits of the project.

What I hold in *Utopophobia* is that projects whose value depends on certain things, such as levels of compliance that are contrary to fact, are valueless, and so such a proposal is, in that sense, worthless as well, whether or not proposing it might yet be good.<sup>10</sup> In the typology of two kinds of “idealism” in political theory offered by Berteau, then, my position cannot be located in the “moderate” camp, which would mean that it objects to *proposing* things (proposals in one sense)

---

5 It might not properly be a project until it is underway, only a possible or upcoming project.

6 In ordinary language to propose something might be only to propose it for discussion, but to keep things clear I won't call that a kind of proposing.

7 At one point in *Utopophobia*, I gloss “a project” as “a goal to set out for,” which works as well (2019, 10). Berteau speaks at one point of blueprints, and I understand him to mean a detailed description of a project. The blueprint might be a physical object or piece of digital information, while the project so described is always an abstract object.

8 Partial pursuit of a project can be defined as comprehensive pursuit of a subproject, say, the first three stages out of five of the full project, or whatever. If it is said that it would be good to pursue a project because even though it will fail (or be bad in some other way), part of it will succeed, we can more carefully state this case as involving a part of the overall project that is not bad—a subproject. The overall project would still be bad because comprehensive pursuit of it will fail (or be bad in some other way).

9 There is a special case, discussed in the piece by Prendergast in this symposium, where the proposing might have effects which make it a sounder project by inducing more expected compliance, and so on.

10 Of course, a project's value might not depend on its full success, but that's consistent with this claim.

unless *what is proposed* (proposals in the other sense) is suitably realistic. As I say, I agree with Berteau's point that, for reasons he partly canvases, there might be nothing wrong with constructing, publicly stating or even proposing such projects even though they are not in themselves sound.

In the context of my argument in *Utopophobia*, my point was that even though unrealistic proposals—that is, unrealistic projects—are bad for that reason, unrealistic principles are not. That is to say, in the more specific terms I've introduced, any proposal that is unrealistic—that is, calling for comprehensive pursuit of an unrealistic project—is bad as a proposal,<sup>11</sup> as a project proposed. That's pretty obvious if to propose it is to call for its comprehensive pursuit. If it is sufficiently unrealistic then it would not be good to comprehensively pursue it. That point says nothing about whether it might yet be good to propose a bad project—to make a bad proposal in that sense. As Berteau points out, elsewhere in *Utopophobia* I emphasize the great value there can be in thinking in (at least what seem to us at the time to be) unrealistic terms. This is partly because what seems unrealistic—such as abolishing slavery, electing a Black president—has often turned out not to be unrealistic after all. Importantly, that isn't yet either to defend or object to proposing what is really, for all time, unrealistic. But as Berteau emphasizes, that, too, might be a good thing to do. It might induce pursuit with partial success, or it might have various other vindicating consequences.

In Prendergast's (2023) piece, to which I'll turn more fully below, she focuses on one practical benefit in particular of unrealistic acts of proposing, a point that is very much in the spirit of Berteau's piece: By offering what I take to be an unrealistic (and so unsound) proposal I might, Prendergast observes, contribute to its feasibility—its realism. But, formulating things more precisely, it's not clear if Prendergast is advocating offering proposals that one genuinely takes to be unrealistic *even in light of* the small effect on feasibility of proposing it. In that case, where I have no reason to think my own proposing the unrealistic thing would make it realistic, nothing Prendergast says is, after all, any reason for proposing it. Then, what about the other case, where I believe that it would be unrealistic except for my proposing it, but proposing it *will* make it realistic? Well, then I would not, after all, be offering a proposal that I take to be unrealistic, only one that I believe would be unrealistic unless I proposed it. This might well be what Prendergast means. It is certainly important to take account of the effects one's own acts of proposing things have on the feasibility of what is proposed, especially if proposing it might tip the balance. For most individuals most of the time that's highly unlikely, but there might still be a non-consequentialist moral reason of some kind to perform such a non-decisive contribution to what could be a set of

---

<sup>11</sup> See *Utopophobia*, page 40.

actions that will tip the balance. Similar cases include voting, reducing one's personal carbon footprint, making a free throw, etc.

