

Idealization and Abstraction in Models of Injustice

LEIF HANCOX-LI

Charles Mills has argued against ideal theory in political philosophy on the basis that it contains idealizations. He calls for political philosophers to do more nonideal theory, namely political theory that pays more attention to the most visible oppressions in society, such as those based on race, gender, and class. Mills's argument relies on a distinction between idealization and abstraction. Idealizations involve adding false assumptions to one's model, which is unacceptable, whereas abstractions merely leave out details without undermining descriptive power. By studying formal models of injustice, I argue that the idealization/abstraction distinction is unhelpful. Either the distinction exists only relative to one's modeling purposes, or all models in political theory contain idealizations. Either way, the distinction does not help Mills's cause. Furthermore, there are arguments from philosophy of science for the epistemic benefits of idealizations. However, Mills's call for greater emphasis on the most visible mechanisms of oppression can be supported without relying on an idealization/abstraction distinction. I provide three alternative reasons for why we should prefer political theories that place more emphasis on race-, class-, and gender-based oppression.

I. INTRODUCTION

In his essay on ideal theory in political philosophy, Charles Mills calls for political philosophers to theorize in a way that is more in line with the “distinctive experience of the oppressed” (Mills 2005, 166). Such theorizing is underrepresented in Anglophone political philosophy, which focuses on ideal theory. John Rawls's theory of justice, for example, comes in for critique on this front, since it does not incorporate actual race- and gender-based oppression. Mills takes the distinction between ideal and nonideal theory to be one between “idealizing and nonidealizing approaches to ethical theory” (166). He argues that ideal theory is too unrealistic to guide our actions. Instead, he suggests, ideal theory is an ideology perpetuated by those who benefit from its dominance.

Hypatia vol. 32, no. 2 (Spring 2017) © by Hypatia, Inc.

Mills relies on a distinction between idealization and abstraction first drawn by Onora O'Neill (O'Neill 1987). Feminists have often resisted idealizations in ethics. One common criticism is that idealized theories claim to be universalist in application, but ignore the distinctive experiences of oppressed groups (O'Neill 1990; Schwartzman 2006b). In particular, political theory has been charged with ignoring facets of human life that are coded feminine, such as dependency and care issues (Jaggar 1988; Kittay 1998). However, no theory can capture the particular experiences of every individual, so some abstraction is necessary. For Mills and O'Neill, abstractions are acceptable, but idealizations are not. I criticize the utility of this distinction by examining Mills's analogy with scientific models. By studying formal models of injustice, which describe the oppressive mechanisms that Mills thinks are important, I argue that we are faced with two options: either the abstraction/idealization distinction is context-sensitive, or all social-science models contain idealizations. Either way, the abstraction/idealization distinction does not help Mills's cause. Furthermore, arguments in philosophy of science indicate that idealized models can be more explanatory than realistic models.

Despite these criticisms of Mills's idealization/abstraction distinction, I sympathize with Mills's call for political philosophers to place more emphasis on the most visible mechanisms of oppression. I provide three reasons to support his call independently of a distinction between idealization and abstraction. First, one could acknowledge that certain idealizations are better than others, based on ethical considerations. Second, one could argue for placing more emphasis on models of injustice in political philosophy than there currently is in order to promote cognitive diversity. Third, some formal models of injustice have successfully passed tests of robustness that theories like Rawls's may not.

In short, the idealization/abstraction distinction that Mills relies on is either untenable or unhelpful, but there are still good reasons for political philosophy to pay more attention to race- and gender-based oppression.

II. DEFINING IDEAL THEORY

In defining ideal theory, Mills starts by disentangling the different senses of "ideal" at play. The first, most harmless sense is that of *ideal-as-normative*: the idea that ethics is a normative theory that describes how the world could be better. This is not the sense of "ideal" he is attacking, since it is one that is assumed by any ethical theory (Mills 2005, 166). Next is the *ideal-as-descriptive*: a model of a system may be an ideal model in the sense that it abstracts away from some details of the concrete system but still retains descriptive power (166). This, we will see, corresponds to O'Neill's idea of what abstraction is.

Finally, we have what Mills calls *ideal-as-idealized* models (167). These consist of

1. models that represent some limiting situation that does not obtain in reality, such as the ideal gas model, where gas particles don't interact;

2. models that represent an instance of a machine with an impossibly “perfect” standard of functioning, such as a perfectly efficient engine.

These models, we will see, correspond to what O’Neill thinks of as idealized models.

Mills thinks that in political philosophy, ideal-as-descriptive models are acceptable, but that ideal-as-idealized models model society as what it *should* be like, analogous to how an ideal-as-idealized model of an engine may model it as having 100% efficiency. This is unacceptable because ideal-as-idealized models neglect most actual mechanisms of injustice, such as those based on racism. Furthermore, users of ideal-as-idealized models do not sufficiently explore how different their models are from ideal-as-descriptive models (170). Mills thinks that marginalized groups tend to be more critical than dominant groups about ideal-as-idealized models in political philosophy because of ideology. Ideal-as-idealized models serve the interests of dominant groups by focusing philosophers’ attention on the kinds of oppression experienced by these groups. Meanwhile, more visible kinds of oppression, such as those experienced by racial or gender minorities, are decentered from philosophers’ attention.

