Unpublished draft 3-3-22.

Uncommon Features:

Defending Ideal Theory with Model-to-World Inference[1]

By Olúfẹ́mi O. Táíwò

Abstract

Defenders of ideal theory in social and political philosophy insist that its abstractions and idealizations are justifiable or even indispensable; detractors maintain that these will irretrievably distort reality, pointing out important aspects of political life that are un- or misrepresented by theory. Charles Mills provides an even more pointed and worrying version of this criticism, arguing that ideal theories and theorists act as unwitting or negligent accomplices in oppressive structures. But philosopher Kwasi Wiredu provides an example of philosophy resembling ideal theory and crafted as such in response to racial and colonial domination. I distinguish between models and "model-to-world inferences", the conclusions about the world or our actions that we take them to license. Using this distinction, I argue contra Mills and in favor of Wiredu: that ideal theory and associated strategies of idealization and abstraction can be useful tools for philosophy that opposes social injustice.

I.   Introduction

Defenders of ideal theory in social and political philosophy insist that its abstractions and idealizations are justifiable or even indispensable; detractors maintain that these will irretrievably distort reality, pointing out important aspects of political life that are un- or misrepresented by theory.[2] They argue that ideal theory works on a picture of the world too removed from how the world actually works to guide efforts to fix its problems, thus failing to meet a core obligation they take political theory to have.[3]

---

[1] Thanks to: Serene Khader, David Estlund, Bryce Huebner, James Mattingly, Leif Hancox-Li, Simon Căbulea May

[2] For a compelling example of this sort of argument, see Farrelly 2007. Colin Farrelly, "Justice in Ideal Theory: A Refutation," *Political Studies* 55, no. 4 (2007): 844–64.

[3] See, for example, the argument provided by Knight on disability, which argues that we should "move away from ideal theory and grapple with complex power relations that arise in the real world in order to produce theories capable of guiding action toward emancipatory change." Amber Knight, "Disabling Ideal Theory," *Politics, Groups, and Identities*, 2018, 1–17.

Unpublished draft 3-3-22.

One particular cohort of these detractors, most prominently Charles Mills, advance an even more pointed and worrying version of this criticism: they argue that ideal theories and theorists, by virtue of their failure to take into account specific features of various forms of actual historical domination, act as unwitting or negligent accomplices in oppressive structures, both within professional philosophy and society more broadly.[4] Mills bases his argument on Onora O'Neill's distinction between abstractions and idealizations, arguing both that the going ideal theories don't merely fail to represent as much as they could (abstraction) and also that they introduce falsehoods (idealization).

This argument sits in implicit tension with some other available approaches in Africana philosophy. I argue that Ghanaian philosopher Kwasi Wiredu's writing on "consensual democracy" serves as a prominent example of ideal theorizing within African philosophy, and aim to defend it here from Mills' objection.

Key to the defense from Mills' argument against ideal theory is a point made by Leif Hancox-Li in "Idealization and Abstraction in Models of Injustice", where he argues that many researchers in the literature have gotten something fundamentally wrong.[5] Whether our theories are ideal or non-ideal, whether they involve "idealizations" or "abstractions", in the strictest sense, they just don't themselves do what the participants in the debate accuse them of doing: make claims about the actual world.[6] Then, these particular criticisms of ideal theory rest on something of a category mistake. Alongside Hancox-Li, I argue that many of the supposed problems with ideal theory (and benefits to non-ideal

---

[4] Charles W Mills, "'Ideal Theory' as Ideology," *Hypatia* 20, no. 3 (2005): 165–83.

[5] Leif Hancox-Li, "Idealization and Abstraction in Models of Injustice," *Hypatia* 32, no. 2 (2017): 329–46.

[6] The distinction is O'Neill's. Roughly, abstractions refer to things that are ignored by a theory (the danger being that they may be relevant to real-world instances of the target phenomena) and idealizations refer to falsehoods about phenomena introduced by a theory, perhaps for simplicity or another sense of convenience. Onora O'Neill, "Abstraction, Idealization and Ideology in Ethics," *Royal Institute of Philosophy Lecture Series* 22 (1987): 55–69.

Unpublished draft 3-3-22.

theory) are not problems with based on the content of our theories at all, but how we take those features to relate to features of the world that the model is built to investigate – that is, what I call *model-to-world inference*.[7]

On the view argued for here, the better target for Mills' criticisms are bad practices and norms of model-to-world inference, and that the proffered distinction between ideal and non-ideal theory is at best orthogonal to the concerns that Mills and other scholars invoke to motivate abandoning ideal theory. In part I I will clarify the difference between Mills and Wiredu's metaphilosophy with respect to ideal theory, arguing that defenders of non-ideal theory have an overly narrow conception of how models can inform. In part II I will explain why the criticisms of ideal theory fail and suggest a different diagnosis of the problems in political philosophy attributed to ideal theory by Mills. In part III I conclude with some preliminary suggestions for different ways forward in Africana social and political philosophy, again taking cues from Wiredu's example.

I.   Theories, Models, and Common Features

a.   Contrasting Mills and Wiredu

Many concerns about ideal theory rest on the relationship critics that ideal theory has with the real world. Mills worries that these abstractions either tacitly represent the problems of oppression as simple deviations from the ideal that don't require serious consideration or theorization, or that

---

[7] Cartwright is already wise to this – for example, her distinction between models as parables and as fables appeals to differences in what I would call model-to-world inference: "The advantage of thinking of what happens here in terms of Lessing's account of morals and fables is that it makes clear that there is nothing wrong with the initial experiment. What is wrong vis-à-vis applicability elsewhere is the level at which the conclusion is described." Nancy Cartwright, "Models: Parables v Fables" (Beyond Mimesis and Convention, Springer, 2010), 28.

theorizing about changing the world must start with ideal theory.[8] Others are concerned that will lead to misdiagnoses of social problems, since the idealized version of the world may as such exclude the features of the world that we need to understand flaws – including the history and basic nature of historically specific forms of oppression.[9] Thus, ideal political theories like Rawls' as such involve either a claim about what aspects of the real world are of core importance to understanding how it operates or a claim about what aspects of the world one must keep track of to improve it.

While Rawls' theory of justice can be interpreted as a verbal model aiming to succeed on this sort of criterion, the general aim pursued by ideal theory should not be construed as solely the province of white bourgeois analytic philosophers in the Global North. As political theorist Adom Getachew chronicles in her work *Worldmaking After Empire*, the African anti-colonial activists of the 60s and 70s were also thinking about the abstract rules of a new world order organized around non-domination as an ethical principle and a New International Economic Order as a concrete set of political principles.[10] They had an immediate struggle to fight and win – the fight for national independence from colonial powers. But they didn't want to stop there – they also wanted to use the power they hoped to win to reshape their political structures and lives.

How best to do this restructuring? Getachew refers to their practices of answering this question as a "worldmaking" view, and it's a natural fit with the normative understanding criterion. Moreover, worldmaking was part and parcel of the overall political struggle in which the anti-colonial activists were engaged. As such, the non-ideal theorists have no special claim to the stance of social justice (other than

---

[8] Mills, "'Ideal Theory' as Ideology," 168.

[9] See, for example: Elizabeth Anderson, *The Imperative of Integration* (Princeton University Press, 2010), 4–5; Knight, "Disabling Ideal Theory."

[10] Adom Getachew, *Worldmaking after Empire: The Rise and Fall of Self-Determination* (Princeton University Press, 2019).

Unpublished draft 3-3-22.

the rhetorical advantages caused by the erosion of cultural memory that followed the global defeat of

left projects for justice over the tumultuous years that followed).