*Jochen Bojanowski* (2023) raises some objections to my treatment of facts and principles, with special reference to views of John Rawls and G. A. Cohen. Cohen (2008) argues that any normative principle that derives partly from facts cannot be normatively fundamental because those facts could only have the alleged normative relevance in virtue of a more fundamental principle. In *Utopophobia* I argue, not that Cohen was mistaken, but that the point is normatively inert as deployed against Rawls (1999), for the following reasons. We can grant to Cohen that Rawls's two principles of justice depend, as Rawls himself says, on certain facts of human life. This relevance of facts is represented in Rawls by the parties to the original position (OP) reasoning in light of those facts. So, granting this also to Cohen's argument, there must be a more basic principle about how those particular facts bear on justice—a principle that stands apart from those facts. In response, suppose Rawls were to grant that there is a principle that is more fundamental in Cohen's sense, and it is this:

The unearthed Rawlsian principle

A basic social structure ought to meet principles (which are in this way less fundamental than this principle itself), whatever they are, that would be chosen in the OP, with its sensitivity to the relevant facts, whatever they might be.

This should satisfy Cohen's insistence on a more fundamental principle explaining the bearing of the facts on the two principles of justice. But, I argue, as compared with the less fundamental (in his sense) two principles of justice, this theoretical structure makes no difference to what a basic social structure must be like in order to count as just—it is normatively inert in that way. That is not an objection to Cohen's point about facts and principles, but it denies that the point grounds any normative critique of Rawls's two principles of justice.

Bojanowski suggests that this argument of mine about the unearthed principle is not germane, because Cohen was only talking about normative principles, and the unearthed principle is not a normative but a "methodological" one (Bojanowski 2023, 269).<sup>12</sup> We can tell this—or so I take the suggestion to be—by noting that it is normatively inert just as I have admitted, or insisted. However, being normatively inert in that way isn't the same as being non-normative, contrary to what Bojanowski must be suggesting. Bojanowski doesn't offer a general

---

<sup>12</sup> In *Utopophobia*, I also mention these issues about when an ostensibly deeper principle is normative or merely methodological, or "constitutive," at page 47. I cite there a piece by Mark Schroeder (2005).

account of what makes a principle normative, and I don't have one to propose, but at the very least it should not seem obvious that the unearthed principle is not a normative claim (whether or not it is also in some sense "methodological"). For example, Cohen writes, "A normative principle, here, is a general directive that tells agents what (they ought or ought not) to do" (2008, 229). The unearthed principle specifies what criteria principles, and so institutions on pain of injustice, "ought" to meet.<sup>13</sup> So, unless it can be explained (and perhaps it can) in what way the unearthed principle is not relevantly normative for Cohen's purposes, Bojanowski's objection seems unavailable.

Cohen gives the following argument for his claim that a deeper principle is involved in such explanations of how facts support principles, the defense being, "simply to challenge anyone who disagrees to provide an example in which a credible and satisfying explanation of why some F supports some P invokes or implies no such more ultimate [normative]<sup>14</sup> principle" (2008, 236). Cohen is confident then that anything that satisfactorily explains how certain facts support certain principles *will* itself be a normative principle. It's hard to see how the unearthed principle would not qualify. It seems to provide just such an explanation: What's unjust about unjust basic social structures is that they fail to meet principles that address people's interests impartially in the way modeled in the OP. Why, then, doesn't the unearthed principle satisfy Cohen's demand, while also blunting the significance of his claim, even if it's right (about which more shortly) that this makes Rawls's two principles less than fundamental? And, to be clear, all my argument requires is that the unearthed principle explains the bearing of facts on principles in Rawls's theory. I don't even need to claim that it's normative, whatever that might mean here.