Mills relies on a distinction between idealizations, which distort reality, and abstractions, which omit some details but remain descriptive of the target system. This reiterates O’Neill’s distinction between idealization and abstraction (168). O’Neill characterizes abstract reasoning as reasoning that “leaves out a great deal” (O’Neill 1987, 55) and idealized models as those in which “much (too much) that is false of human agents is *added*” (56; emphasis hers).¹ Thus, idealizations involve explicit statements of falsehoods, whereas abstractions involve mere omissions that are neutral on whether the target system has the omitted property. O’Neill names the following as examples of idealizations of human agents: complete information, infallible powers of calculation, and independence from institutional and ideological context (56).

Mills’s distinction between ideal theory, which is ideal-as-idealized, and nonideal theory, which is ideal-as-descriptive, is only one possible usage of “ideal” and “nonideal” in political theory. Laura Valentini outlines three different uses of these terms, each tracking a different distinction. She cites Mills as an example of using “ideal” to label less realistic theories, and “nonideal” to label more realistic theories (Valentini 2012, 658). Given that Mills thinks nonideal theories are more descriptively accurate than ideal theories, I think her classification of Mills is correct. I will use “ideal” and “nonideal” in Mills’s sense throughout this article without committing to whether it fits with how some of his targets, like Rawls, use the terms.

III. DESCRIPTIVE FACTS IN NORMATIVE MODELS

Mills’s discussion of idealization and abstraction uses scientific models as an analogy. The main body of my article discusses the idealization/abstraction distinction in models of injustice. Like scientific models, these are descriptive models. One might

wonder how the abstraction/idealization distinction operates in the *normative* models that are Mills's target.

I take no stance on the precise role of realistic assumptions in normative theory. I make only a minimal assumption that many normative theories take some facts about the world as a constraint. Rawls, for example, wants his theory to be consistent with basic principles of moral psychology (Rawls 1999, 498). Philosophers disagree on how far facts should constrain normative theory—Valentini reviews theorists running the gamut from those who think that Rawls's theory incorporates too many realistic assumptions to those who think it is insufficiently realistic (Valentini 2012). Mills falls into the latter camp. He thinks ideal theory makes ontological assumptions, for example, that are false about the world (Mills 2005). Similarly, we saw above that O'Neill objects to false assumptions about human agents that appear in some normative theories. It is enough for me that many political theorists think that normative theory should rely on *some* descriptive facts, and that Mills aims his discussion of idealization at these facts. I argue that the *descriptive component* of any normative theory ought not to be criticized on the basis of a context-independent idealization/abstraction distinction. Thus this article is relevant to theorists like Wiens, who think that social science has to be considered concurrently with moral theory in normative political theory (Wiens 2015).²

Perhaps more careful consideration of normative theories would lead us to conclude that the considerations of robustness, the epistemic benefits of idealizations, and so on that I discuss later have very different implications than the ones they have for descriptive theories. However, given that many philosophers want normative theories to include some descriptive assumptions, it seems at least noteworthy that mere falsehood of descriptive assumptions does not by itself undermine the descriptive strength of the theory. How this then affects the normative recommendations of the theory is beyond the scope of this article.

Another possibility is that Mills's criticism is intended to apply only to idealizations in comprehensive doctrines, and not to those in more partial theories. However, the arguments I make will suggest that it is likely impossible to construct a comprehensive doctrine that does not contain idealizations in some contexts. In that case, Mills's critique could be reconstrued as the pretense of these doctrines to be comprehensive in the first place. It is not individual falsehoods that are problematic, but the desire for comprehensiveness in light of these falsehoods.³

In the next section, I look at whether the idealization/abstraction distinction works for formal models of injustice, concluding that either it doesn't, or all social-science models are idealized. Lisa Schwartzman has made a similar argument that there is no such thing as "value-free" abstraction (Schwartzman 2006a). I think this point appears starker when we look at formal models of injustice, which model the factors that Mills thinks are important in political philosophy, but which are highly idealized. Furthermore, as I argue in section V, there are arguments in philosophy of science for why idealized models may sometimes be preferred to more realistic ones.

IV. FORMAL MODELS OF INJUSTICE

SCHELLING SEGREGATION MODELS

Thomas Schelling pioneered a class of formal models on how segregation occurs (Schelling 1971). He considers mechanisms of segregation that are based solely on decisions made by individuals to move or not based on the “color” of the inhabitants of a certain environment.⁴ He then considers a process of iterated turns. In each turn, individuals decide to move or not based on the color conditions of their environment. There are further rules about where they would move if they do move. Schelling finds that segregation easily occurs on the basis of such individual decisions.

What idealizations or abstractions might have gone into the model? One potential idealization is the limitation of segregation to one mechanism. By focusing on individual decisions based solely on neighboring color conditions, Schelling omits mechanisms that he acknowledges are probably more important in real-life segregation, such as organized action, and the general process by which the poor get separated from the rich (Schelling 1971, 144).⁵ Furthermore, Schelling admits that there is no clear distinction between segregation based on individual choices and coerced segregation (145). Households may feel “coerced” by social pressure into choosing segregation, for example. So it’s not just the omission of mechanisms that’s false, but also the assumption that these mechanisms can be separated.