Ghanaian philosopher Kwasi Wiredu provides an important example of one worldmaking sort of

enterprise that philosophers responding to the anti-colonial political situation have used ideal theory to

deliberate about. Throughout a series of books and papers, Wiredu articulates and defends a view of the

ideal democratic structure for African societies.[11] He takes "consensual democracy", to be a governance

form that is rooted in African traditions and that was prevalent before European colonialism, but also

takes it to respond to the continent's new challenges for governance left in the wake of the European

colonial era. This theory, and the debates it proliferated, have been a fruitful site for African theorists

and theorizing.[12]

A number of reasons count in favor of interpreting Wiredu's argument as an instance of ideal

theory. First, Wiredu's account is that of the ideal African democratic form – it does not analyze any

particular African society, though it is informed by historical analysis of African traditions. Both he and

his dissentors discuss idealizations and/or abstract relations between abstract classes or groupings of

people, considering how organizing around "lineages" might contrast fruitfully with party systems, and

how the outcome of consensual deliberation might better position "minority" groups for social

cooperation than under a party system.[13] While Wiredu makes reference to some actual history, the

---

[11] Kwasi Wiredu, "Democracy by Consensus: Some Conceptual Considerations," *Philosophical Papers* 30, no. 3 (2001): 227–44; Kwasi Wiredu, "Democracy and Consensus in African Traditional Politics: A Plea for a Non-Party Polity," *The Centennial Review* 39, no. 1 (1995): 53–64; Kwasi Wiredu, *Cultural Universals and Particulars: An African Perspective* (Indiana University Press, 1996).

[12] See, for example, this recent edited volume: Emmanuel Ifeanyi Ani and Edwin E. Etieyibo, eds., *Deciding in Unison: Themes in Consensual Democracy in Africa* (Vernon Press, 2020).

[13] I agree with Hamlin and Stemplowska that the difference between abstraction and idealization is insufficiently clear to apply to complex cases. For a critical discussion of Wiredu's point on minority groups in particular, see Etieyibo. Alan Hamlin and Zofia Stemplowska, "Theory, Ideal Theory and the Theory of Ideals," *Political Studies Review* 10, no. 1 (2012): 50–

Unpublished draft 3-3-22.

scope of the specific historical structures it considers does not come close to matching the scope of

societies the argument aims to prescribe actions for – Wiredu's historical investigations largely revolve

around Ashanti social structure, a society that occupied much of what is now Ghana. For the people of

Benin or Guinea-Bissau to adopt governance principles mined from this history is not obviously a non-

ideal theoretic response to their particular political situation. From their perspective, the kind of

theorizing Wiredu is doing is fact-insensitive, which counts in favor of viewing it as ideal theory on

Volacu's conception thereof, which treats fact sensitivity as a consideration relevant to the classification

of a theory as ideal or non-ideal.[14]

Then, non-ideal theorists' insistence that ideal theory must make reference to specific

categories and histories of oppression gains artificial plausibility because it takes place in the domain of

Anglophone political philosophy, in which the defenders of ideal theory are associated with old white

men like Rawls. But the substance of his criticism seems to throw those of us interested in doing the

kind of theory Wiredu was doing under the same dialectical bus. Wiredu's highly idealized approach is

not a failure to understand the complexities of the particular political histories that structure political

life on the continent. Far to the contrary: it is motivated by those very complications. Wiredu does not

theorize about the particular history of, say, the social categories of Hutu and Tutsi relations in Rwanda

because he is trying to guide political action on the scope of the entire continent, which calls in favor of

the kinds of abstractions that will guide decision-making in Togo as well. The kind of generality and

abstractions that ideal theory trucks in, in this context, don't constitute an attempt to ignore political

context but, one might imagine, a constructive response to the intractability of the alternative:

51; Edwin Etieyibo, "African Consensual Democracy, Dissensus and Resistance," *Deciding in Unison: Themes in Consensual Democracy in Africa*, 2020, 135.
[14] Alexandru Volacu, "Bridging Ideal and Non-Ideal Theory," *Political Studies* 66, no. 4 (2018): 887–902.

attempting to navigate what is by some measures the world's most diverse continent by a political approach that avoids generalizing across different groups and histories.[15]

b. Models and Theories

If the preceding is right, then there's something to be said for abstraction and generalization in general. But what role should these play in theorizing, and to what extent should we rely on them? I turn here to a more pointed discussion of theories and theorizing, to help bring out the stakes of the disagreement between Mills and Wiredu.

Theories, whether about racial justice or about the migratory patterns of buffalo, involve or are constituted by abstract relationships between abstract concepts.[16] We build theories because there's something in the world we want to understand or navigate practically, and we think the network of concepts we build through theory will help. Theories generate principles that we can use to guide our investigation and observation of phenomena in the world.

But it takes some work to apply the principles to the world and things within it. One reason for this is that theory is not connected to the world in the same sense that our empirical observations are. One important sense in which theory is disconnected from the world gives rise to some of the anxieties that inform Mills' attack on ideal theory. Theory involves abstractions and idealizations from the messy details of the real world: some features of the target phenomena and its worldly context are left in and

---

[15] "A Revealing Map of the World's Most and Least Ethnically Diverse Countries," *Washington Post*, accessed July 27, 2021, https://www.washingtonpost.com/news/worldviews/wp/2013/05/16/a-revealing-map-of-the-worlds-most-and-least-ethnically-diverse-countries/.

[16] Nancy Cartwright, "Models and the Limits of Theory: Quantum Hamiltonians and the BCS Models of Superconductivity," *IDEAS IN CONTEXT* 52 (1999): 242.

some are left out. Theorists' decisions about which to leave in and which to leave out are quite

meaningful, as both defenders and detractors of ideal theory have rightly maintained. Our choices at

the levels of abstraction at which political philosophers tend to operate (whether ideal or non-ideal

theorists) will especially constrain and load the deck for various downstream effects of the theory,

whether its effects on our concrete political projects or even just the attention hierarchies of philosophy

as an academic discipline.

To establish a connection to the world that we build theories to help us understand and

navigate, theory makes use of models and functionally similar tools. A model is an "interpreted

structure": a device that purports to inform us about "target systems": phenomena in the world or the

aspects of them we're interested in learning about.[17] There's a tight relationship between theories and

models: on one candidate positive view of theories applying to (at least) the domain of the natural

sciences, the "semantic" view of theories, theories simply are sets of models.[18] Less ambitiously, we

might content ourselves with observing that theories make use of models as tools that mediate between

theory and world.[19]

When we see models discussed as models, they are often formal mathematical models (e.g.

systems of equations) or computations models. These are important and scientifically useful classes of

models, but they are not the only sort. We can build replicas of natural phenomena with physical

materials (concrete models) or verbally articulate chains of ideas that stand in for models that could be

---

[17] Michael Weisberg, *Simulation and Similarity: Using Models to Understand the World* (Oxford University Press, 2012), 15, 172.

[18] I do not take a stance here on the precise relationship of theories to models, as it is enough for my purposes that there is one. Cartwright, "Models and the Limits of Theory: Quantum Hamiltonians and the BCS Models of Superconductivity," 241; Robert W Batterman and Collin C Rice, "Minimal Model Explanations," *Philosophy of Science* 81, no. 3 (2014): 349–76.

[19] Margaret Morrison and Mary S Morgan, "Models as Mediating Instruments," *Ideas in Context* 52 (1999): 10–37; Nancy Cartwright and Robin Le Poidevin, "Fables and Models," *Proceedings of the Aristotelian Society, Supplementary Volumes* 65 (1991): 55–82.

fleshed out with any of our other modelling approaches (narrative models). This last sense of "model" is the sense that most obviously applies to political philosophy.

