Against Ideal Guidance, Again:
A Reply to Erman and Möller

David Wiens

Eva Erman and Niklas Möller (2022) have recently presented a trenchant critique of my (2015) argument that ideal normative theories are uninformative for certain practical purposes. Their criticisms are largely correct. In this note, I develop the ideas behind my earlier argument in a way that circumvents their critique and explains more clearly why ideal theory is uninformative for certain purposes while leaving open the possibility that it might be informative for other purposes.

Eva Erman and Niklas Möller (2022) have recently presented a trenchant critique of my (2015) argument that ideal normative theories are uninformative for certain practical purposes: namely, for specifying the normative principles we should realize (the "Target View"), and for specifying a set of principles that we should use to normatively rank feasible options (the "Benchmark View").[1] Their criticisms are largely correct: my model is at turns obscure and clumsy and my arguments contain several missteps. Even still, we can develop the ideas behind my original argument in a way that circumvents their critique. In this note, I do just that in an effort to more clearly explain why ideal theory is uninformative for certain purposes while leaving open the possibility that it might be informative for other purposes.

1. IDEAL THEORY: AN UPDATED MODEL

My original exposition was unnecessarily abstruse, but the intuition behind that model is simple. An ideal is represented by the set of social possibilities that rank highest according to a theorist's "evaluative criteria" (W 437) and are consistent with her assumptions about the background circumstances in which the institutional schemes or social practices under consideration are imagined to be implemented (W 438). An ideal theory is a set of principles that are fully implemented by the ideal (W 434–35); what ideal-theoretic principles say ought to be the case is in fact the case in an ideal state of affairs (cf. EM 527, note 11).

1 Hereafter I cite Erman and Möller (2022) as EM and Wiens (2015) as W.

David Wiens

Let's now specify a model that deviates from the original one in its details while preserving its spirit. Let a *candidate theory* be a set of normative principles that a theorist considers to be a potential answer to the question, "Which principles would be realized by a normatively ideal state of affairs?" We assume that each candidate theory has numerous possible *implementations*: each theory can be realized by various configurations of institutions, social practices, and interpersonal behavior and across a wide range of circumstances. We *accept* a candidate theory *as our ideal theory* by comparing the expected costs and benefits of implementing each candidate theory, with certain qualifications to be specified shortly.[2] To render implementations of candidate theories determinate for the purposes of comparing them, we represent each possible implementation of a candidate theory with a *model scenario*, which is a selective description of a social possibility (or perhaps a set of possibilities) that focuses our attention on those aspects that are central to a theorist's particular theoretical aims, whether these be a hypothetical society, a particular component of a society's institutional arrangements (e.g., its distributive component or its decision-making component), or a network of relations among societies. The *set of candidate scenarios* represents the set of all possible implementations of the theorist's candidate theories. As I argued earlier, ideal theorists do not accept a candidate theory as their ideal theory by canvassing the set of candidate scenarios in its entirety. Instead, they focus their attention on a limited range of candidate scenarios, usually those that are consistent with certain idealistic assumptions about the circumstances in which the candidate theories are imagined to be implemented; for example, that individuals have "an effective desire to comply with the existing rules and to give one another that to which they are entitled" (Rawls, 1999b, 274f) or that markets are perfectly competitive and free of corruption and fraud. Let the *set of idealistic scenarios* be the (proper) subset of candidate scenarios that are consistent with a theorist's assumptions about the circumstances in which candidate theories are imagined to be implemented.

In the spirit of my earlier model, we define an *ideal* as the idealistic scenario that ranks highest according to a theorist's standard for comparatively evaluating candidate scenarios. To make this thought more precise, let an *evaluative criterion* specify a basic value or consideration that informs our normative evaluation of candidate scenarios. For example, suppose the extent to which a candidate scenario realizes personal liberty informs our normative evaluation of candidate scenarios; then the criterion of personal liberty specifies what it is we are looking for when we judge the extent to which a candidate

---

2 "The evaluation of [candidate] principles must proceed in terms of the general consequences of their public recognition and universal application, it being assumed that they will be complied with by everyone" (Rawls, 1999b, 119; see also Simmons, 2010, 8).

scenario realizes personal liberty (similarly for other potential criteria). We can think of an evaluative criterion as providing (the basis for) a measure of a basic consideration that affects our normative assessment of candidate scenarios. An *evaluative standard* specifies the set of relevant criteria and their relative importance in our evaluation of candidate scenarios. For example, an evaluative standard might tell us that personal liberty and material well-being are equally important; or it might tell us something more complex: for instance, that personal liberty matters more than material well-being when the latter exceeds some threshold level but that liberty matters less than well-being when the latter falls below that threshold.[3] So an ideal is the idealistic scenario that optimally realizes certain basic evaluative criteria given their relative importance as specified by the theorist's evaluative standard.