I say that the unearthed principle is, insofar as it is normatively inert, "metaethical," (2019, 176) and Bojanowski complains that I thereby neglect Cohen's explicit avoidance of metaethics. But, as Bojanowski notes, Cohen only claims to avoid a certain subset of metaethical issues, those concerning the objectivity or reality of values. Indeed, Cohen says that his thesis about facts and principles is itself, "a piece of neglected, and routinely denied, metaethical truth" (2008, 268). So, Bojanowski must mean that my use of the unearthed principle is not only metaethical, but also implicated in the cluster of metaethical questions

---

<sup>13</sup> There is an issue about whether a principle about what is just or unjust is actually normative in that "action-guiding" way. Whom does it tell to do what? But that doesn't apply to the unearthed principle any more than it does to the two principles of justice themselves, whose normativity Bojanowski clearly grants.

<sup>14</sup> Cohen writes, "This chapter concerns the relationship between facts and normative principles (or, as I shall call them, for short, principles)" (2008, 229).

about the objectivity of values.<sup>15</sup> And this might be the import of his saying that I hold “that Cohen’s FIC [fact-insensitivity claim] turns on the question of whether the principles of justice are true or justified “because God says so, or because there is a quasi-Platonic fact-free principle” (Bojanowski 2023, 278). But I think that misunderstands the passage of mine quoted. What I argued is rather that, as is characteristic of Rawls, *irrespective* of those kinds of metaphysical issues (even though the unearthed principle is metaethical in a broader sense), Rawls “may still assert that the principles of justice are justified by their appeal to the parties in the OP in light of the facts of human life” (2019, 176–77). That’s what the unearthed principle implies after all.

Must Rawls nevertheless retract his claim that the principles selected in the OP are “first principles” of justice? Well, it’s not clear that the unearthed principle is a principle of justice at all.<sup>16</sup> If it isn’t, then we haven’t turned up any challenge to Rawls’s claim that the two principles are the “first principles” of justice, or that the content of justice, at the fundamental level, depends on the “facts of human life.”

Cohen may have thought that the deeper explaining principle, once excavated, would turn out to conflict with the OP’s “constructivist” methodology, to the latter’s detriment. If that’s the idea we would need more than the general point about the regress of principles. We would need it to be explained *which* deeper principle serves in the explaining role if not the constructivist unearthed principle itself, and how the real explaining principle conflicts with the OP method.

Lars J. K. Moen (2023) focuses on my rejection, in *Utopophobia*, of what seems to be a common thought, namely that a proper standard of justice must be something whose satisfaction we ought not only to hope for, but to actually strive for in practice. Put the other way around, if some supposed standard of justice would be hopeless to meet, given people’s actual expected behavior, that standard is (this view says) not correct about what justice requires. Stated yet one more way, a proper standard of justice must be shaped or bent<sup>17</sup> to fit our bent—our natures and proclivities in such a way that it is indeed something that we ought to set out for. I call that the “bent” view of justice, and I criticize it.

---

<sup>15</sup> What Cohen says, as Bojanowski notes, is only that “the question pursued here is distinct from those that dominate the meta-ethical literature,” namely, “controversies about the objectivity of principles” (both at page 230). Cohen also says, “My thesis, that principles that reflect facts reflect principles that don’t reflect facts, is, if correct, of interest ... as [among other reasons] a piece of neglected, and routinely denied, metaethical truth” (2008, 268).

<sup>16</sup> Similarly, perhaps, on a Kantian view, the categorical imperative is more fundamental than the prohibition on lying in the relevant sense, but it is not itself a principle of honesty.

<sup>17</sup> I don’t mean shaped or bent by any agent, but in the sense in which a found stone might be shaped like a heart, or a discovered tree might be severely bent.

Moen argues that those like me who reject the bent view must go beyond pursuing our dissenting intuitions, much less merely disputing about terminology such as the word “just.” Instead, he says, we must offer a more substantive critique of the reasons that are offered for accepting that bent view. Before arguing that, in any case, I do so in the book, it’s not obvious that he’s right about this. Certain reasons for rejecting the bent view might be quite weighty even without considering the reasons in favor of it. For example, as I point out in *Utopophobia* (2019, 144) the bent view appears to have the following implication. Imagine a society in which most people are self-consciously and virulently racist. They broadly accept the institutions that are most appropriate under those motivational conditions even though, presumably, these would not all be friendly to their racism. It’s not that the racists couldn’t do otherwise than be racist, but only that predictably most will not do otherwise. Institutions can only do so much, and across the races there is in this hypothetical society massive segregation, widely divergent prospects for health and success, frequent dismissal of opinions of the targeted race in public discussion, and so on. Sure, it might be that if institutions are chosen properly then they will change people in the right ways over time (or maybe not). But in the society I sketch they are not changed yet (2019, 144).