Applying O’Neill’s definitions, leaving out coercive mechanisms does involve leaving out a great deal that is true, which would count as an abstraction. But this omission involves an assumption that no interaction occurs between “individual” and “coercive” mechanisms, and this does seem to be *adding* something false, thus making it an idealization. Furthermore, O’Neill states that the assumption of “independence from the institutional and ideological context we inhabit” is an idealization of human agents (O’Neill 1987, 56). Schelling’s assumed noninteraction between individual and coercive mechanisms does seem to be such an assumption. Turning to Mills’s definitions, the omission of mechanisms counts as representing a limiting situation, akin to omitting intermolecular interactions in the ideal gas model. He takes these limiting models to be ideal-as-idealized models—those that pervade ideal theory and ought to be avoided. Thus, Schelling’s omissions would count as idealizations under both O’Neill’s and Mills’s definitions.

Having established that Schelling’s models contain idealizations, I now consider whether they contain abstractions that are not idealizations, that is, abstractions that leave out details but do not “assert falsehoods.” One such candidate simplification is the omission of the internal structure of the household. Each household is modeled as an “individual” with no internal structure. But it is plausible that members of households have different attitudes toward segregation, and that power relationships internal to the household determine how those attitudes interact to produce a household decision. Perhaps single-person households would behave very differently from

nuclear families. Perhaps intra-household power relations change according to how segregated the neighborhood is, whether a household is immediate neighbors with people with similar attitudes, and so on. Someone interested in these factors could use Schelling-type models with additional complexity added into moving decisions to incorporate the effects of internal household structure. For someone investigating how single-person households might contribute to segregation differently from nuclear family households, abstracting away from internal structure is an idealization, since it would fail to describe the difference between the two kinds of households, which is the very purpose of the investigation. Whether an omission in a model counts as an idealization or an abstraction therefore depends on one's modeling purposes.

We can make a similar point about another omission in Schelling's models. The models consider household decisions made purely on the basis of color, a proxy for race. This fails to incorporate intersectional factors, such as how the gender, age, class, and household structure of one's neighbors interact with race to influence moving decisions. Perhaps one would be less likely to move in the presence of neighbors who are middle-class nuclear families of a minority race, as compared to neighbors who are single, poor men of the majority race. Is this failure to incorporate intersectional factors an abstraction or idealization? To answer this we need to determine if the omission is descriptively harmful. But this depends on one's modeling purposes. If one is interested in isolating the causal contribution of color alone to segregation, then perhaps it is a harmless abstraction. But if one is interested in modeling segregation in neighborhoods where there are good reasons to think that intersectional considerations like those described above apply, then ignoring intersectional considerations counts as an idealization. Again, abstraction and idealization are not terms that apply to a model *simpliciter*, but depend on one's modeling purposes.

Even in natural science, modeling purposes can determine which omissions matter. The ideal gas model can be viewed as ideal-as-descriptive, even though Mills classifies it as ideal-as-idealized. Intermolecular interactions do not matter for many properties of certain gases, so ignoring them does not undermine the descriptive power of the ideal gas model *for the purposes of these properties, for certain gases*. But the ideal gas model could be ideal-as-idealized when modeling situations where intermolecular interactions matter. Similarly, the omission of intersectional factors in segregation models may have nonnegligible effects if our research question concerns the causal contribution of these factors. The question of whether the effects of a simplification are negligible, then, depends on our modeling purposes. Since models in political philosophy can have different purposes, there cannot be an absolute distinction between idealization and abstraction in political philosophy.

Given the contextual dependency of what counts as an idealization, one might be tempted to draw an absolute distinction between abstractions and idealizations by saying that abstractions are those omissions that are descriptively harmless in all contexts. However, this move would render all social-science models idealized. Given the complexity of human societies, it is hard to think of an omission that would not be descriptively relevant in some context. In short, either the distinction between idealization and abstraction is purpose-dependent, or all social-science models are

idealized. If the former, then Mills's distinction between ideal and nonideal theory cannot depend on the idealization-abstraction distinction unless a modeling purpose is fixed. If the latter, then the distinction is unhelpful for his argument, since every model of society would be idealized.

Is it possible, however, that Mills might already subscribe to the thought that the idealization/abstraction distinction is purpose-dependent? Is he doing so when he writes that the question of how useful an ideal-as-idealized model will be cannot be answered *a priori* (Mills 2005)? There, he uses the analogy of an inclined plane: an ideal-as-idealized model of a frictionless plane may work for modeling a smooth, Teflon-coated plane in a vacuum, but not for a Velcro-covered plane. I have two responses to this hedge of Mills's. First, he is claiming roughly that ideal-as-idealized models work if they are "close enough" to reality, so there is a continuous spectrum from ideal-as-descriptive models to ideal-as-idealized models. However, neither Schelling models nor the game theory models I next consider are effective in virtue of being close to reality—at least not to the extent that a frictionless plane is similar to a Teflon-coated plane in a vacuum. In fact, I argue in section V that being distant from reality can be a virtue in a model. Second, if we take Mills to agree with me that the idealization/abstraction distinction is purpose-dependent, then many of the claims he makes become unintelligible. For example, he claims that no one could seriously say that the "ideal-as-descriptive model has approximated the ideal-as-idealized model" in ethics (Mills 2005, 171). This claim is made without any reference to modeling purposes, contrasting with my view that the same model can be both ideal-as-descriptive and ideal-as-idealized, depending on one's modeling purposes. Mills also invokes "ought is supposed to imply can" as a reason to reject ideal-as-idealized theories, because they attribute inhuman capabilities to humans (171). This assumes an absolute distinction between what humans can do and what they cannot do, a distinction independent of modeling purposes. Thus, I think my claim that the abstraction/idealization distinction is purpose-dependent is not consistent with Mills's article.