Crucially, an evaluative standard is not the same as a candidate theory. If that were the case, then an argument that we should accept a candidate theory as specifying the content of an ideal theory risks begging the question. To illustrate, suppose Rawls's two principles of justice constitute both a candidate theory and the evaluative standard used to comparatively evaluate idealistic candidate scenarios (cf. EM 534–35). Given that we rank candidate scenarios using Rawls's two principles (and the associated "priority rules"; see Rawls, 1999b, 266), it would hardly be surprising to find that the highest ranked idealistic scenario depicts an implementation of Rawls's two principles. In this case, if someone asks why we should accept Rawls's two principles as specifying the content of our ideal theory of justice, our answer must be "Because, among the idealistic scenarios, the scenario depicting an implementation of Rawls's two principles optimally realizes Rawls's two principles." This is hardly a satisfying answer for someone who is unsure whether we should accept Rawls's two principles as specifying an ideal theory of justice.[4] For our comparative evaluation to provide a compelling explanation for why we should accept a particular candidate theory as specifying the content of our ideal theory, our evaluative standard must stand at some remove from our candidate theories.[5]

---

3 The evaluative criteria and evaluative standard correspond, respectively, to what Erman and Möller call "the conceptual aspect" and "the axiological aspect" of the "process of ideal theory" (EM 537).

4 Stating the answer within Rawls's framework brings out the point: "Because individuals situated in the original position would judge a well-ordered society governed by my two principles as optimally realizing my two principles." Of course, Rawls argues no such thing. Instead, he argues that individuals situated in the original position would judge a well-ordered society governed by his two principles as optimally realizing their fundamental interests in things like liberty, social equality, welfare, community, and so on (see Rawls, 1999b, secs. 20–1). As I argued earlier (W 436), it is natural to interpret the design of the original position and the specification of the parties' interests as representing Rawls's evaluative standard for comparing candidate theories.

5 Erman and Möller misunderstand me on this point when they say, for example, "we typically do not

In sum, an ideal theory is the candidate theory that is fully implemented by the idealistic scenario that ranks highest according to a theorist's evaluative standard. In the arguments that follow, we will say that an ideal theory is *informative for the purposes of the Target View* if its content provides non-redundant information when specifying the principles that characterize the normatively salient attributes of an appropriate long-term objective for real-world political action. Similarly, we will say that an ideal theory is *informative for the purposes of the Benchmark View* if its content provides non-redundant information when specifying the normative principles we should use to rank feasible options for improving upon the social status quo in non-idealistic circumstances. Otherwise, we will say an ideal theory is *uninformative* for the specified practical purpose(s).[6] We limit our attention to determining whether the content of ideal theories is informative for these two purposes.[7] We will briefly address whether ideal theory is informative for other purposes in a concluding section.

2. AGAINST IDEAL TARGETS: A REVISED ARGUMENT

Let's assume that a theorist accepts a particular candidate theory as an ideal theory, which we label T^* to fix a concrete referent. In accordance with the model above, our hypothetical theorist accepts T^* as her ideal theory because it is fully implemented by the highest ranked idealistic candidate scenario, which we label S^*. Our question in this section is whether T^* provides non-redundant information when specifying the candidate theory we should accept as characterizing the normatively salient attributes of an appropriate long-term objective for political action.