Rawls' theory of justice, for instance, is particularly amenable to re-description as a series of verbal models. Rawls begins with an initial, guiding model of the "well-formed society". The well-formed society resembles the world on some dimensions, in that it is populated by people with moral powers that we take each other to generally have, and is governed by familiar sounding structures like a political constitution. It is also helped along by some important and load-bearing abstractions (e.g. from the contingencies of a particular political history or demographic composition) and idealizations (e.g. that the society is "closed" to outside forces) from things that are true about the actual world.

From there, other models are introduced that extend some of the initial assumptions and suspend others: perhaps most notably, the original position and the second, international tier of Rawlsian theory, which eliminates the assumption of a closed society which was used to develop the conception of domestic justice). The end goal of these was to investigate the principles of justice that would govern Rawls' well-formed society and the supposedly separate principles that would govern a global consortium of such societies, which may inform our orientations towards the less than perfectly formed societies we find in the world.

c. Desiderata for models and theories

One way of characterizing ideal theory in political philosophy is to point out its regulative function: ideal theory there produces a characterization of the just society, which we can use as an evaluative standard for the various versions of non-ideal reality that history puts on the table.[20]

---

[20] Ingrid Robeyns, "Ideal Theory in Theory and Practice," *Social Theory and Practice* 34, no. 3 (2008): 346.

Unpublished draft 3-3-22.

Valentini insists that there are three separate debates misleadingly grouped under this heading: 1)

whether full or partial compliance is assumed by a theory; 2) the extent to which a theory is "utopian"

rather than "realistic"; and 3) whether the theory is conceived as a conception of the "end-state" at

which justice is achieved or is a "transitional" account of how we might make our society more just.[21]

Hamlin and Stemploska treat the distinction between idealization and abstraction as a fourth.[22]

Whichever characterizations of the debate we accept, the crux of the issue is the extent to which

theories match up to the target phenomena in the world, and what justifications succeed or fail to

vindicate the use of these theories when they do not.[23]

The question of what justifications we can give for theories and models turns on the prior

question of what it is that we can accomplish with them. There are a multiplicity of desiderata for

theories and models, each implying different standards of evaluation. Since not all of them require or

even involve relationships between features of the world and corresponding features of the model, the

different desiderata imply differences in which model-to-world inferences are licensed – that is, what

descriptions of the world we can take a theory or theorist to be committed to because of what their

theory or model looks like.

The desiderata considered here adapt the distinctions offered in Michael Levins' seminal paper

"The Strategy of Model Building in Population Ecology" which explains these as a response to the

---

[21] Laura Valentini, "Ideal vs. Non-ideal Theory: A Conceptual Map," *Philosophy Compass* 7, no. 9 (2012): 654–64.
[22] Hamlin and Stemplowska, "Theory, Ideal Theory and the Theory of Ideals."
[23] Volacu suggests a two-axis schema: fact sensitivity/insensitivity on one axis and prioritization of desirability/feasibility on the other. I more or less agree with Volacu's schema here and take it that my sentence here roughly approximates these axes. Volacu, "Bridging Ideal and Non-Ideal Theory," 893.

Unpublished draft 3-3-22.

complications of "overlapping but not identical goals of understanding, predicting, and modifying

nature" to reflect the kinds of considerations relevant to political philosophy and normative theorizing.[24]

One possible desideratum is descriptive accuracy. This desideratum licenses the sort of models

that we build to get a deeper understanding of how target systems in the world function. Levins uses the

example of a tendency in fishery biology to build models tightly around the variables relevant to the

short term behavior of the organism of interest. When we improve a model with respect to this

criterion, we make the model itself resemble the target system in the world more strongly. This sounds

like what non-ideal theorists have in mind when they complain that ideal theory distorts reality, and that

non-ideal theory is better suited to studying the real world effects of oppression. In the extreme case,

we an adopt an ideal on which we try to make the model match the target system as perfectly as

possible, which philosopher of science Michael Weisberg calls "completeness".[25]

Since features of a model built according to this criterion are attempts to match features of the

world, such a model may license model-to-world inferences from features of the model to descriptive

commitments about what the target system is actually like. This seems to fit Mills' implicit picture of

how political theories ought to relate to political realities, and would provide a stable ground for

interpreting a theory's silence about race and gender as an affirmative statement that these are

unimportant axes of oppression in the actual world.

But there are other desiderata that we might use to evaluate models and refine them, which

wouldn't license such model-to-world inferences. Another is predictive accuracy, a desideratum geared

and predicting the behavior of a target system or elements within it. Fulfilling this criterion is often in

---

[24] Richard Levins, "The Strategy of Model Building in Population Biology," *American Scientist* 54, no. 4 (1966): 422.

[25] Michael Weisberg, "Forty Years of 'The Strategy': Levins on Model Building and Idealization," *Biology and Philosophy* 21, no. 5 (2006): 626.

Unpublished draft 3-3-22.

tension with the previous one. Levins gives the example of "clearly unrealistic" models: in biology, he describes mathematical models that investigate biological phenomena with equations that neglect critical aspects of the descriptive story like the effect of a species' population density on its own rate of increase, which he likens to models in physics that assume frictionless surfaces or "perfect" gases.[26] These models nevertheless achieve other goals: for instance, they allow us to predict population dynamics across many species.

We refine models with respect to this criterion when their *output* – predictions – more strongly matches the behavior of the target system or its constituent parts. But that may not involve making the model *itself* more representative of the real world at all – it might require introducing more or different idealizations, or changing the relative contribution of unrepresentative parts of the model to its output. Accordingly, there's no stable relationship between the elements of models built to excel at this desideratum and any particular description of the world – then, model-to-world inferences that treat a model's content or omissions as statements about what the target real world phenomenon is like are unlicensed.

Action guidance is a third potential desideratum. To meet this criterion, a model should provide useful and actionable predictions about the world, or otherwise inform practical projects in which we are engaged. This is tightly related to the previous criterion, since the most obvious ways of using models practically hinge on how well they do at predicting behavior of or in a target system. But this criterion focuses on a subset of predictions: the ones relevant to our potential practical intervention. A model that succeeds in this way gives us actionable intelligence, which could be increased even by a refinement of a model that degrades overall predictive accuracy.

---

[26] Levins, "The Strategy of Model Building in Population Biology," 422.

Unpublished draft 3-3-22.

For example, a political communication model may recommend a series of ads that genuinely increase a candidate's vote share in an election, despite (or even because!) it makes incorrect predictions about how the total viewership will respond. This might correspond, for example, to a situation where the model predicts that an ad will help out a candidate by convincing "independent" voters not to vote for the other party's candidate, when it in fact simply convinces a higher share of them (whose minds were already made up) not to skip election day. This would be a case where the model succeeds despite resting on a blatantly false "description" of the world – yet another case where we are not licensed to infer directly from the features of the model to descriptive claims about the corresponding feature of the world.

A fourth criterion for models: normative understanding.[27] This criterion is also related to the previous criterion, but perhaps the most peculiar to political philosophy and ethics.[28] The action guidance criterion uses the practical aims we already have to evaluate the models: but those practical aims, themselves, ought to be evaluable. What interventions ought we be trying to make, and what would their consequences be? What should our long term goals be, supposing we succeed in the short term at eliminating the obstacles that prevent us from long term control of our social environment? Several of these are among the reasons Rawls gives for his method of inquiry in Justice as Fairness: A Restatement.[29]

---

[27] This corresponds to something like what Sen labels the "transcendental" approach to justice. Amartya Sen, "What Do We Want from a Theory of Justice?," *The Journal of Philosophy* 103, no. 5 (2006): 215–38.