MODELS OF COLLABORATION

Another example of formal modeling of social injustice is Justin Bruner and Cailin O'Connor's work on how power relationships affect the benefits that accrue from academic collaborations. I will describe the models they use and explain how they also demonstrate the futility of a context-independent distinction between abstraction and idealization.

Bruner and O'Connor model interactions between possible collaborators using a game-theoretical model known as the stag hunt. They describe the payoffs for players in the game as follows:

Two hunters can either choose to hunt stag or hare. The payoff for successfully hunting hare is lower than the payoff for successfully hunting stag

(less meat). But two hunters are needed to take down a stag, while one alone can catch a hare. If one actor hunts stag and her partner does not, she gets nothing. (Bruner and O'Connor forthcoming, 4)

Figure 1, from Bruner and O'Connor (forthcoming), presents the payoffs of this game. Hunting a stag represents collaboration, and hunting a hare represents writing a single-authored paper.

If the players decide to collaborate, then the prize is divided according to a second game known as the Nash demand game:

The Nash demand game is a model of simultaneous bargaining. Two actors want to divide a resource, and each may demand some portion of it (50%, for example, or 10% or 92%). If the demands do not exceed the total resource, both actors receive their demand. If the demands are too high, though, the actors receive payoffs called the disagreement point (often nothing). This corresponds to a scenario in which bargaining fails. Figure 2 shows a payoff table for a Nash demand game with restricted payoffs. In this game actors can demand either 4, 5, or 6 of a total resource of 10 (we refer to these demands as Low, Med, and High). In the rest of the paper, we will model bargaining with a restricted game of this sort. It is tempting to interpret these demands literally when thinking about bargaining in collaborative scenarios. High is first author, Low is second author. Given that levels of effort tend to differ in collaborative work, it

		Player 2	
		Stag	Hare
Player 1	Stag	3,3	0,2
	Hare	2,0	2,2

Figure 1 Payoff table for the stag hunt (Bruner and O'Connor forthcoming). Payoff for Player 1 is listed first; payoff for Player 2 second.

		Player 2		
		Low	Med	High
Player 1	Low	4,4	4,5	4,6
	Med	5,4	5,5	0,0
	High	6,4	0,0	0,0

Figure 2 Payoff table for the Nash demand game (Bruner and O'Connor, forthcoming). Payoff for Player 1 is listed first; payoff for Player 2 second.

is more fruitful here to think of these demands as relative to effort. The High demand can be thought of as a demand for a relatively prestigious author position compared to effort exerted on the paper. The Low demand can be thought of as agreeing to a relatively poor author position compared to effort. The Med demand can be thought of as an attempt to distribute credit equitably. (Bruner and O'Connor forthcoming, 5)

The stag hunt, followed by the Nash demand game, combine to form what Bruner and O'Connor call a collaboration game. Agents play the collaboration game multiple times and learn winning strategies by favoring strategies that maximized their own payoffs in the past and imitating strategies that worked for their peers.

The next step is to assume that the players in the game come from two populations. One then investigates how players' strategies evolve when they can make decisions depending on which population the other player in the game comes from. Suppose co-authors of the same gender tend to negotiate more equitable outcomes than co-authors of different genders. In such conditions, Bruner and O'Connor find that minority group members learn to make smaller demands from majority group members, but majority group members learn to make bigger demands from minority group members (Bruner and O'Connor forthcoming, 8). This suggests that when people condition their collaborative behavior on group membership, minorities lose out. Indeed, O'Connor and Bruner find that minority types are incentivized to hunt hare in such a situation, avoiding collaboration. This is consistent with empirical evidence that in many fields, women prefer to collaborate with women (9).

A more complex model incorporating hierarchical interactions produces similarly interesting results. Bruner and O'Connor find that junior academics end up in weaker bargaining positions if certain collaborations are relatively unimportant for senior academics but relatively important for junior academics (15).

Bruner and O'Connor admit to making idealizations in their models. They list the following two:

1. Actors in the games choose strategies concurrently. In real life, actors may choose strategies through mutual discussion. One actor might choose a strategy based on what they know the other has chosen.
2. Their models assume that actors learn strategies that maximize their own benefit. Actual humans may not be this selfish, and even those who are may not be as efficient as the game players are in learning the most self-beneficial strategy. (19)

In addition, in their models incorporating hierarchical interactions, Bruner and O'Connor assume that actors meet their partners randomly (7). This is an idealization because racial and gender groups often cluster in certain subdisciplines. However, they point out that in similar models, they have shown that introducing nonrandom encounters does not substantially affect the results. Bruner and O'Connor's results are robust in other aspects, such as when parameter settings in the games are varied

(Bruner and O'Connor working paper, 20).⁶ However, there is no suggestion of robustness against idealizations 1 and 2 above.