Let the *set of feasible scenarios* comprise the model scenarios that represent practically possible social arrangements. We use "feasible" interchangeably with "practically possible" and allow the reader to insert whichever notion of practical possibility they think serves as an appropriate constraint on normative requirements, subject to one condition: we assume that the set of practically possible social arrangements does not

know which feasible world is morally best independent of moral (or in this case political) theories" (EM 530). From the context, it seems clear that they mean to claim that we cannot comparatively evaluate candidate scenarios except on the basis of what I am calling a candidate theory. This interpretation is confirmed by the fact that their alleged "direct counterexample" (EM 534–35) has the question-begging structure discussed in the text.

6 Erman and Möller treat one of my earlier statements about the conditions under which an ideal theory is informative (W 441) as a premise that requires justification (EM 531–33). I meant the statement in question to be true by my definition of "informative", which I had left implicit. I regret my lack of clarity on this point.

7 See Erman and Möller's distinction between the "content" and "process" of ideal theory (EM 535–38). I regret that my earlier arguments did not clearly observe this important distinction.

expand over time, it only contracts (if it changes at all).[8] Regarding the relationship between the set of feasible scenarios and the set of idealistic scenarios, we assume the following: (1) There are some feasible scenarios that are not included in the set of idealistic scenarios; (2) Idealistic scenarios that are included in the set of feasible scenarios (if any) are not necessarily among the feasible scenarios that are ranked highest by the theorist's evaluative standard. These assumptions will be crucial to my arguments, so let's pause to reflect on their plausibility.

Assumption (1) simply acknowledges that some feasible scenarios are inconsistent with a theorist's idealistic assumptions about the circumstances in which candidate theories are imagined to be implemented. For example, suppose a theorist (e.g., Rawls) assumes that citizens of a hypothetical society have an "effective sense of justice"; (1) simply acknowledges that citizens lack an effective sense of justice in some feasible scenarios.

Assumption (2) implies that some feasible scenarios could be ranked higher, according to the theorist's evaluative standard, than any idealistic scenario — that is, than any scenario that is consistent with a theorist's particular assumptions about the circumstances in which candidate theories are imagined to be implemented.[9] Some theorists might deny this assumption by claiming that the point of ideal theory is precisely to identify "the upper limit for what can be realized" (EM 529, note 22; also 536). This is too quick. Idealistic scenarios are such in the sense that they assume circumstances that are thought to be better than the circumstances that obtain in the actual world. But this does not yet imply that the assumed idealistic circumstances are better than any feasible circumstances since the set of feasible scenarios includes model scenarios that represent non-actual social possibilities. To establish that no feasible scenario is normatively better than the highest ranked idealistic scenario, an ideal theorist must comparatively evaluate their idealistic scenarios with the set of feasible scenarios. Merely asserting that idealistic

8 Eschewing a substantive conception of practical possibility avoids Erman and Möller's concern that my argument rests on a controversial feasibility constraint on normative requirements (EM 528). The stated assumption rules out any definition of "practical possibility" that implies that a state of affairs S is not practically possible if it cannot be realized given the status quo circumstances C, even though S could become practically possible because it can be realized given different circumstances $C' \neq C$ and we can take actions to transition from C to C' (cf. Gilabert, 2017, 118–23; also EM 535f). It seems most plausible to say, from the perspective of C, that realizing S is practically possible because we can take actions to bring about circumstances that are conducive to realizing S. This is what I will assume in lieu of a compelling reason to assume otherwise.

9 Recall my earlier point that a theorist's evaluative standard for ranking idealistic scenarios is distinct from the normative theory that is implemented by the highest ranked idealistic scenario. See footnote 4 and the associated text.

scenarios are better than any feasible scenario while ignoring questions about feasibility is unpersuasive.

We now turn to our argument. For concreteness, let $\hat{S}$ denote the best feasible scenario (i.e., the feasible scenario that a theorist would rank highest according to her evaluative standard) and let $\hat{T}$ denote the set of normative principles that are implemented by $\hat{S}$. Plausibly, realizing $\hat{S}$ and, consequently, $\hat{T}$ is the normatively appropriate long-term objective for political action.[10] Realizing T^* is either feasible or not; that is, there is a feasible scenario that implements T^* or there is not.[11]

Suppose realizing T^* is feasible. Does T^* specify the appropriate long-term objective for political action? Only if $T^* = \hat{T}$. Given assumptions (1) and (2), the only way to establish that $T^* = \hat{T}$ is by identifying $\hat{S}$ and determining whether it implements T^*.[12] Given this and the fact that identifying and analyzing $\hat{S}$ is sufficient to specify $\hat{T}$, it is pointless to identify S^* and, in turn, T^* — doing so provides no relevant information beyond that provided by our identification and analysis of $\hat{S}$. Thus, T^* is uninformative for the purposes of specifying $\hat{T}$, the theory that characterizes the normatively salient attributes of the social possibility we should realize in the long term.

