Hypatia (2026), 1–22
doi:10.1017/hyp.2026.10060

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Alienation, Idealization, and Modeling

Catharine Saint-Croix[1,2]

[1]Philosophy, University of Minnesota Twin Cities, Minneapolis, USA and [2]African Centre for Epistemology and Philosophy of Science (ACEPS), University of Johannesburg, Johannesburg, South Africa
Email: cat.saintx@gmail.com

(Received 23 July 2025; revised 8 December 2025; accepted 11 December 2025)

Abstract

Models allow us to expose, explore, and excavate our assumptions. This is especially true for the sorts of assumptions that concern scholars of marginalization and resistance, who aim to show how patriarchy, colonialism, and related structures shape our lives. Yet within this scholarship, there is a recurrent skepticism toward practices of modeling, particularly formal or idealizing practices. These critiques portray modeling as anything from necessarily sexist to perniciously universalizing to oppressive. While these claims are mistaken, they nevertheless share a crucial and underappreciated undercurrent: alienation. I argue that alienation provides a more fundamental and more precise framework for evaluating modeling practices. This reframing clarifies the normative stakes of exclusionary modeling while preserving formal techniques for the projects that they are especially well-positioned to promote: exposing entrenched assumptions, resisting "obfuscatory accommodation," and facilitating engagement across epistemic and political divides.

Exploring the eastern foothills of the Rocky Mountains, Lewis and Clark lost their way. This was not because they lacked a map. Rather, as Binnema (2001) explains, Lewis and Clark had the benefit of quite generous help provided to their forerunners: members of the Blackfoot people created hand-drawn maps for them, such as Fig. 1. But Lewis and Clark found themselves confused. They, along with the British cartographers who copied the Blackfoot map and attempted to incorporate it into their own, could not understand it because they failed to realize that it used the *mistakis* as its central orienting feature.

This term, *mistakis*, refers to the horizon line or "spine" of the Rockies when viewed from the eastern side (Fig. 1). This was radically different from Lewis and Clark's expectations. They were expecting that any map would use the continental divide, which demarcates the east–west watershed of the Americas, as its central orienting feature. The continental divide, however, is difficult to discern from the plains, and therefore much less useful for navigation than the *mistakis*. So it was not the central orienting feature of a map made by people who live on the plains. As Binnema (2001, 212) puts it, "Blackfoot

© The Author(s), 2026. Published by Cambridge University Press on behalf of Hypatia Inc. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

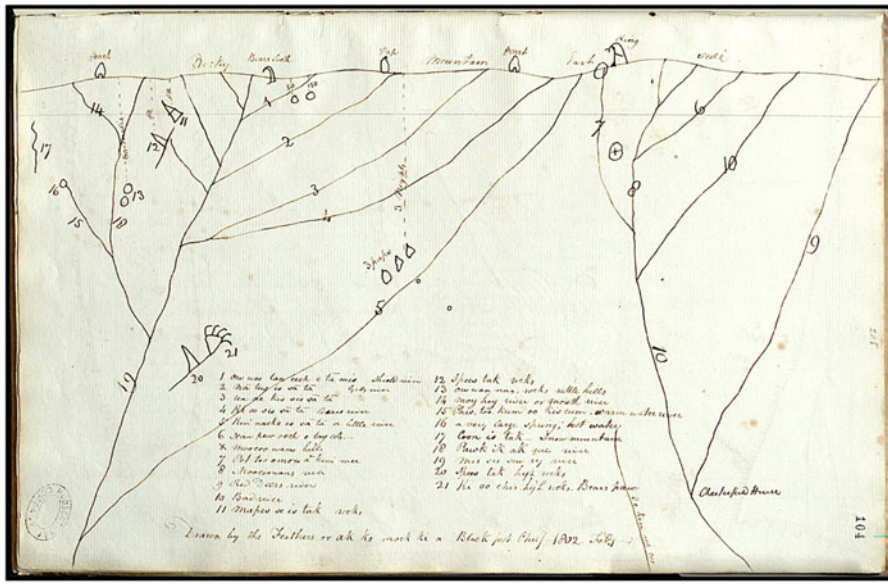


Figure 1. Peter Fidler's copy of Old Swan's 1802 map. The mistakis is the line along the top. Image from Hudson's Bay Company Archives, Provincial Archives of Manitoba, E.3/2 fo. 104.