Here is another possible idealization in Bruner and O'Connor's models. Their models do not just assume that actors learn from their peers and from experience, but that they learn only from their peers' behavior *in previous games* and their own experience *in the same types of games*. This is an idealization since, for example, women might learn how to collaborate with men not just from previous attempts at collaboration, but also from other aspects of social interaction, such as how men behave toward them as students and teachers. They may not learn the same bargaining strategies from these other interactions that they learn from the collaboration game. For example, it is possible that from other interactions with men, women learn that they have to make more forceful demands to get what they want. What women learn from these other interactions may counteract lessons that they learn from the collaboration game, where O'Connor and Bruner find that minority members are incentivized to ask for less.

Is the restriction to learning from iterations of the same kind of game an abstraction or idealization? I contend that this depends on the research question of interest. If one is interested in the causal contribution of collaborative interactions to a person's bargaining strategy, then O'Connor and Bruner's results suggest specific mechanisms underlying this contribution. Given this research question, the restriction to learning from the collaboration game is an abstraction, because it ignores details that are irrelevant to the research question. Instead of saying that the model introduces the false assumption that actual agents learn to bargain only from collaborative games, one can reframe the situation as one in which the model *remains silent* on how other sources of learning affect the agent. It omits these other sources without claiming that they do not exist—it merely claims that they do not exist *in the mechanism of interest*, which is true. In a context where we are interested only in a specific causal mechanism, omitting these other sources does not introduce falsehoods, because our research question stipulates that those other sources are not part of our investigation. However, one may have a different research question, about how a range of interactions with majority group members affect bargaining strategies in collaboration. In this case, ignoring other interactions outside the collaboration game may be an idealization, since one is interested in more than just the causal contribution of a specific kind of interaction.

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. 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. 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 deplors.⁷ This does not help Mills's argument, since he would like nonidealized but abstract models to exist. There are simple models, however, that investigate specific causal mechanisms in isolation from others. O'Connor and Bruner's models fall into this class. I contend that we can either:

1. view their models as being abstractions relative to one research question (that of how people learn to bargain from collaborative interactions alone) and idealizations relative to another (that of how people learn to bargain in collaborations from social interactions more generally); or
2. concede that any social-science model omits some causal factors that exist in reality, and therefore all social-science models count as "ideal theory."

Either alternative suggests that the idealization/abstraction distinction Mills uses is unhelpful for his purposes.

I will now describe some lessons from philosophy of science that explain why formal models like Schelling's and Bruner and O'Connor's can be helpful despite containing idealizations.

V. THE BENEFITS OF IDEALIZATION

Limiting idealizations can help distinguish between different mechanisms and different levels of description. For example, thermodynamics is a very useful level of description that depends on the idealizing assumption of infinitely large systems, such as heat reservoirs with infinite heat capacity. It can be understood as a limiting case of statistical mechanics where the number of particles goes to infinity, and the ratio of the number of particles to volume goes to infinity (Styer 2004). Thermodynamics is particularly useful for describing macroscopic systems, and less so for sufficiently small systems.

Another example is that of classical mechanics. Classical mechanics deploys idealizations in different domains to achieve different purposes. In continuum mechanics, the internal properties of materials, such as elasticity and stress, are of modeling interest. Because of the difficulties of modeling such properties at the atomic level, continuum mechanics models materials as continua. This is an idealization, not an abstraction, since we know that materials are composed of particles. But it is an epistemically invaluable idealization for much of engineering. On the other hand, there is rigid-body mechanics, which contains an idealizing assumption that the bodies being modeled are perfectly rigid—that they retain their shape and size regardless of the stress put on them. This is the basic framework for models of seesaws on fulcra, for example, that we see in introductory physics textbooks. The "plank" and fulcrum of the seesaw are modeled as perfectly rigid, when in reality they undergo some distortion under the forces involved. However, with the seesaw, we are often more interested in the rigid-body behavior than in the small distortions the plank must

undergo. Thus, the rigid-body idealization serves a fine epistemic purpose in this context (Wilson 2013). Blurring the sharp distinction that Mills draws between idealization and abstraction, here it is acknowledged that rigid-body mechanics leaves out details that affect the behavior of the material (wood does bend in a seesaw), but these details may not matter in some contexts of interest.

These examples suggest that in many areas of science, idealizations are epistemically essential because they isolate mechanisms of interest for certain modeling purposes. If we wanted to find out how to balance a seesaw, it would not be epistemically useful to include a complete model starting from the atomic level, going up to material properties, and on to how these manifest on the macroscopic scale as an almost-rigid body. Computational impossibility aside, this model would also include a lot of irrelevant detail, whereas a rigid-body model would bypass such details. On the other hand, if we wanted to find out how the fulcrum of a seesaw behaves under pressure, or how a thin plank would bend if used as a seesaw, we would have to use continuum mechanics, and rigid-body mechanics would be irrelevant. Idealizations can thus be seen as capturing different kinds of limiting behavior in science, which are of interest in different contexts.

From another perspective, idealizations can also be seen as isolating different mechanisms at work in a complex system: the macro-mechanism of a rigid body pivoting on a fulcrum is more important, in some contexts, than the micro-mechanism of atoms impinging on one another where the plank and the fulcrum meet.