[28] See, for example, Estlund's distinction between "concessive" and "aspirational" theory – the former makes concessions in theorizing about how we can expect people to respond, the latter does not. Both, Estlund thinks, are kinds of theory that are potentially worth doing. David Estlund, "Utopophobia," *Philosophy & Public Affairs* 42, no. 2 (2014): 113–34.

[29] John Rawls, *Justice as Fairness: A Restatement* (Harvard University Press, 2001), 2–5.

Unpublished draft 3-3-22.

Models can also be used to investigate these questions, since they represent ways both of clearly stating and investigating the causal connections of the world we hope to intervene in. We may be interested in setting targets, evaluating long-term objectives that may be a ways off from an immediate practical perspective, or simply directly evaluating current aspects of our current world without commitment to any particular practical response.[30] The models or theories we construct for such purposes may diverge even further from a description of the actual world than with the previous criterion, as we wheel in heavily idealizing assumptions like "full compliance", the assumption that everyone will do as our moral standard or rule requires.[31] The models or theories involved with reaching this criterion are perhaps least likely to license any particular inferences from the content of the models to any particular descriptive claims about the real world.

II.   Beyond Common Features

a.   Directly Answering the Non-Ideal Theorists

I now focus on negative arguments that switching to non-ideal theory is unlikely to help us diagnose or solve the problems Mills and other theorists have attributed to ideal theory. The main problem with the arguments against ideal theory as such is that the problems attributed to ideal theory as such are better described as problems with model-to-world inference.

First, a brief refresher of Mills' case against ideal theory. In "Ideal Theory as Ideology", Mills asks, echoing much of the non-ideal theory literature that followed:

---

[30] Estlund makes the last point about the separability of evaluation and (any particular) practical intervention in: David Estlund, *Democratic Authority: A Philosophical Framework* (Princeton University Press, 2009), 265–67.

[31] Hamlin and Stemplowska, "Theory, Ideal Theory and the Theory of Ideals."

Unpublished draft 3-3-22.

Can it possibly serve the interests of *women*, white and nonwhite, to ignore female

subordination, represent the family as ideal, and pretend that women have been treated as

equal persons? Obviously not. Can it possibly serve the interests of *people of color* to ignore the

centuries of white supremacy and to pretend that a discourse originally structured around white

normativity now substantively, as against just terminologically, includes them? Obviously not.

Can it possibly serve the interests of the *poor and the working class* to ignore the ways in which

an increasingly inequitable class society imposes economic constraints that limit their nominal

freedoms and undermine their formal equality before the law? Obviously not.[32]

But, as this paper suggests, I don't think this is quite as obvious as advertised. It *is* obvious that

there is a non-accidental connection between the tendencies of ideal theories and ideal theorists to and

the forms of oppression and domination Mills points out. What is not obvious is that the connection is

causal in the particular way needed for the argument to go through: that the privileging of ideal theory

over non-ideal theory *itself* exacerbates these structures of domination, whether in the domain of

philosophy or more generally.

Of course, we could give an alternative etiology of the problem, which would in turn suggest a

different functional relationship between philosophers' reluctance to engage real world forms of

oppression and domination and the social structures they are thereby ignoring. Perhaps the tendency

of the literature to ignore actual injustice reflects the preferences selected for by an unfair world, but

exerts no causal influence on that world's unfairness. For example: perhaps it is true that classical

music aficionados are overrepresented amongst philosophers. But if it were also true that both classical

music and academic philosophy select for a certain class background we could explain the covariance

---

[32] Mills, "'Ideal Theory' as Ideology."

Unpublished draft 3-3-22.

here without suspecting that we had uncovered evidence that classical music appreciation among

academic philosophers is a significant aspect of the class-based disadvantage inherent to capitalism.

Without the veneer of analogy: perhaps it is simply the case that philosophers' reluctance to

engage real world forms of oppression and domination and their penchant for ideal theory are simply

both symptoms of the same disease – that philosophy selects for the people least likely to take it upon

themselves to research oppression - rather than the latter causing or enabling the former. Then, the

reluctance of philosophers to engage real world forms of oppression and domination is explained by the

sociological forces that produce *philosophers* in general, not the theoretical decisions that produce ideal

theory, and ideal theory plays no role in the explanation.

A more serious challenge to Mills' framing of the issue here involves returning to the two key

assumptions referred to in the introduction: the common features account of models' role in knowledge

production, and the related assumption that models should be taken to make descriptive claims about

real-world systems. Recall that a *common features account* of what makes theories or models

informative requires that they share common features with the aspects of the real world they

investigate.[33] This would ground the model-to-world inference that Mills makes here, when he takes it

that the targeted aspects of the ideal theory "model" are committed to describing real-world political

systems.

For all he says here, Mills could even be read as presupposing a much stronger claim: Weisberg's

completeness ideal for models, the representational ideal on which every aspect of the target system

must be represented in the model).[34] However, my arguments against his position work equally well

against the weaker "common features" reading of his argument's implicit commitments, which seems

---

[33] Batterman and Rice, "Minimal Model Explanations," 351–57.

[34] Weisberg, "Forty Years of 'The Strategy': Levins on Model Building and Idealization."

the more charitable interpretation of his position. For the remainder of the paper I will assume the

common features reading in my response.

Mills' discussion here involves not only this narrow view of models, but the second mistake as well, which is helped along by the first. Reading a common features account working in the background would make sense of why Mills identifies the assumption of equality among persons as a distortion of the actual world rather than simply a potentially innocuous or even informatively distortive feature of the relevant models as such. Such an account of how political theories work, after all, would take them to be making implicit claims about the real-world counterparts of their component features. That is why kind of idealizations Rawls makes about the well-formed society amounts to pretending that "women have been treated as equal persons" in the real world rather than simply a description of the model world under construction.[35]

In "Idealization and Abstraction in models of Justice" Hancox-Li introduces an effective response to this point by discussing an abstract "collaboration game" model developed by Justin Bruner and Cailin O'Connor.[36] Agents play the game multiple times and maximize their payoffs by using strategies that have proved effective for them or other players in the past. Bruner and O'Connor show that, when the population is divided into a majority and a minority group, it effects what strategies the players employ and learn from: majority players make larger demands of minority players, and minority players prefer to collaborate with other minorities or not at all. Hancox-Li points out that the model predictions are consistent with a real world phenomenon: women preferring to collaborate with women in multiple fields of inquiry.[37]

---

[35] Mills, "'Ideal Theory' as Ideology."
[36] Justin Bruner and Cailin O'Connor, "Power, Bargaining, and Collaboration," 2016.
[37] Hancox-Li, "Idealization and Abstraction in Models of Injustice," 337.

Unpublished draft 3-3-22.

But Bruner and O'Connor's model of abstract "minorities" maximizing numerical payoffs does not reflect a number of complications that actual women face. One idealization Hancox-Li focuses on is that players in Bruner and O'Connor's model only learn from previous versions of this same collaboration game. Women researchers, on the other hand, have presumably learned about gender in a variety of contexts before they ever became researchers (in home, in early education, etc.). Is the model distorting reality in an objectionable way? Hancox-Li's answer:

> It might seem that I have made a dodgy move in claiming that the model can *remain silent* on other sources of learning instead of *introducing the falsehood* that they do not exist [emphasis his]. I confess that I find it hard to distinguish clearly between these two alternatives, since "remaining silent" and "introducing a falsehood" *may only be metaphorically applied to models, as opposed to speakers or texts* [emphasis mine]. Some readers may have the intuition that omitting other sources of learning in the collaboration game counts as introducing a falsehood even if we are interested only in the question of how people learn from collaboration alone. But if we follow these readers' intuitions and take the exclusion of real-world causal factors from a model to always be introducing falsehoods, we are led to the conclusion that all social-science models introduce falsehoods, and are thus idealized in the way that Mills deplores.[38]

Hancox-Li's point involves distinguishing between the content of a model and a description of a target system. Empowered with the previous section, we can also flesh this point out further. First, Mills seems to assume that we should treat models as descriptions of the target system and hold them to the criterion of descriptive accuracy. However, three out of the four possible model desiderata considered here would license viewing the model as potentially informative or useful without saddling it with the

---

[38] Hancox-Li, 339.

expectation that it should be built on common features shared with the part of the real world we aim to use it to learn about.