Suppose realizing T^* is not feasible. Nonetheless, T^* might be informative if we can infer the content of $\hat{T}$ from T^*. On what grounds could we do this? The most tempting thought is that we should realize the feasible set of principles that is most similar to T^* when realizing T^* is not feasible (e.g., Valentini, 2012, 42). The idea is roughly this. We start by identifying S^* and, in turn, T^*. Upon discovering that realizing T^* is not feasible, we look for the feasible set of principles that is most similar to T^*, which we label $T^\dagger$ to fix a concrete referent.[13]

---

10   Why not take T^* to specify the appropriate long-term objective for political action? Recall my remarks on assumption (2): we can't simply assume without argument that T^* is implemented by the best feasible scenario (as opposed to the best idealistic scenario). But perhaps we can rightly take T^* as the appropriate long-term objective for political action without showing that it is implemented by the best feasible scenario? This seems in tension with the widespread thought that ideal theory is meant to specify a "realistic utopia" (e.g., Rawls, 1999a, 11). My point here is not to challenge the thought that the purpose of ideal theory is to specify the principles that would be implemented by the best feasible scenario. My point is instead that we have no reason to accept the claim that a theory that characterizes the best-case scenario consistent with a theorist's (idealistic) assumptions about background circumstances in fact characterizes the best feasible scenario without taking up the kinds of questions about feasibility that are typically set aside by ideal theorists — indeed, set aside *because* one is doing ideal theory rather than non-ideal theory.

11   Our assumption, above, that candidate theories have numerous possible implementations implies that there can be feasible scenarios that implement T^* that are not S^*.

12   Anticipating a potential objection from Erman and Möller (see EM 529, 530), the point here is not that we need to identify $\hat{S}$ to determine whether realizing T^* is feasible but, rather, to determine whether T^* specifies the appropriate long-term objective for political action on the hypothesis that we should realize $\hat{S}$.

13   My argument here makes no use of any particular similarity measure (see EM 531), so we simply assume

Now, does $T^\dagger$ specify the appropriate end for political action? As above, only if $T^\dagger = \hat{T}$. Given assumptions (1) and (2), the only way to establish that $T^\dagger = \hat{T}$ is by identifying $\hat{S}$ and determining whether it implements $T^\dagger$. Given this and the fact that identifying and analyzing $\hat{S}$ is sufficient to specify $\hat{T}$, it is pointless to identify S^* and, in turn, T^* then $T^\dagger$—doing so provides no relevant information beyond that provided by our identification and analysis of $\hat{S}$. Thus, again, T^* is uninformative for the purposes of specifying $\hat{T}$, the theory that characterizes the normatively salient attributes of a social possibility we should realize in the long term.

In conclusion: we can't specify a normatively appropriate long-term objective for real-world political action without identifying and analyzing $\hat{S}$, which requires engaging questions about feasibility. Whether realizing T^* is feasible or not, doing ideal theory while bracketing feasibility issues is uninformative for the purpose of specifying the principles that characterize the normatively salient attributes of a social possibility we should realize in the long term.