Extending this lesson to the social sciences is straightforward. Schelling's models emphasize one mechanism of segregation, that which happens in the limiting case of uncoerced individual choices, and are useful for studying this mechanism. Although perhaps no individual choice is entirely uncoerced, they provide a partial account of what goes on at the less-coerced end of the spectrum. The workings of this particular mechanism may be obscured if we include the full messy reality of actually coerced choices, organized action, and so on. It may be more useful to start by investigating each mechanism separately using separately idealized models for each. This is not to say that there is no benefit to then combining these models, but rather that having individual idealized models also serves a purpose, a kind of decomposition of the causes of segregation.

The same might be said of Bruner and O'Connor's models. Bruner and O'Connor investigate one causal contribution to bargaining strategies. Although no actual agent may be influenced by *just* this causal factor, the models illuminate how this particular causal mechanism works. Including all other causal contributions to bargaining strategies may result in a less explanatory model because analytic difficulties may obscure how individual mechanisms work. As Bruner and O'Connor point out, their idealizing assumptions "provide insight into causal processes in these very complex situations that is difficult to obtain otherwise . . . they allow us . . . to list preconditions for these effects to occur that would not have been obvious without these models" (Bruner and O'Connor forthcoming, 19). Furthermore, Bruner and O'Connor think that the models suggest that the social dynamical effects hypothesized are possible, and that they give us reason to watch out for bargaining disadvantages in our interactions (19).

William Wimsatt has a list of possible benefits of idealized models that go beyond the benefits described above. The following are applicable to idealized models in the social sciences:

1. An incorrect simpler model can be used as a reference standard to evaluate causal claims about the effects of variables left out of it but included in more complete models, or in different competing models to determine how these models fare if these variables are left out." (Wimsatt 2007, 104)
2. Two or more false models may be used to define the extremes of a continuum of cases in which the real case is presumed to lie, but for which the realistic intermediate models are too complex to analyze or the information available is too incomplete to guide their construction or to determine a choice between them. In defining these extremes, the 'limiting' models specify a property of which the real case is supposed to have an intermediate value." (104–105)
3. A family of idealized models can be used to test the robustness of hypotheses: hypotheses that are derivable in all these models are robust because they are relatively independent of the differences in assumptions between the models. For nonrobust hypotheses, one can also learn which assumptions they depend on. (105)

Although most of Wimsatt's examples come from biology, his methodological prescriptions are based on the inherent complexity of biological systems and the futility of making complete models of any biological system. Since these properties also hold for systems in the social sciences, one would expect the same epistemic benefits from an appropriate use of idealized models in the social sciences.

VI. SO WHAT'S WRONG WITH IDEAL THEORY?

I have given a two-pronged argument for why Mills's argument against ideal theory fails. Either the abstraction/idealization distinction does not exist independently of modeling purposes, or every social-science model is idealized. Either way, the distinction does not help Mills's cause.

I have also argued that idealizations can be epistemically important. Complex causes underlying social injustice can be decomposed through simplified models that isolate details that are relevant to particular contexts. These models inevitably contain idealizations.

However, I still sympathize with Mills's disapproval of the state of Anglophone political philosophy. I propose some other reasons for why this state is deplorable. Some of these are latent in or implied by Mills's essay, but they are distinct from his explicit argument, which is based on the absolute idealization/abstraction distinction that I criticize.

Before I suggest some reasons to prefer nonideal theory, however, it is worth addressing the proposal that nonideal theory, even if it contains idealizations,

contains them to a lesser extent than ideal theory. This proposal might give us a reason to prefer nonideal theory even if it contains idealizations. However, I think this proposal does not work. Given how idealized formal models of injustice are (and necessarily so, for computational tractability), this could easily lead to models like Schelling's being ranked as more idealized than theories like Rawls. In any case, it's hard to "count" idealizations since some distortions could count as multiple falsehoods (Jones 2005, 183–84). Not to mention questions of measuring *distance from reality*: calculating how far a single distorting factor takes us from reality, and then combining distorting factors that are measured in different units into one homogeneous measure of overall distance. Finally, if we accept my argument that some simple, idealized models have epistemic benefits, then measuring degree of idealization would disfavor these models, whereas I think they are extremely useful for certain purposes. The idealization of the thermodynamic limit, which models systems as having infinitely many particles, seems to be an "infinitely great" idealization, so to speak, but it is very useful.⁸

I now turn my attention to other reasons for preferring models of race- and gender-based oppression, reasons that I think are more promising than trying to measure degree of idealization.

COGNITIVE DIVERSITY

In section V I argued that idealizations in models can be seen as capturing limiting cases in which certain mechanisms obtain a dominance over others. One benefit of having idealized models, then, is the ability to study mechanisms in isolation, figuring out what could be relevant for a particular mechanism without including a multitude of irrelevant parameters.

Given this epistemic benefit, we can say that to the extent that an ideal theory does isolate an actual mechanism in society, it ought to be studied. It is problematic, though, if certain mechanisms are studied to the near-exclusion of others. This is part of the situation that Mills is bemoaning: dominant theories in Anglophone political philosophy tend to exclude common mechanisms of oppression, such as racism and sexism. They discuss hypothetical worlds in which such oppression has been overcome. Yet I do not want to simply deny that these dominant theories are useless, since I think models that tackle the oppression of minorities also contain idealizations, and there might be some contexts in which the theories that Mills critiques are helpful. Perhaps we could combine lessons from conventional political philosophy models with lessons from models that emphasize race- and gender-based oppression. The problem is not the existence of such idealized conventional models, but the fact that they dominate Anglophone political philosophy.