maps were designed by a plains people and they truly represent a view … from the plains."[1]

Cartography is a species of modeling. As models, maps embed perspective, assumptions, and context into a simplified, symbolic representation intended to prioritize, communicate, and facilitate certain ideas and inferences. More generally, modeling generates formal objects (sometimes concrete, sometimes abstract) that can embody or expose the gaps between our own understandings and those of others.[2] In the Blackfoot/Explorer case, for example, these gaps are a matter of expectations and literal perspective. And, as we will see, the articulation of these gaps can be alienating or enlightening (perhaps both), depending on the context. But these assumptions aren't necessarily built in intentionally; rather, they become visible *as assumptions* because using the model—the map, in this case—requires a host of expectations and understandings. In their absence, the model is unrecognizable, inscrutable. You get lost in the Rockies. In this way, and more generally, models are a valuable tool for exposing, exploring, and excavating assumptions.

This is especially true for the sorts of assumptions that worry political philosophers, especially scholars of marginalization and resistance who seek to expose the ways that patriarchy, colonialist perspectives, and the like pervade our lives. But, within this scholarship, there is a recurrent theme of rejecting practices of modeling as antithetical to its very goals. This rejection spans not only logic and mathematical modeling but also the less formal tools of abstraction and idealization common across academia.

[1]Cited in Lowan-Trudeau (2021).

[2]On the other hand, some models can also help us realize that our understandings are surprisingly similar to others' because they share structural features. Thank you to Liam Kofi Bright for pressing this.

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

According to these analyses, formal modeling can be anything from necessarily sexist (Nye 1990) to perniciously universalizing (Crary 2018; Harding and Hintikka 1983; Nye 1990; Walkerdine 1988) to oppressive (Mills 2005; O'Neill 1987; Nye 1990; Walkerdine 1988).

These criticisms, I argue, share a crucial and underappreciated undercurrent: alienation. Drawing out this undercurrent is the focus of this paper. I argue that it is alienation, not formal modeling itself, that is at the root of these criticisms. Where critics like Nye (1990), Mills (2005), and O'Neill (1987) present their objections in very different ways (Nye objects to logic itself as an inherently sexist enterprise, while Mills and O'Neill concentrate on the ideal/nonideal distinction), focusing on the sense of alienation underlying these approaches draws our attention to a more fundamental and precise critique. This approach not only evades standard objections to Nye, Mills, and O'Neill, but also furnishes a normative framework through which to assess not only when modeling practices may cause harm, but also when they can support emancipatory aims. This leads back to the dynamic of exposing, exploring, and excavating assumptions, all tasks with special relevance for scholars of marginalization and resistance.

My argument begins with a deeper look at the power of modeling for pursuing these aims in section 1. Section 2 examines objections to formal modeling, focusing on the role of idealization and abstraction as they appear in Mills (2005), O'Neill (1987), and Nye (1990). Section 3 argues that the undercurrent of alienation running through Nye (1990) also runs through the objections to formal modeling discussed in section 2. In light of this, section 4 argues that it is alienation, rather than modeling itself, that should be central to our evaluation of particular modeling practices. Section 5 turns to the value of modeling. I argue that formal modeling is useful for three important projects: (1) exposing both our own and others' assumptions, (2) avoiding what I call "obfuscatory accommodation," and (3) facilitating engagement across epistemic and political divides. The core claim of this paper is that alienation—not formalism, abstraction, logic, or idealization *per se*—is the underlying concern in many critiques of formal modeling, and that recognizing this enables a more fundamental and more precise account of when and why modeling can be harmful. This matters because modeling has valuable work to do, especially for those of us inclined toward these very critiques.