What does the bent view imply here? It must apparently say that this could be a just society—not only just to some degree, and also unjust to some degree, but fully just, not unjust.<sup>18</sup> That’s because any standard that requires a social structure that lacks such racial depredations is, under the sad circumstances, not one we should set out for. It would, we are stipulating in the example, not only be doomed, but such a utopian project might only make matters worse. Therefore, the bent view continues, this execrable social structure might meet proper standards of justice because such standards must be goals we should actually set out for given how we expect people to behave.

Now, it is clear and agreed that what we ought to actually strive for is a social structure that is appropriate and sufficiently feasible given the facts, however sad they might be. What is in dispute is not that, but the bent view’s implication that the society I have described would count as *just* if this is the best that could be done given the intransigent racism. Call this my argument from apartheid: Since the described racist society is not just, but the bent view says that it is, the bent view is false.

Moen responds to this argument by suggesting that a theory of justice might decline, as maybe Rawls does, to make any claims about such a badly deformed

---

<sup>18</sup> I say “could be” since whether it is just might also depend on some other considerations not specified in the example. But whatever those factors might be, the bent view says they might be configured so that this is indeed a just society.

society. But, in reply, it is easy to adjust the example by steps in the direction of less severe racism. If the theory closes its doors and declines even to address the adjusted cases just at the point where the society would be patently unjust even though the bent view would imply otherwise, then that reticence would simply appear to implicitly refuse, as I recommend, to accept the implications of the bent view. If so, I could rest my case.

So far, I grant that I have not here addressed any reason that might be given for adopting the bent view, as Moen argues that I must. To be clear, though, what would need addressing for the argument from apartheid to be a strong case is, I think, not a response to simply whatever reasons someone might offer for the bent view, but only to reasons that might credibly be held to outweigh or defeat the argument from apartheid. No such argument would be offered merely by pointing out, as Moen often does, that the justice-benders might be unpersuaded and stick to their bent view, perhaps leaving nothing but a terminological dispute about “just.” (For example, should that word mean standards that are bent or unbent?) It’s true, of course, that anyone can use “just” to mean whatever they wish, if they make their meaning clear. But that by no means guarantees that they will be addressing questions of *justice* at all—for example, if they were to stipulate that “just” shall, as they use it, mean popular. It’s also true that, paraphrasing Moen, “justice benders [could] keep clinging to the view that justice must be the value we should aim to realize, [and] they can continue to claim, contra Estlund, that justice principles should be adjusted in response to human motivation” (2023, 298). The question, though, is what strong *reason* there is to keep clinging to that bending principle in light (for one thing) of its apparent implications in the apartheid example above.

I believe the main or perhaps the only argument Moen considers in favor of the justice-bending view—as against my rejection of it (but which he offers his own reasons for rejecting), and it is an influential one—is that without the standards being bent to meet human proclivities (good, bad or indifferent) a theory of justice would fail to have much practical significance. That is, one might defend the bent view on the basis of a position I call “practicalism,” namely that no knowledge is valuable unless it has practical value. Thus, unbent theories of justice, supposing for the moment (but which I dispute in *Utopophobia*) they would have little or no practical value, themselves yield no valuable knowledge. As it happens, I devote much space in *Utopophobia* to assessing practicalism, including the whole of chapters 16 and 17, and so I plead “not guilty” to the charge of failing to address the main arguments that might be offered for the bent view.