In short, Mills's call for more "nonidealized" models can be reread as a call for more models studying how minorities are oppressed, even if the latter are also idealized. We should, in short, have a diversity of idealizations in the models that we study. This resembles Helen Longino's argument that to promote objectivity, a

scientific community needs a diversity of viewpoints. Since, she argues, there is no way of doing science without using some unproven background assumptions, we ought to have a diversity of such assumptions in the community (Longino 1998, 184–85). Idealizations can play the role of such assumptions.

Given this, the restricted range of theories considered in Anglophone political philosophy could constitute the “ideology” that Mills speaks of. These theories may not be problematic in themselves, but their dominance of the field is.

VALUE-LADEN IDEALIZATIONS

Another possible reason to favor models of oppression is that some idealizations may be ethically preferable to others, even if all models contain idealizations. Mills, for example, suggests that modeling immediate mechanisms of oppression is a better way to reduce social injustice than working out theories like Rawls’s is (Mills 2005, 181).

This reason to favor models of oppression is not based on how unidealized they are—we’ve established that models of oppression often contain idealizations and that it seems impossible to argue that they are somehow less idealized than theories like Rawls’s. Rather, it’s based on emphasizing the value of effective action. This might or might not amount to the same thing as offering the best theory of justice. The best theory of justice may not be the best theory for effective action, just as a “realistic” model of gases may not be the most useful for thermodynamics. Indeed, some commentators read Rawls as acknowledging that nonideal theory is essential for effective short-term action against injustice, although they think that nonideal theory still requires a background ideal theory (Simmons 2010). This view of Rawls can, in my language, be seen as one in which the idealizations of ideal theory better serve long-term goals, while those of nonideal theory serve shorter-term goals. Theorists may differ on how to weight these different goals, and pick their models accordingly. But given my view that many instances of what Mills would call nonideal theory are highly idealized, the difference between ideal theory and nonideal theory would not lie in their closeness to reality, but in the purposes they serve.⁹

Another value-laden way to choose idealizations may be based not on combating social injustice in general, but on more specific goals, such as antiracist or antisexist goals. These more specific purposes may suggest idealizations different from those that Rawls makes, for example. Of course, these purposes may be simply part of the general one of reducing social injustice, but one can imagine researchers weighting certain sub-purposes more or less in their work, and thus preferring certain idealizations more or less than they would if they were simply committed to social injustice in general.

ROBUSTNESS

Another possible reason for preferring some models of oppression over those found in conventional political philosophy could be the presence of analysis proving that the

former are robust. Robustness analysis investigates whether the model's main conclusions continue to hold when we alter certain assumptions. Schelling-type models have passed such robustness tests (Muldoon, Smith, and Weisberg 2012). One might think that much of conventional political philosophy has not passed such tests. It is still unclear if the empirical constraints that Rawls assumes are robust under varying conditions, where such variation may include the addition of known mechanisms of oppression.

One area in which Rawls's theory might not be robust is its treatment of dependent citizens and their caregivers. Eva Feder Kittay argues that Rawls's assumption that "all [citizens] are capable of honoring the principles of justice and of being full participants in social cooperation throughout their lives" (Kittay 1998, chapter 3) is not an innocent idealization, but one that would cause dependents and their caregivers to end up in a worse-off position. In the actual world, dependents and their caregivers exist. This means that Rawls's account of justice fails once we insert real-world conditions into the theory—it is not robust when we vary one of its incorrect assumptions.

Again, however, whether a model passes a robustness test is not a function of how idealized it is. Schelling-type models are extremely simple, glossing over details that might matter for many purposes, but very robust. Indeed, philosophers of science have argued that so-called minimal models, which contain relatively low levels of detail, may be particularly immune to variations in the details, even if they contain idealizations (Batterman and Rice 2014). Schelling-type models can be considered to be minimal models in this sense.

Importantly, the robustness-based justification for models of oppression is a contingent one, based on those models passing robustness tests. Formal models of oppression may fail robustness tests. There might also be Rawlsian models that pass robustness tests of some kind. For example, Rawls's concept of an overlapping consensus may be reinterpreted as a kind of robustness of liberalism in response to changes in basic assumptions (Rawls 2005, 482–83). Rawls's idea is that liberalism is possible because members of a community can have different "comprehensive doctrines" about nonpolitical virtues and values while still agreeing on political principles. Although Rawls does not frame overlapping consensus as an epistemic thesis, it can be interpreted as an epistemic claim about the robustness of liberalism to variations in nonpolitical beliefs. If people can start from different comprehensive ethical doctrines and still converge on the same political principles, this suggests that the latter are somewhat insensitive to errors in the former.

Although I sympathize with Mills's call to theorize about the experience of the oppressed, I do not think this call should rely on a distinction between idealization and abstraction. I argue that either the idealization/abstraction distinction in formal models of injustice is context-relative, or all social-science models contain idealizations. Furthermore, there are epistemic reasons to prefer simple, idealized models. The problem with ideal theory, then, cannot lie merely in the fact that they use idealizations rather than abstractions. I suggest instead that Mills's critique of ideal theory can be reformulated as a critique of the lack of diversity in the idealizations made in political philosophy, the lack of robustness in the conclusions of ideal theories, or

the ethical values that ideal theory emphasizes when it chooses some idealizations over others.