I don't mean here to advocate for this particular construal of "normative understanding" models as a reading of Rawls or anyone else in particular. Instead, I mean to point out two things. First, if Rawls failed to adequately theorize about race, his use of ideal theory is not the reason why. The claim that racial justice (or any other sort) is not an important category from the standpoint of justice is in no sense implied or licensed by its omission from a proffered model of justice.[39] Models, unlike researchers, do not speak. As such, we should be slower to decide what it means that x or y is not represented in a model.

Secondly, if those of us invested in racial justice (or any other sort) mean to accomplish it, we shouldn't decide what the possibilities for an entire approach to knowledge production are by way of these sorts of arguments, which I argue misunderstand the possibilities for models and modelling in our moral and political life. If we, for example, took the relevant criterion for models serving an "ideal" political theory to be something like "normative understanding", then the models produced to provide normative understanding could play the role of adjudicating between the appropriateness of different racial justice-pursuing strategies, which could be planned out using still other models.

b. Theory in the Natural Sciences Goes Far Beyond Common Features

Reflection on the use of models in the natural sciences helps explain a broader point about them that explains why they might fit *worldmaking* projects: that they can be informative even when they do not resemble the aspect of the world that they are meant to inform us about. Sometimes they are informative about the world precisely because of their dissimilarities or distortions. Discussions

---

[39] Tommie Shelby argues for something like this in: Tommie Shelby, "Racial Realities and Corrective Justice: A Reply to Charles Mills," *Critical Philosophy of Race* 1, no. 2 (2013): 145–62.

Unpublished draft 3-3-22.

about models amongst our colleagues, especially on the subject of explanation in philosophy of science, are potentially informative for political philosophers on this issue.

Morrison and Morgan describe models as "partially independent" of both the world and the theories they help precisify. Models themselves and their constituent parts represent aspects of the theory that informs their construction, things or relationships in the world that motivated the inquiry in the first place, or both at once.[40] They are inspired by, guided by and are sometimes even logically deducible (or otherwise directly derivable) from the theory or theories they derive from.[41] But the extent of their independence is important: they often respond to or contain elements meant to represent empirical features of the world that have been observed but have no counterpart in a guiding theory of a target system, and even in some cases include elements that contradict the principles of the guiding theories.[42] Independence, then, can be thought of as a constraint on the inferences that are licensed by features or behavior of the model: a model is independent of the world or a theory to the extent that it features relationships with no real-world or theoretical analogue.

This is no mere cop-out caveat designed to save time and ward off objections about this or that hastily designed feature of some model. The independence of models and theories from the real world can be a functional aspect of their role in knowledge production. For instance, Nancy Cartwright gives the example of "Galilean experiments" in which "unrealistic assumptions are not a hindrance but a necessity", when a researcher's model ignores all other factors to maximize potential understanding of

---

[40] Recall that on the semantic view theories simply are sets of models, in which case this argument works just as well for "theories". But, as before, I decline to take a stance here on the precise relationship between theories and models. Morrison and Morgan, "Models as Mediating Instruments," 11.

[41] Stephan Hartmann, "Models and Stones in Hadron Physics," *Models as Mediators: Perspectives on Natural and Social Science* 52 (1999): 2.

[42] Michael Redhead, "Models in Physics," *The British Journal for the Philosophy of Science* 31, no. 2 (1980): 147.

the factor of interest.[43] Speaking more generally in "Models as Mediators", Morrison and Morgan

explain that it is the partial independence from both theory and world that allow models to perform the

function that they do – that is, it's *because* the model isn't fully determined by either the theory or that

data that either expected or unexpected predictions or other output of the model or other behavior of

its components is informative.[44] We compare the behavior or predictions of the model with the

relevant, non-identical features of the real to adjust our model, our background theory, our

understanding of the real world, or some combination of these.

Batterman and Rice even identify an entire class of models that are overall neither similar to nor

representative of the real world, which they call "minimal models".[45] Even such "thoroughgoing

caricatures of real systems" can be informative, if the features the model does have nevertheless inform

us about the phenomenon of interest.[46] They use the Lattice Gas Automaton computational model of

fluid flow to illustrate their point. In this model, a set of point particles are arranged on a hexagonal

lattice. A set of rules dictates how these particles move, when and if they collide, and how collisions

alter the trajectory of the colliding particles. This is not how molecules move in actual fluids – the

model-to-world inference that real particles move in hexagonal lattice-like fashion would involve a

misunderstanding of the model. Yet, the large-scale behavior of the model predicts large scale

behaviors of actual fluids.[47]

Minimal models are just one way to put pressure on a common features account as a total

explanation of what makes models informative. Lauren Ross and Mazviita Chirimuuta identify still other

examples of models in neuroscience whose explanatory utility doesn't depend on similarity to the target

---

[43] Cartwright, "Models: Parables v Fables," 23.
[44] Morrison and Morgan, "Models as Mediating Instruments," 17.
[45] Batterman and Rice, "Minimal Model Explanations."
[46] Batterman and Rice, 350.
[47] Batterman and Rice, 360–61.

Unpublished draft 3-3-22.

system, some of which may be "minimal models" in Batterman and Rice's sense and some of which may

be yet another method to learn about the world other than rebuilding it in miniature.[48]

All the kinds of models these authors consider share *some* common features with their target

systems, but this observation by itself is trivial – we could always engineer "common features" by simply

relaxing our standards of what counts as common. What defenders of the common features account

would need to invalidate the point being made here is to establish that the common features between

the model and target system explain why the model is informative. That does not seem to be the case

in either of Batterman and Rice or Chirimuuta's cases.

I don't rely on any of these authors' specific descriptions of how their specific models of interest

explain and inform. What's important about their examples is what they reveal about the role of

models in explanation quite generally: that similarity to the world is not a limitation on how, why, or

whether models inform. If this is so, then this takes some of the wind out of the sails of Mills' objection

against ideal theory, that ideal theories are too dissimilar from the messy unjust world to inform us

about them.

c.   Reconsidering the Role of Common Features in Ideal Theory

The discussion in philosophy of science is not a perfect parallel to the discussion in political

philosophy on ideal theory – but that is in part because the discussion in philosophy of science on

models and explanation is live to a fuller spectrum of ways we can learn about the world than the ideal

theory debate is. David Kaplan's "3M constraint" requires both that the components of a model

correspond to components of the target system and that relations between these also correspond to

---

[48] Lauren N Ross, "Dynamical Models and Explanation in Neuroscience," *Philosophy of Science* 82, no. 1 (2015): 32–54; Mazviita Chirimuuta, "Minimal Models and Canonical Neural Computations: The Distinctness of Computational Explanation in Neuroscience," *Synthese* 191, no. 2 (2014): 127–53.

Unpublished draft 3-3-22.

causal relations between the parts of the target system – a particularly strict version of a common features account of models and a clear competitor to the "minimal models" approach to understanding why models explain.[49] But Kaplan and Carver) are clear that this is a constraint on *explanation*, in particular, at most demarcating those models that explain from those that do not.[50] They treat as open the well established possibility that even models that do not properly explain (in the particular mechanistic sense they defend) might nevertheless meet some other standard of informativeness (e.g. "empirical adequacy").