Be that as it may, Moen also favorably cites Ryan Muldoon’s critique of my argument against practicalism, in which I point out that “if you think math can

have nonpractical value then you cannot deny a piece of political theory value on the bare ground that it has no practical value” (2019, 310). Muldoon complains that

Proofs in pure math have their epistemic standing precisely because all of those methods and definitions are settled. But that is not the case in political philosophy. “Justice” (etc.) is up for grabs ... It is not the proofs of pure math. It is closer to work in string theory. (2021, 595)

Suppose what Muldoon implies is right, that string theory in physics involves concepts and other elements that are contested in their very meaning. Still, I don’t see anything troubling for my view in that analogy. Indeed, in *Utopophobia* I actually gave two examples as challenges to practicalism: higher mathematics and cosmology (2019, 36). The latter happens to be intimately engaged with string theory.<sup>19</sup> Unless Muldoon invites us to doubt that research by physicists working on string theory is valuable unless it should turn out to have practical value then the example is a fine fit for my critique of practicalism,<sup>20</sup> which is one of the main grounds that gets offered for bending standards of justice to fit our proclivities.

*E. R. Prendergast* (2023) identifies what she calls a “paradox of political feasibility,” namely that, as discussed briefly above, by judging some desirable political project to be infeasible due to insufficient public support, I thereby—presumably via behavior that it induces in me—contribute causally to that lack of support, and thus to its infeasibility.<sup>21</sup> In light of that, since there is reason to contribute to the feasibility of desirable projects, there is reason *not* to judge that it is infeasible. If one does initially judge that it is infeasible, this is, he says, roughly a duty to change that judgment and engage in a kind of wishful thinking. This issue is not pressed against anything in *Utopophobia*, but I offer a few thoughts before turning to some criticisms he does make.

The conclusion of Prendergast’s argument is that in such conditions we have “duties to engage in wishful thinking” (2023, 307). She adds a qualification, anticipating a natural objection “to whatever extent wishful thinking is under one’s control” (307). The objection would be that whether or not I judge a project to

---

<sup>19</sup> For example, see the Wikipedia article on “string cosmology.”

<sup>20</sup> Muldoon precedes this point with a related argument that the fact of political disagreement about fundamental values is at least as important a reason to doubt the practical value of unbent theories of justice as is human moral deficiency, on which my arguments tend to focus. Such disagreement is suggested to be what makes ideal theory rather like string theory. Let me just briefly note that a party to a disagreement about some topic such as justice can hardly deny the value of inquiry addressing whether they are correct in their contested view of justice or not, a point I make in *Utopophobia* at pages 47ff. Of course, there are *also* crucial questions about what to do about disagreement.

<sup>21</sup> Prendergast never states the puzzle explicitly, but I’ve tried to do so in a way that is true to her intentions.

be feasible is, indeed, not under my control. I don't have the ability to come to believe it just because there are practical reasons to do so if I could, anymore than I can choose to believe it's raining while seeing through the window that it isn't, even if so believing would win me a lot of money.

At the very least, whether our beliefs are under our voluntary control is philosophically controversial, so Prendergast might be happy to simply hitch his wagon to the "voluntarist" side in that debate. But there is an alternative route it seems natural for her to consider, which is that, while there is no duty to judge one way or another (which may not be possible for the agent anyway, as I say), there is a duty, roughly, to *act as if* one judged the project to be feasible.<sup>22</sup> That endeavor is under voluntary control. Prendergast seems to fluctuate between the duty to believe and the duty to act as if one believed, when she writes (emphasis added),

Political theorists ought to proceed in civic life with some hope for their preferred political outcomes and an *expectation* that political projects can succeed ... Political theorists ... should proceed with *making arguments* that lead to conclusions that say that we ought to do that which seems politically infeasible in those cases when they hope a politically infeasible project will succeed.

(2023, 317)

I leave the matter here in order to turn briefly to Prendergast's engagement, for related purposes I won't describe here, with my example of Doctors Slice and Patch. (The example is offered by me for a specific purpose which also won't concern us here.) In *Utopophobia* I speak of what I call the "global prime requirement," which purports to conjoin actions by different agents under a single requirement. What kind of requirement is that? We sometimes say things like, "You all ought to build that barn." But except in special cases that collection of agents is not an agent and so can't be under a standard obligation. Nor is it clear in such cases that each, or even any, individual has any obligation to do her part. This is especially clear when doing her own part would be quite bad and clearly impermissible unless enough others did theirs, and they could be in the same boat. So, what could we mean that vindicates this undefined idea of a kind of plural requirement? For a sharp illustration of the troubling possibility I offer this example:

#### Slice and Patch Go Golfing

Suppose that unless the patient is cut and stitched he will worsen and die (though not painfully). Surgery and stitching would save his life. If there is surgery without stitching, the

---

<sup>22</sup> Or perhaps to "accept" that proposition in a way that is voluntary, unlike belief. See Chignell (2018).

death will be agonizing. Ought Slice to perform the surgery? This depends, of course, on whether Patch (or someone) will stitch up the wound. Slice and Patch are each going golfing whether the other attends to the patient or not. It is clear that each of them, callous as they are, is morally deficient in her motives, then. But, does anyone act wrongly?  
(2019, 211)

Patch ought to stitch the patient if and only if Slice will perform the surgery. Stitching is possible, but pointless and harmful if there is no wound that needs stitching. But suppose that Slice will not perform the surgery. Patch might as well go golfing. Ought Slice to cut? No, because no one, including Patch, will stitch, and so the surgery will only make the patient's death more painful. Slice might as well golf. Neither has acted (or omitted) wrongly, despite the fact that the patient will needlessly die. Many of us respond to this case with the intuition that there is some moral violation here, but the puzzle is to find an agent who has committed it.

Prendergast doubts the “intuitive pull of the case,” because in any realistic story similar to that one the problem is bound to be down to one of the agents violating a duty. Now, even if situations that fit the violation-free model of Slice and Patch are exceedingly rare, or even never actually occur (which I by no means concede), that does not address the central point of the example. It is not a practical question, but a philosophical question about whether we really know what we're talking about when we say that a group of people ought, together, to build a barn, or save a life, or do their part in various important projects. If I've come up with an example where you think the set of individuals would be required to carry out the project, but where it's also clear that none of the individuals is obligated to do their part, then surely your view is under challenge. What kind of requirement do you mean if no agent is so required? It is no answer to say that such examples never occur in the real world. Even if that were so, what do you mean by their being required? And then, perhaps you ought to wonder whether the way you've been using that idea of requirement in the real world can be given any definite and defensible meaning?<sup>23</sup>

Prendergast says that even in the story as I tell it, Slice and Patch have each acted wrongly by going golfing at that time (actually, by “planning to” do so, but my reply is the same). But unless doing so must (not might) violate their contract or a promise or

---

**23** Prendergast hasn't, I think, complained that the Slice and Patch example is incoherent, as if it couldn't possibly arise in real life. We see that in her granting that the example would succeed if only this or that implausible but possible thing happened to be the case. For example, it's clearly conceivable that the doctors either have no way to communicate or coordinate, or doing so would be too dangerous, and so on. This responds to Prendergast's supposition, with no basis offered, that in realistic cases the parties would be able to communicate, and the wrong would be the failure by one or both to do so. But I would add (as I discuss in *Utopophobia* at page 224) that we could easily assume that the doctors did indeed communicate their plans to each other, which would not have changed anything relevant to the puzzle.

something (a convenient empirical stipulation for a critic, but on no discernible basis), why shouldn't each go golfing given that the other will not be at the hospital anyway?

But, in any case, and whether or not it matters, this is an occasion to illustrate how realistic examples can indeed fit the model, making the case deeply puzzling for our real-world normative thought. One example I do give in the book, by the way, is mutual assured destruction in a nuclear arms race. It could certainly happen that no negotiations succeed in removing both threats, and that each has sufficient reason to keep up the threat in order to deter a unilateral attack from the other. Neither head of state acts wrongly in keeping up the threat of gruesome violence on a mass scale. Many have a strong sense that this condition of mutual threat is wrong, but it appears neither side is necessarily doing anything wrong, and might even be acting as morally required. In any case, here is another example:

#### Generals Spice and Hatch

The enemy in this war has heinous designs on the world if they were to prevail. In a pivotal move for the course of the war enemy forces have occupied a large working industrial warehouse, preventing the dozens of non-combatant workers from leaving. To prepare for an attack on those enemy forces General Spice has the power to first see to the removal of many of the confined workers, but some will be killed, others will be injured, and Spice himself will be at significant risk. General Hatch has the power, at some risk to herself, to follow up by attacking the warehouse and incapacitating the enemy occupiers. As it happens, Spice, out of selfish caution, would not perform the removal even if Hatch would follow up with the attack. And Hatch would not attack even if the removal were performed. Indeed, on the appointed day, there is no removal and no attack, and, due to generals Spice and Hatch's behavior, the evil enemy gains an irreparable advantage in the war.