NOTES

I thank Liam Kofi Bright for suggesting the formal models I used, and for other extensive comments. Samantha Hancox-Li, Olufemi Taiwo, two anonymous referees, and the audience at the 2016 Vancouver Summer Philosophy Conference also provided valuable feedback.

1. O'Neill also defines a second sense of idealization that corresponds to Mills's definition of ideal-as-normative. The similarity I'm pointing to lies between Mills's concept of ideal-as-idealized and O'Neill's first sense of idealization as adding too much.

2. This contrasts with the traditional view that social science helps only with implementing directive principles, after moral theory has determined the latter.

3. I thank Eric Swanson for suggesting this.

4. Here, an "individual" means an agent in the model, so it could stand for a household, an organization, and so on.

5. Thus, Schelling also omits processes involving interactions between segregation by color and segregation by income.

6. One example of a parameter setting is the specific values of payoffs in the payoff tables.

7. To maintain analytic tractability, all social-science models exclude some real-world causal factors.

8. A more restricted possibility for measuring degree of idealization is one that indexes degree of idealization relative to a modeling purpose. One might say, for example, that the ideal gas model is relatively unidealized for the purposes of modeling the properties of certain gases, but that it is highly idealized when we want to model other gases. In this case, degree of idealization could perhaps be given a coherent meaning as "accuracy of prediction" for a certain purpose. If our purposes are explanatory rather than predictive, perhaps we could also introduce some measure of explanatoriness. I do not want to rule out this possibility of defining degree of idealization relative to a purpose. If we use this measure to decide which models to prefer, however, then we need some way of deciding which purposes are to be favored, since even simple models like Schelling's can serve many purposes. I discuss the possibility of preferring models based on an ethical evaluation of the purposes they serve in the second section of part VI.

9. A notable disanalogy with the physical-science case is that idealizations in social-science models may be more pernicious because we can be socially primed to forget them in accordance with structural biases.

REFERENCES

Batterman, Robert W., and Colin C. Rice. 2014. Minimal model explanations. *Philosophy of Science* 81 (3): 349–76.

- Bruner, Justin, and Cailin O'Connor. Forthcoming. Power, bargaining, and collaboration. In *Scientific collaboration and collective knowledge*, ed. Thomas Boyer, Conor Mayo-Wilson, and Michael Weisberg. Oxford: Oxford University Press. <http://philpapers.org/archive/BRUPBA-2.pdf> (accessed November 4, 2016).
- Bruner, Justin, and Cailin O'Connor. Working paper. Dynamics and diversity in epistemic communities. https://www.academia.edu/9796362/Dynamics_and_Diversity_in_Epistemic_Communities_Working_Paper (accessed November 4, 2016).
- Jaggar, Alison M. 1988. *Feminist politics and human nature*. Totowa, N.J.: Rowman & Littlefield Publishers.
- Jones, Martin R. 2005. Idealization and abstraction: A framework. *Poznan Studies in the Philosophy of the Sciences and the Humanities* 86 (1): 173–218.
- Kittay, Eva Feder. 1998. *Love's labor: Essays on women, equality, and dependency*. New York: Routledge.
- Longino, Helen. 1998. Values and objectivity. In *Philosophy of science: The central issues*, ed. Martin Curd and J. A. Cover. New York: W. W. Norton & Company.
- Mills, Charles W. 2005. "Ideal theory" as ideology. *Hypatia* 20 (3): 165–83.
- Muldoon, Ryan, Tony Smith, and Michael Weisberg. 2012. Segregation that no one seeks. *Philosophy of Science* 79 (1): 38–62.
- O'Neill, Onora. 1987. Abstraction, idealization and ideology in ethics. *Royal Institute of Philosophy Supplements* 22: 55–69.
- . 1990. Justice, gender and international boundaries. *British Journal of Political Science* 20 (4): 439–59.
- Rawls, John. 1999. *A theory of justice*, Revised ed. Cambridge, Mass.: Belknap Press.
- . 2005. *Political liberalism*. 2nd ed. New York: Columbia University Press.
- Schelling, Thomas C. 1971. Dynamic models of segregation. *Journal of Mathematical Sociology* 1 (2): 143–86.
- Schwartzman, Lisa H. 2006a. Abstraction, idealization, and oppression. *Metaphilosophy* 37 (5): 565–88.
- . 2006b. *Challenging liberalism: Feminism as political critique*. University Park: Pennsylvania State University Press.
- Simmons, A. John. 2010. Ideal and nonideal theory. *Philosophy & Public Affairs* 38 (1): 5–36.
- Styer, Daniel F. 2004. What good is the thermodynamic limit? *American Journal of Physics* 72 (1): 25–29.
- Valentini, Laura. 2012. Ideal vs. nonideal theory: A conceptual map. *Philosophy Compass* 7 (9): 654–64.
- Wiens, David. 2015. Against ideal guidance. *Journal of Politics* 77 (2): 433–46.
- Wilson, Mark. 2013. What is "classical mechanics" anyway? In *The Oxford handbook of philosophy of physics*, ed. Robert Batterman. Oxford: Oxford University Press.
- Wimsatt, William C. 2007. *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Cambridge, Mass.: Harvard University Press.