Two remarks before moving on. First, the argument above does not rely on any optimism about the prospects for identifying $\hat{S}$ and, in turn, $\hat{T}$ (cf. EM 529). If we cannot identify $\hat{S}$, then we cannot establish that $T^* = \hat{T}$ nor that $T^\dagger = \hat{T}$; thus, we cannot establish that T^* or $T^\dagger$ characterize an appropriate long-term objective for political action. So much the worse for the Target View. This is one reason to think real-world political action should not be guided by a distant ideal (also Gaus, 2016). Second, the argument above is about whether the content of an ideal theory provides a basis for accepting a candidate theory as specifying the principles we should in fact realize rather than the principles we should *try* to realize or otherwise *pursue*.[14] Assuming $\hat{T}$ refers to the principles we should in fact realize, we may nonetheless do best by aiming to realize a different theory T' because, for example, T' might be a good intermediate goal on the way to realizing $\hat{T}$, or because aiming to realize T' might be the most effective approach to in fact realizing $\hat{T}$ due to certain features of human psychology. Some might now claim, in this vein, that we do best by trying to realize an ideal theory T^* even if we should in fact realize $\hat{T}$. That may be so. But ideal theorists need to provide an argument for that claim, which requires comparing the likely consequences of aiming to realize T^* versus aiming to realize some alternative theory. In any case, my argument here is not about which principles we should try to realize but about the principles that in fact characterize the appropriate long-term objective for political action.

---

that such a measure can be defined without providing any details. If not, we can't make sense of the idea that we should realize the set of principles that is most similar to T^*.

14 On potential reasons to pursue (unattainable) ideals, see Southwood and Wiens (forthcoming).

3. AGAINST IDEAL BENCHMARKS: A REVISED ARGUMENT

By definition, an evaluative standard specifies which criteria are relevant for evaluating social possibilities, as well as their relative importance. It thus specifies the normatively appropriate balance of criteria as well as the appropriate rate at which competing criteria should trade-off against each other when they come into conflict. It will be significant in what follows that the appropriate balance and trade-offs among criteria can be sensitive to the context in which the evaluative criteria are realized (see Farrelly, 2007; also Arvan, 2014). For example, where social inequalities are vast, it can be appropriate to impose substantial restrictions on certain freedoms, whereas such restrictions would be inappropriate in situations of rough social equality; thus, it can be appropriate to impose trade-offs between freedom and equality in the first scenario that would be inappropriate in the second scenario (see, e.g., Rawls, 1999b, 215–16).

   The standard we should use to comparatively evaluate feasible options is the one that codifies the balance and trade-offs among evaluative criteria that are normatively appropriate for the circumstances in which those options are to be implemented. To fix ideas, let $\hat{T}$ denote the set of principles that satisfy this description. Let x and y denote two feasible options for improving upon the status quo in non-idealistic circumstances. By assumption, $\hat{T}$ implies the normatively appropriate ranking of x and y, which is sensitive to the non-idealistic circumstances in which x and y would be implemented. Again, letting S^* denote the best idealistic scenario, with T^* being the candidate theory implemented by S^*, our question is: can T^* provide a basis for specifying the content of $\hat{T}$ as it applies to the ranking of x and y?

   Recall that we accept T^* as our ideal theory because we judge that, among idealistic scenarios, S^* realizes the optimal balance of our basic evaluative criteria as specified by our evaluative standard (which, recall, is distinct from the normative theories that are candidates for acceptance as our ideal theory). Given this, we don't (yet) have a reason to accept that T^* codifies the optimal balance and trade-offs among evaluative criteria for the non-idealistic circumstances in which x and y are to be implemented. Indeed, in view of the points above about the normatively appropriate balance and trade-offs among criteria being sensitive to the context of realization, we should suspend judgment about whether T^* implies the appropriate ranking of x and y. The only way to resolve our uncertainty about whether T^* implies the appropriate ranking of x and y is by inquiring directly about the balance and trade-offs of evaluative criteria that are appropriate for non-idealistic circumstances. Suppose, then, that we undertake this inquiry to resolve

our uncertainty.[15] By assumption, we would conclude that $\hat{T}$ (at least, those aspects of $\hat{T}$ that apply to the circumstances in which x and y would be implemented) codifies the appropriate balance and trade-offs among criteria for the purpose of ranking x and y. We could then compare T^* with $\hat{T}$ to determine whether we have reason to accept that T^* implies the appropriate ranking of x and y. But that last step is pointless — we already discovered what we sought to know by inquiring directly about the balance and trade-offs among criteria that are appropriate for non-idealistic circumstances. So T^* is uninformative for the purposes of comparatively evaluating non-idealistic feasible options.[16]

To be clear, it is consistent with my argument that T^* can provide a constraint on our specification of $\hat{T}$; in particular, our specification of $\hat{T}$ had better imply the same ranking of idealistic scenarios that explains our acceptance of T^* as our ideal theory. But proponents of the Benchmark View seem to have something more ambitious in mind: namely, that T^* can be used directly to comparatively evaluate non-idealistic scenarios (e.g., Robeyns, 2012; Swift, 2008). My argument shows that this ambition is misguided.