Nothing assumed in the example is bizarre. If one objects that each general is a culprit after all because of their objectionable selfish dispositions, in *Utopophobia* I deny that this is relevantly true.<sup>24</sup> Briefly, each general's moral character is a preexisting condition, so to speak, and so that character itself is no worse on the appointed day than it was a month earlier. Therefore, despite a common sense that their behavior on that day is wrong, no additional wrong on the appointed day is explained by reference to that ongoing deficiency of character.<sup>25</sup> Nor are the

---

<sup>24</sup> Niko Kolodny offers an objection broadly of this kind in his forthcoming book *The Pecking Order* (2023). He says the victim of Slice and Patch has a moral complaint against each for acting on the wrong reasons, even if they did the right thing. What I say in the text above indicates my reply to Kolodny as well. The victim has no basis for complaint against either doctor for their role in the harm, which would have happened even if that doctor had all the right motives. Then we can't explain how the patient has any more basis for complaint—which is not to say none—than he or anyone would have on this and on any emergency-free day.

<sup>25</sup> There is more to say in anticipation of other replies in *Utopophobia* at pages 225ff.

enemy's later victims, in particular, especially warranted in such a complaint against either general. Not against Spice, because no one undergoes any harm they would not have undergone even if Spice had perfectly good motives, since that would rightly have led him to stand down, wrecking the plan anyway. Same for Hatch; given that Spice wouldn't have acted, Hatch rightly would not have attacked the enemy holdout. The sense persists, at least for me, and only provisionally of course, that there is wrong in Spice and Hatch not doing the combined maneuver, and also that there is no agent of that compound maneuver, and so no agent that has committed any such wrong. The philosophical import is whatever the import of the Slice and Patch case is, with the addition, in case it matters, of a perhaps more realistic and politically significant case.

## References

- Berg, A. 2023. "Which Limitations Block Requirements?" *Moral Philosophy and Politics* 10 (2): 229–48.
- Berteau, S. 2023. "A Defence of Robust Idealism in Political Philosophy." *Moral Philosophy and Politics* 10 (2): 249–66.
- Bojanowski, J. 2023. "Bad Facts and Principles: Finding the Right Kind of Fact-Insensitivity." *Moral Philosophy and Politics* 10 (2): 267–84.
- Chignell, A., Spring. 2018. "The Ethics of Belief." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. Also available at <https://plato.stanford.edu/archives/spr2018/entries/ethics-belief/>.
- Cohen, G. A. 2008. *Rescuing Justice and Equality*. Cambridge: Harvard University Press.
- Estlund, D. 2016. "Reply to Wiens." *European Journal of Political Theory* 15 (3): 353–62.
- Estlund, D. 2019. *Utopophobia: On The Limits (If Any) of Political Philosophy*. Princeton: Princeton University Press.
- Kolodny, N. 2023. *The Pecking Order: Social Hierarchy As a Philosophical Problem*. Cambridge: Harvard University Press.
- Moen, L. J. K. 2023. "How Do You like Your Justice, Bent or Unbent?" *Moral Philosophy and Politics* 10 (2): 285–98.
- Muldoon, R. 2021. "Review of Estlund: Utopophobia: On the Limits (If Any) of Political Philosophy." *Review of Politics* 83 (4): 593–6.
- Prendergast, E. R. 2023. "Defense of Wishful Thinking." *Moral Philosophy and Politics* 10 (2): 299–320.
- Rawls, J. 1999. *A Theory of Justice*, Revised Edition. Cambridge: Harvard University Press.
- Schroeder, M. 2005. "Cudworth and Normative Explanations." *Journal of Ethics and Social Philosophy* 1 (3): 1–28.
- Wiens, D. 2016. "Motivational Limitations on the Demands of Justice." *European Journal of Political Theory* 15 (3): 333–52.