1. Modeling as resistance

In 2014, The Firelight Group began the Indigenous Mapping Workshop, a conference providing indigenous peoples with access to tools and training that facilitate mapping their lands using modern technology. More recently, this project has given rise to the Indigenous Mapping Collective, a virtual space for sharing knowledge. Similar projects include "mapping back" (The MappingBack Network 2022) and "counter-mapping" (Hunt and Stevenson 2017). These projects demonstrate that, although cartography has been a tool of colonialist oppression, it can also be a tool of reclamation, legitimization, and cultural communication. Projects like these do not eschew the formal, abstract tools of mapmaking—they embrace them. Some engage in straightforward decolonial interventions, such as replacing or superimposing indigenous names on Western maps. Others create new maps altogether, setting aside or revising Western cartographic practices in the process. Others draw together indigenous and mainstream practices, with the intention of bringing to light the assumptions and tensions of each.[3]

[3]See Lowan-Trudeau (2021) for an overview of these projects.

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

The revelatory capacity of such tools is not unique to intentional anti-colonialist interventions. Focusing on the role of logic in shaping our reasoning, Plumwood (1993) argues that classical logic reinforces oppressive reasoning, especially around gender, and develops a non-classical relevance logic in its stead. Similarly, Kosten (2025) devises an intuitionistic logic for reasoning about gender identities that might appear inconsistent or incoherent in classical logic. Focusing on the nature of logic itself, Russell (2024) and Saint-Croix and Cook (2024) draw on recent work in philosophy of science to argue that the project of feminist logic is an *anti-exceptionalist* project—this is the view that logic and mathematics are, like physics and biology, studies of the empirical, natural world rather than rarefied exceptions. Within standpoint epistemology, Wu (2022) uses agent-based modeling to demonstrate the consequences of ignoring testimony from marginalized groups, while Saint-Croix (2020) uses modal logic to model how standpoint occupants respond to evidence.[4] In a similar vein, Bowman and Cook (forthcoming) use formal logic to model "knowing-what-it's-like"—a crucial concept for much of feminist standpoint theory.[5] These projects help to vindicate the claims of standpoint epistemology by embodying them in a formal structure. Additionally, such formal structures may allow us to make new observations or find otherwise obscure conflicts among our assumptions.

Where examples and applications are called for, I will focus on cases from feminist formal logic. Formal logic is, after all, a clear example of formal modeling. Importantly, however, feminist formal logic occupies one end of what is a spectrum: modeling may be more or less formal. While logic, like cartography, is quite formal, I take the sort of narrative modeling characteristic of much of political philosophy (especially, e.g., Rawls (1999), of which more in the next section) to be within the remit of this discussion as well.

I understand modeling in terms similar to those Giere (2010) sets out: there are no necessary and sufficient conditions for what counts as a model. Rather, modeling is a practice and models are defined relationally within that practice. On Giere's view, "Agents (1) intend; (2) to use model, M; (3) to represent a part of the world, W; (4) for some purpose, P," (2010, 269). A pencil may be a model for particle physics just as well as a set theoretical construction can be a model for decision-making. What matters is agents' *intentional use* of what they designate as the model, *whatever it may be*, to represent the world for some purpose.[6]

2. Ideal theory's discontents

Anxiety over formal modeling has taken many forms, but the most influential is Charles Mills's (2005) and Onora O'Neill's (1987) narrower criticism of ideal theory. Both

[4]For discussion of standpoint epistemology in general, see Allen and Howard (1997) and, more recently, Wylie (2013) and Toole (2021).

[5]Cook and Yap (forthcoming) gathers such projects together under the heading of feminist formal logic.

[6]This pluralistic approach is further buttressed by the inclusive conclusions of this paper and the general direction of the literature on modeling. As Callender and Cohen (2006) and Frigg and Nguyen (2017) argue, necessary and sufficient conditions for what constitutes a model are unlikely to be unearthed. Different contexts may call for radically different degrees of formality, abstraction, and idealization and radically different implementations. Nevertheless, the there is an extensive literature on the topic of models and modeling; see Frigg and Hartmann (2024), for an overview. Notably, on this view, Rawls's (1999) work Mills's (2005) criticism of it is squarely within the scope of modeling. This will be important as we turn to section 2.