4.   ARE IDEAL THEORIES TOTALLY USELESS?

No. I conclude something more limited (as I did in my earlier article; see W 434): the content of an ideal theory fails to provide non-redundant information for the purposes of specifying the principles (i) that we should accept as characterizing the appropriate objective for practical political action (EM 535) or (ii) that we should use to comparatively evaluate feasible options for improving upon the status quo in non-idealistic circumstances (EM 536). There might be other ways in which ideal theories can be useful or informative. Erman and Möller gesture at a potentially fruitful idea: that ideal theories (and the process of specifying their content) might be useful in the way that models are useful in the natural and social sciences, namely, as devices for exploring candidate analyses of the concepts we use to articulate propositions about social and political life — in particular, as devices for investigating the implications of applying candidate conceptual analyses to actual or hypothetical scenarios (EM 531–32, 536, 537; cf. Ismael, 2016; Johnson, 2014). This conceptual exploration could be especially useful for clarifying

---

15   Should T^* guide this inquiry? Following Mills (2005), we should be wary that T^* can distort or mislead this inquiry, given the idealistic assumptions that framed our inquiry and subsequent acceptance of T^*. But see my concluding remarks.

16   I meant my earlier discussion of Rawls (W 442–44) to provide a concrete illustration these more general points, in contrast with Erman and Möller's interpretation of that discussion as "a good old-fashioned first-order normative argument" (EM 534).

our ideas when specifying our evaluative standard (EM 537). This idea should be worked out in more detail. But pointing out that ideal theories can be useful in *other ways* is no reason to think, contrary to my conclusions, that ideal theories are informative for the purposes of the Target and Benchmark Views.

REFERENCES

Arvan, Marcus. 2014. "First Steps Toward a Nonideal Theory of Justice." *Ethics & Global Politics* 7 (3): 95–117.

Erman, Eva, and Niklas Möller. 2022. "Is Ideal Theory Useless for Non-Ideal Theory?" *Journal of Politics* 84 (1): 525–540.

Farrelly, Colin. 2007. "Justice in Ideal Theory: A Refutation." *Political Studies* 55(4):844–864.

Gaus, Gerald. 2016. *The Tyranny of the Ideal: Justice in a Diverse Society.* Princeton and Oxford: Princeton University Press.

Gilabert, Pablo. 2017. Justice and Feasibility: A Dynamic Approach. In Kevin Vallier and Michael Weber, eds., *Political Utopias: Promise or Peril?* New York: Oxford University Press.

Ismael, Jenann. 2016. "A Philosophy of Science Looks at Idealization in Political Theory." *Social Philosophy and Policy* 33 (1–2): 11–31.

Johnson, James. 2014. "Models Among the Political Theorists." *American Journal of Political Science* 58 (3): 547–560.

Mills, Charles W. 2005. "'Ideal Theory' as Ideology." *Hypatia* 20 (3): 165–184.

Rawls, John. 1999a. *The Law of Peoples.* Cambridge, MA: Harvard University Press.

Rawls, John. 1999b. *A Theory of Justice.* 2nd ed. Cambridge, MA: Harvard University Press.

Robeyns, Ingrid. 2012. "Are Transcendental Theories of Justice Redundant?" *Journal of Economic Methodology* 19 (2): 159–163.

Simmons, A. John. 2010. "Ideal and Nonideal Theory." *Philosophy & Public Affairs* 38 (1): 5–36.

Southwood, Nicholas, and David Wiens. Forthcoming. "Devoting Ourselves to the Manifestly Unattainable." *Philosophy and Phenomenological Research.*

Swift, Adam. 2008. "The Value of Philosophy in Nonideal Circumstances." *Social Theory and Practice* 34 (3): 363–387.

Valentini, Laura. 2012. *Justice in a Globalized World: A Normative Framework.* New York: Oxford University Press.

Wiens, David. 2015. "Against Ideal Guidance." *Journal of Politics* 77(2):433–446.