The lesson for political philosophy, then, is that the approach assumed by non-ideal theorists is not actually required. Having a background theory that pre-judges which aspects of the real world are causally relevant, and then populates the model with just those features that matter is only one way to proceed – the only way that Mills, Knight, and others seem to hold that political ideal theory could work. Political philosophy ought to take a page from this book and avoid tying its own hands by demands for representation – which involves casting a critical eye on the criticisms of ideal theory that presuppose similarity between theory and models.

Pair Rawls' and Wiredu's theory with Batterman and Rice's discussion of large scale fluid flow, for example. Rawls does not suppose that his toy society in *A Theory of Justice* is much like any society in the real modern world, much less all of them: for instance, Rawls explicitly concedes that we may live in a "corrupt society" rather than the "well-ordered" one he prefers to theorize about, but sees value in this very fact. Ideal theory can tell us ways in which our social world might have been different and

---

[49] David Michael Kaplan, "Explanation and Description in Computational Neuroscience," *Synthese* 183, no. 3 (2011): 347.
[50] David Michael Kaplan and Carl F Craver, "The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective," *Philosophy of Science* 78, no. 4 (2011): 602–3.

Unpublished draft 3-3-22.

might well become different in the future.[51] Similarly, while Wiredu does seem to claim that many

African societies that have had consensual democratic bases, he can neither charitably be read as

making this claim about literally all of them nor as equating Zulu political history of consensual

structures with that of Ashanti's political history with the same. Then, to recommend consensual

structures for all African societies is not to treat Asanteman as a universally and trans-historical

description of all African societies but is better understood as an act of idealization past relevant

sociohistorical differences.

Also similarly, Batterman and Rice claim we needn't (and in fact shouldn't, given what

subsequently became proven about sub-atomic particles - like the fact that they exist) think that the

model of fluid flow provided by the Lattice Gas Automaton (LGA) model resembles how particles in

flowing fluid actually move. Nevertheless, a wide variety of engineering contexts make use of the

model.[52] To criticize the LGA model because it represents particles as moving along a hexagonal lattice

(and particles are not in fact constricted in this way) would involve a fundamental misunderstanding of

how the model is attempting to inform us about the world. Using the LGA needn't involve licensing the

model-to-world inference that real-world particles move in a lattice-like fashion – just that representing

the movement of particles this way can help the model make useful predictions about the flow of actual

fluids.

Similarly, defenders of Wiredu's theory of consensual democracy need not argue that the entire

African continent (comprised of well over a billion people) will return to exactly those structures of

conversation and social organization that structured the much smaller Ashanti society. Nevertheless,

ethical principles developed with reference to that society and a set of abstract values might well be

---

[51] Rawls, *Justice as Fairness: A Restatement*, 37–38.
[52] Batterman and Rice, "Minimal Model Explanations," 357.

Unpublished draft 3-3-22.

informative for the continent. Likewise, defenders of Rawlsian theory can concede that neither the

members of the Rawlsian original position nor the interactions among them are much like real world

interactions between people who find themselves enmeshed in histories of domination, yet Rawls'

model could still be informative.[53]

d. Answering an Objection: Is Normative Theory Special?

One distinction that might seem particularly useful to the anti-ideal theorist is the distinction

between theories with descriptive and normative ends. So far, I have mostly confined my discussion to

a treatment of models as such, leaving many possible distinctions by the wayside. I've cut wildly across

different environments and domains of inquiry that feature models, and different kinds of models and

model-like reasoning. But perhaps there are moral norms governing normative theorizing that do not

apply to other kinds, and these norms either independently require something of a "common features"

approach to theory and model building in political philosophy or otherwise account for the success of

arguments in favor of common features-based theories that might otherwise fail.

The objection gets going something like this: some theories and acts of theorizing are primarily

about understanding the world. Maybe Galileo was just curious about whether or not Jupiter had any

moons, and that by itself was a good enough reason to find out. This sort of theorizing is perhaps more

common in the natural and social sciences, which licenses their ways of building and evaluating theories

and models, or at least differentiates their practices from those that are appropriate in more

normatively oriented disciplines.

But other theories are - or at least should be - in the business of making the world better. It

would be callous and objectionable to spend a good deal of social resources developing a theory of a

---

[53] For a brief defense of Rawls whose basic strategy makes use of this possibility, see David Wiens, "Against Ideal Guidance.," *Journal of Politics* 77, no. 2 (April 2015): 433–46.

Unpublished draft 3-3-22.

disease that we did not intend to cure or treat. This last analogy is especially on the nose, since this seems something like the charge laid at the feet of ideal theory.

What exactly is different about theorizing that is normatively laden? One possibility is that our acts of theorizing are in and of themselves morally laden. The objection that stems from this position is that ideal theory is objectionable because of what it *does* – abstracts from real world conditions of oppression and injustice – rather than what comes of its so doing.

But it's hard to read those who attack ideal theory as having meant this version of the objection. Mills, for example, asks whether the idealizations of ideal theory can possibly "serve the interests" of women and of racially oppressed people, which implies a consequential argument. But the plausibility of this argument hinges far too much on the particular examples. It's a good deal easier to say how the idealizations of Wiredu's ideal theory might serve the interests of, say, ethnic minorities on the African continent who might be ignored altogether in contexts like Nigeria's, where explicit reference to ethnicity (particularly by the larger groups) has historically driven domination and conflict.[54] It is less likely that Mills has pointed out a crucial problem with ideal theory *as such* here and more likely that Mills has identified particularly important aspects of social domination that any theory or model, ideal or not, could benefit from including.

It's also worth taking a second look at how the burden of proof has been assigned here. It is less than obvious what the practical payoff of an opposition to ideal theory in favor of explicit theorizing about oppressed groups will be. Adolph Reed and Merlin Chowkwanyun argue that a spate of social science inspired by motivations like Mills' has uncovered a great deal of evidence about a great many specific racial disparities in specific contexts, but this very specificity comes with a considerable cost:

---

[54] Usman Mohammed, "Corruption in Nigeria: A Challenge to Sustainable Development in the Fourth Republic," *European Scientific Journal* 9, no. 4 (2013).

obscuring the very causal factors that we would need to intervene on to prevent them.[55] In the aptly

titled article "Fuck Nuance", sociologist Kieran Healy argues persuasively that nuance is only superficially

attractive as a criterion of social theory, and may serve as an impediment to learning when valued

inappropriately.[56] Pointing out racial disparities may be good argumentative fodder against those who

deny the existence, reach, or importance of racism. But suppose we already were persuaded that

racism exists and pervades society. Does it materially and significantly help the racial justice struggle to

spend social resources proving that discrimination pervades kindergarten education slightly more in

Tulsa than it does in Kansas City? Would such granularity "serve the interests" of Black and brown

students in either case, who are confronting a system of racial injustice that cuts across the differences

in their contexts? It is precisely the kind of unity across differences that motivates the use of

idealizations and abstractions. Non-ideal theorists seem to take it to for granted that more empirical and

historical analysis of the racial origins of such disparities is of practical use in responding to them, but it

is far from clear how such a focus helps, which undermines the advantage of practical relevance they

take themselves to have over ideal theorists. This challenge is particularly pressing if the

attentional/resource cost or research design that allows us to uncover the disparity precludes

investigation of the sort of factors that might explain what is different about Tulsa rather than simply

reveal that Tulsa is different.[57]

The possibility that replacing ideal theory with empirically laden non-ideal theory could be

counterproductive pushes back effectively against even the more deontological framing of the problem:

that it's somehow intrinsically immoral or unjust to theorize without explicit reference to axes of

---

[55] Adolph L Reed and Merlin Chowkwanyun, "Race, Class, Crisis: The Discourse of Racial Disparity and Its Analytical Discontents," *Socialist Register* 48 (2012): 149–75.
[56] Kieran Healy, "Fuck Nuance," *Sociological Theory* 35, no. 2 (2017): 118–27.
[57] Reed and Chowkwanyun, "Race, Class, Crisis: The Discourse of Racial Disparity and Its Analytical Discontents," 152–53.

oppression like race, gender, and disability. It seems plausible that whatever principle would recommend making explicit reference to these would be a subsidiary principle in a larger set of principles oriented around the broader themes of racial, gender, and disability justice. How could it be more morally weighty to follow the subsidiary principle that recommends that we socially signal via our models that we care about, say, racial justice, than to actually advance the political project of dismantling structural racism? Shouldn't a higher level principle override in this sort of instance? If not, what could count in favor for constructing our set of racial justice principles in so self-defeating a fashion? Opponents of ideal theory should tell us.