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

O'Neill and Mills take themselves to be rescuing modeling—to some extent—from a slew of critical political theorists, such as Burke, Bentham, Hegel, and Marx (among many others), who argued that political theories formed on the basis of abstract reasoning fail because, for example, they "assume capacities for reasoning and choosing which human agents simply lack" (O'Neill 1987, 56). In the wake of such criticisms, there was growing suspicion of any abstraction in political theory, ethics, and related fields.

O'Neill and Mills push back against the idea that avoiding abstraction entirely is even possible, let alone worthwhile. Instead, they argue, when we take a more discerning eye to the pernicious practices in which formal theorists sometimes engage, we see that it is actually the distinct category of "ideal theory" that is problematic.[7] Both of them worry that ideal theory disconnects itself from the real world because the assumptions are too far beyond the actual capacities of human agents to be relevant or revelatory. Mills puts an even finer point on this, arguing that focusing on these idealized models all but ensures that ideal theory will perpetuate, rather than combat, oppressive structures. He writes:

> mainstream political theory has not, until very recently, thought about and taken seriously what would be necessary to achieve genuine racial and gender equality. I suggest that this is a perfect complement, in the more empirical realm of political science, to the abstraction in the more rarefied realm of ethics and political philosophy. In both cases, an idealized model is being represented as capturing the actual reality, and in both cases this misrepresentation has been disastrous for an adequate understanding of the real structures of oppression and exclusion that characterize the social and political order. The opting for "ideal" theory has served to rationalize the status quo. (Mills 2005, 181)

What, then, is meant by "ideal theory"? Mills, following O'Neill, defines ideal theory as theory that relies on idealization "to the exclusion, or at least marginalization, of the actual" (Mills 2005, 168). Here, Mills draws a distinction between two forms of modeling: idealization-as-descriptive model (which is what O'Neill calls "abstraction") and idealization-as-ideal model (or simply, "idealization"). Elaborating on this distinction, Mills draws a tight analogy with modeling in classical mechanics:

> **Ideal-as-descriptive-model.** Models that "represent the thing as it is, as best as possible," (Mills 2005, 166–67). Such models purport to be descriptive of the crucial aspects and causal or dynamical structure of the target phenomenon. In this case, the assumptions one makes are simplifying assumptions, with the aim being to produce "a schematized picture of the actual workings and actual nature of P" (Mills 2005, 166–67).

[7]It is worth noting that the ideal/nonideal distinction has been deployed in several different ways, as catalogued by Valentini (2012). For Mills and O'Neill, the ideal/nonideal distinction falls under Valentini's second category, for which "ideal" theory is less realistic and "nonideal" theory is more realistic. Other uses include the full compliance vs partial compliance distinction and the end-state vs transitional distinction. With Mills and O'Neill, I will focus on the less realistic vs more realistic understanding of the distinction.

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

Ideal-as-idealized-model. Models that represent an exemplar of P. These are models that show how the target phenomenon *should* work, where "should" might involve any relevant normative domain.

Examples of this latter kind include ideal gases, perfect vacuums, and frictionless planes. And also, of course, a Rawlsian theory of justice (Rawls 1999). Similarly, in her enumeration of the sorts of pernicious idealizations that mark ideal theory, O'Neill (1987, 56) includes the slew of assumptions expected utility theory employs when characterizing agents: interpersonally comparable and universally commensurable utilities, perfect quantitative capabilities, transparent self-knowledge, etc. For Mills and O'Neill, then, this distinction drawn in analogy with classical mechanics maps onto the ideal/nonideal distinction in political and moral theory: Ideal theory is grounded in ideal-as-idealized models, whereas nonideal theory instead avails itself of ideal-as-descriptive models.