Even if we take Mills to have correctly assigned the burden of proof, the same kind of arguments that undermined the principle based advocacy of abandoning ideal theory also undermine an approach that focuses on the consequences of different kinds of theorizing. We can group the consequences under consideration by domain.

First, consequences internal to academia. Perhaps "ideal theory" of the Rawlsian or Wiredian sort will lead to bad consequences in the discipline itself because ideal theory will act as an ideology or will otherwise signal strong commitments to researchers about what aspects of the world are important to a description of how it works. That is, maybe not explicitly theorizing about gender, disability, or race communicates that a political philosopher need not learn about these phenomena to effectively do their job.

But we could even concede to Mills that the particular ideal theories, as received by the discipline, function ideologically in this sense without being further committed to preferring non-ideal theory or avoiding ideal theory. We could, for instance, adopt a race, ability, and gender-blind (for example) ideal theory and pair it with a non-ideal theory or set of these that were conscious about the aforementioned aspects of identity, thus fulfilling both the desiderata of theorizing about oppression

Unpublished draft 3-3-22.

explicitly and whatever ideal theories might be good for. Or we could campaign to change how

philosophers are educated on axes of oppression directly: by changing the content of graduate

requirements, holding symposia on these issues for mid-career philosophers, and other interventions of

this sort.

If any of these are potentially effective responses, then the criticisms of ideal theory are perhaps

better directed at the lacunae of appropriate, matching non-ideal theories rather than at the content of

the ideal theory. If it is inadequate, we are owed an answer why – and, again, an appeal to what the

ideal theory has or lacks seems orthogonal to the problem, unless it is established that an ideal theory

precludes examination of systems of oppression. That strong claim is not established by observing that

the ideal theory simply doesn't itself examine oppression in the desired ways. It is hard to attribute

whatever negative consequences stem from oppression-silent ideal theory to the content of the ideal

theory itself, rather than the broader epistemic environment in which it intervenes.

Second, the consequences external to academia. Perhaps anti-ideal theorists are concerned

about the effects of ideal theory in the world – either downstream of the previous set of concerns,

because of the influence philosophers' reception of theory has outside of the discipline, or perhaps in

some direct sense that bypasses how other philosophers respond to ideal theory. The latter seems

implausible, given how seldom academic papers are read at all.[58]

Also, the earlier response helped along by Reed and Chowkwanyan puts pressure on the most

plausible explanation of the problem here: that theorists' inattention should be understood as a

negative consequence for social justice. But say we had gotten rid of ideal theory and that resulted in

---

[58] Biswas and Kirchherr claim that the average peer reviewed paper is fully read by "no more than ten people" but do not explain how they arrived at this number. Asit Biswas and Julian Kirchherr, "Prof, No One Is Reading You," Text, The Straits Times, April 11, 2015, https://www.straitstimes.com/opinion/prof-no-one-is-reading-you.

theorists taking axes of oppression more seriously, as the non-ideal theorists advocate for. Would that have been a positive development for the struggles for social justice? Perhaps, but perhaps not – if what researchers do when they study social justice is distracting or otherwise counterproductive. There seems to be at least some evidence that at the very least, despite laudable intentions, researchers' effects on the injustices they study are not uniformly positive. Yet more reason that we ought to focus on the actual effects our theories have, rather than the abstract categories we sort them into.

III.   Conclusion

Ideal theory is a tool – nothing more, nothing less. The most skillfully designed guitar won't play beautifully while used as a door stop; the car with the finest engine will not outrace a lemon while it's in park. We would know straightaway what had gone wrong with the reasoning of any person who looked to find flaws in theguitar or the car themselves in these cases for the explanation of why they failed to be a part of bigger and better things. Similarly, those who are concerned about ideal theory have learned the wrong lesson from its misuses and non-uses. If we want ideal theory, or philosophy in general, to serve the cause of ending oppression then it is our task to appropriate it and set to that purpose, or to find a tool that is better based on *its* potential usess and not the present machinations of its owners.

There are other ways to navigate the world besides rebuilding it in miniature. The debate about the relative virtues of ideal theory vs. non-ideal theory is, at best, simply orthogonal to both the question of how to avoid the risks that have thus far been attributed to problematic idealizations and the more important question of how theory could be emancipatory or progressive. At worst it is a costly and self-undermining distraction from more pressing concerns: if the problems pinned on ideal theory

are actually problems with model-to-world inference, then switching to non-ideal theory is would at best solve the problems indirectly and accidentally, and risks obscuring them or worsening them.

Perhaps ideal theory isn't hegemonic and ideological because of what it *does* contain (problematic idealizations or abstractions) but because of what the broader field *doesn't*: a sustained, systematic or otherwise intentional way of working from our models to practical projects. That would provide guidance each step of the way about which model-to-world inferences were licensed by our various theories. As things stand now, the problems identified with ideal theory and its accoutrement apply more strongly and directly to our interpretation of models than to our construction of them.

To make this point about model-to-world inference is not to defend ideal theory as currently practiced, but to contend that ideal theory itself as an approach is a confused choice of target and it is unclear what a move to non-ideal theory would do to remedy the problem as understood in this way. I criticize the debate as it stands precisely because I share the motivation to avoid the field's current pitfalls, and thus see the presentation of non-ideal theories as a solution to or even necessarily an improvement with respect to these problems as misleading about what theory that takes those problems seriously would look like. Unlike our colleagues in other fields and even other corners of philosophy, model-to-world inference is vastly undertheorized in political philosophy and endangers the relevance and helpfulness of our work as a field.

Even if it is so that race, gender, disability, and other such categories ought to be represented such specifically in "theory" (done by theorists in their professional capacity), this is no argument for them needing to appear in *theory* in the strong sense (as direct components of relevant theories). That is, *that* they should be on the agenda of political philosophers ought to be separated from the question of *how.* If it turns out that we don't need to explicitly think about gender or race to do a specific kind of political theorizing, it does not follow either that these phenomena are unimportant or that we do not

Unpublished draft 3-3-22.

need to think about race to do political philosophy more generally speaking. We could challenge the

wholly unmerited association of abstraction with importance directly by appropriately valuing applied,

translational, and other forms of philosophy other than ideal theory rather than the overcorrection of

demonizing a perfectly legitimate theoretical enterprise.[59]

As mentioned in the introduction, the argument in this paper is not in the least designed to

defend political philosophers' *use* of models, be they Rawlsian or Wiredian or in some other camp. But

ideal theory as such is neither an explanation of nor a defense of this tendency of some philosophers.

It's hard to see why discussions about the adequacy of Wiredu's theory itself should stand or fall on the

good or bad behavior of Rawlsians – or why our appraisal of the general theoretical approach should be

sensitive to either camp's, as opposed to a sober analysis of what the approach could be if it were held

in better hands.

Ideal theory may have a positive role to play. We may find ourselves with question's like Rawls

about what to hope for our of a just social system. We may find ourselves in a context like Wiredu's – he

wrote in the aftermath of the massive transformation of African politics that followed the successful

anti-colonial movements of the 60s and 70s and the unsuccessful attempt to forge a just way forward

from there. He had to abstract heavily from a particular African histories to respond to the total range of

contexts he was interested in: the full African continent. Ideal theory was useful for these enterprises

and is likely useful for still others.

If theory really has the causal role in either supporting or potentially dismantling systems of

oppression that non-ideal theorists speak as though it has, then that is all the more reason to evaluate

---

[59] "Translational" philosophy is "about creating a practical intervention to address a real world problem", relating theory to the real world in a practically oriented sense. The authors distinguish this from applied philosophy, which relates theory to real world phenomena in an epistemically oriented sense. Margaret Little et al., "ETHICS LAB," *A Guide to Field Philosophy: Case Studies and Practical Strategies*, 2020.

Unpublished draft 3-3-22.

our theories with a ruthless focus on their adequacy. If our theories really are genuine ammunition in

the fight for justice, a little idealization is a small price to pay.

Unpublished draft 3-3-22.

Works Cited

Washington Post. "A Revealing Map of the World's Most and Least Ethnically Diverse Countries." Accessed July 27, 2021. https://www.washingtonpost.com/news/worldviews/wp/2013/05/16/a-revealing-map-of-the-worlds-most-and-least-ethnically-diverse-countries/.

Anderson, Elizabeth. *The Imperative of Integration*. Princeton University Press, 2010.

Ani, Emmanuel Ifeanyi, and Edwin E. Etieyibo, eds. *Deciding in Unison: Themes in Consensual Democracy in Africa*. Vernon Press, 2020.

Batterman, Robert W, and Collin C Rice. "Minimal Model Explanations." *Philosophy of Science* 81, no. 3 (2014): 349–76.

Biswas, Asit, and Julian Kirchherr. "Prof, No One Is Reading You." Text. The Straits Times, April 11, 2015. https://www.straitstimes.com/opinion/prof-no-one-is-reading-you.

Bruner, Justin, and Cailin O'Connor. "Power, Bargaining, and Collaboration," 2016.

Cartwright, Nancy. "Models and the Limits of Theory: Quantum Hamiltonians and the BCS Models of Superconductivity." *IDEAS IN CONTEXT* 52 (1999): 241–81.

———. "Models: Parables v Fables," 19–31. Springer, 2010.

Cartwright, Nancy, and Robin Le Poidevin. "Fables and Models." *Proceedings of the Aristotelian Society, Supplementary Volumes* 65 (1991): 55–82.

Chirimuuta, Mazviita. "Minimal Models and Canonical Neural Computations: The Distinctness of Computational Explanation in Neuroscience." *Synthese* 191, no. 2 (2014): 127–53.

Estlund, David. *Democratic Authority: A Philosophical Framework*. Princeton University Press, 2009.

———. "Utopophobia." *Philosophy & Public Affairs* 42, no. 2 (2014): 113–34.

Etieyibo, Edwin. "African Consensual Democracy, Dissensus and Resistance." *Deciding in Unison: Themes in Consensual Democracy in Africa*, 2020, 135.

Farrelly, Colin. "Justice in Ideal Theory: A Refutation." *Political Studies* 55, no. 4 (2007): 844–64.

Getachew, Adom. *Worldmaking after Empire: The Rise and Fall of Self-Determination*. Princeton University Press, 2019.

Hamlin, Alan, and Zofia Stemplowska. "Theory, Ideal Theory and the Theory of Ideals." *Political Studies Review* 10, no. 1 (2012): 48–62.

Hancox-Li, Leif. "Idealization and Abstraction in Models of Injustice." *Hypatia* 32, no. 2 (2017): 329–46.

Hartmann, Stephan. "Models and Stones in Hadron Physics." *Models as Mediators: Perspectives on Natural and Social Science* 52 (1999): 326.

Healy, Kieran. "Fuck Nuance." *Sociological Theory* 35, no. 2 (2017): 118–27.

Unpublished draft 3-3-22.

Kaplan, David Michael. "Explanation and Description in Computational Neuroscience." *Synthese* 183, no. 3 (2011): 339.

Kaplan, David Michael, and Carl F Craver. "The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective." *Philosophy of Science* 78, no. 4 (2011): 601–27.

Knight, Amber. "Disabling Ideal Theory." *Politics, Groups, and Identities*, 2018, 1–17.

Levins, Richard. "The Strategy of Model Building in Population Biology." *American Scientist* 54, no. 4 (1966): 421–31.

Little, Margaret, Elizabeth Edenberg, Sydney Luken, and Jonathan Healey. "ETHICS LAB." *A Guide to Field Philosophy: Case Studies and Practical Strategies*, 2020.

Mills, Charles W. "'Ideal Theory' as Ideology." *Hypatia* 20, no. 3 (2005): 165–83.

Mohammed, Usman. "Corruption in Nigeria: A Challenge to Sustainable Development in the Fourth Republic." *European Scientific Journal* 9, no. 4 (2013).

Morrison, Margaret, and Mary S Morgan. "Models as Mediating Instruments." *Ideas in Context* 52 (1999): 10–37.

O'Neill, Onora. "Abstraction, Idealization and Ideology in Ethics." *Royal Institute of Philosophy Lecture Series* 22 (1987): 55–69.

Rawls, John. *Justice as Fairness: A Restatement*. Harvard University Press, 2001.

Redhead, Michael. "Models in Physics." *The British Journal for the Philosophy of Science* 31, no. 2 (1980): 145–63.

Reed, Adolph L, and Merlin Chowkwanyun. "Race, Class, Crisis: The Discourse of Racial Disparity and Its Analytical Discontents." *Socialist Register* 48 (2012): 149–75.

Robeyns, Ingrid. "Ideal Theory in Theory and Practice." *Social Theory and Practice* 34, no. 3 (2008): 341–62.

Ross, Lauren N. "Dynamical Models and Explanation in Neuroscience." *Philosophy of Science* 82, no. 1 (2015): 32–54.

Sen, Amartya. "What Do We Want from a Theory of Justice?" *The Journal of Philosophy* 103, no. 5 (2006): 215–38.

Shelby, Tommie. "Racial Realities and Corrective Justice: A Reply to Charles Mills." *Critical Philosophy of Race* 1, no. 2 (2013): 145–62.

Valentini, Laura. "Ideal vs. Non-ideal Theory: A Conceptual Map." *Philosophy Compass* 7, no. 9 (2012): 654–64.

Volacu, Alexandru. "Bridging Ideal and Non-Ideal Theory." *Political Studies* 66, no. 4 (2018): 887–902.

Weisberg, Michael. "Forty Years of 'The Strategy': Levins on Model Building and Idealization." *Biology and Philosophy* 21, no. 5 (2006): 623–45.

Unpublished draft 3-3-22.

———. *Simulation and Similarity: Using Models to Understand the World*. Oxford University Press, 2012.

Wiens, David. "Against Ideal Guidance." *Journal of Politics* 77, no. 2 (April 2015): 433–46.

Wiredu, Kwasi. *Cultural Universals and Particulars: An African Perspective*. Indiana University Press, 1996.

———. "Democracy and Consensus in African Traditional Politics: A Plea for a Non-Party Polity." *The Centennial Review* 39, no. 1 (1995): 53–64.

———. "Democracy by Consensus: Some Conceptual Considerations." *Philosophical Papers* 30, no. 3 (2001): 227–44.