Recently, Hancox-Li (2017) and Táíwò (2023) have offered valuable criticisms of this distinction. Focusing on the social sciences, Hancox-Li argues that the sorts of idealizations that Mills and O'Neill point to are either present in every model, because all social scientific models involve idealization to at least some extent, or the notion of an idealization itself must be context-sensitive, so that what counts as an ideal-as-idealized model rather than an ideal-as-descriptive model depends on the research question at hand. Táíwò argues that the issues pinned upon the shoulders of ideal theory are better understood as issues about model-to-world inferences. On this view, Mills's focus on the ideal construction itself is misplaced—instead, we ought to worry about the uses of ideal constructions to infer features, facts, and predictions about the real world. These criticisms are well-placed and cast significant doubt on the viability of the ideal/nonideal distinction, at least as proxy for the ideal-as-idealized/ideal-as-descriptive distinction. Nevertheless, they do not put to rest the worries driving the originators of that distinction. It was developed in response to *something,* and the argument that follows is an effort to locate this common kernel.

2.1. The anxiety of alienation

For the moment, however, we take a brief detour into another critical voice: Andrea Nye. In her landmark *Words of Power*, Nye (1990) argues that formal modeling in all of its guises—from Aristotelian syllogisms to economics to the logic of racism—is inimical to feminist purposes. Within analytic feminism, Nye's arguments were not well-received.[8]

But her thorough dismissal has been too hasty. My purpose in drawing our attention to Nye is to shed light on an underappreciated element of her critique of formal logic. Even in its most inflammatory moments, *Words of Power* offers the clearest expression of the undercurrent of anxiety running through many critical discussions of modeling. That anxiety is alienation.

[8]For example, feminist logician Valerie Plumwood (1993, 437–39) argues that Nye's criticisms are woefully constrictive and also misplaced, being better targeted at reason than logic, MacPherson's (1999) "Three Misrepresentations of Logic" focuses largely, though not exclusively, on Nye (1990), and Weiner (1994) closes her review of *Words of Power* by observing, "accuracy is one of the standards by which Nye has asked to be judged. This is a standard that *Words of Power* conspicuously fails to meet. Nye's characterizations—both of the logic and of the man—are products, not of scholarship, but of fantasy," (p. 681).

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

The opening pages of *Words of Power* recall Nye's own struggle in her introductory logic course. She wonders, "Is it because I, as a woman, had a different kind of mind, incapable of abstraction and therefore of theorizing? Is it because I was 'too emotional'?" (Nye 1990, 2). Trying to comprehend her experience, Nye turned to histories of logic. Summarizing that study, she writes:

> The successes and failures described [by the historians] were only formal: detached from human concerns and as inscrutable and independent of any sensible content as Quine's examples. The problems of logic, historians of logic seemed to insist, are logic's own; they are unrelated to conflictual relations between men and women, between men, or between men and the natural world. Unrelated, that is, to all that might have made them understandable. (1990, 3)

Though Nye does not speak in terms of alienation, it's quite clear from this description that her experience of logic and formal theory is nothing if not one of alienation: The formal tools she is provided make her feel estranged from the ideas they're meant to illuminate.

In her conclusion, Nye inverts this theme, arguing that it is the practitioners of formal logic themselves who are alienated:

> Desperate, lonely, cut off from the human community which in many cases has ceased to exist, under the sentence of violent death, wracked by desires for intimacy that they do not know how to fulfill, at the same time tormented by the presence of women, men turn to logic. Doomed to fail in their Parmenidean flight out of the world, fated tragically never to realize their desire for permanence and purity, can they be condemned, or only pitied, or even admired for their nobility? (1990, 175)

In the paragraph that follows, Nye—perhaps unsurprisingly—relates having found reason for neither pity nor admiration. It is easy to dismiss this passage, and, perhaps, the text along with it. But this would be to dismiss the theme. Here at the end of the text, just as in the beginning, we find the primary anxiety driving the book's critique: a fundamental sense of alienation. The fear that logic and formalization detach us from what is human. The fear that formalization is some alienation brought down upon us by Men, The West, or whatever target tyrant one prefers. This is the central anxiety of Nye's text, whether in the form of confession or accusation.

3. Alienation in abstraction

This same theme of alienation runs through Mills (2005) and O'Neill (1987). For the sake of space, I focus on Mills here. After giving his explanation of ideal theory and listing some of the characteristic features of ideal models—that they idealize social ontology, individual capacities, that they abstract away from oppression, and so on—Mills then entreats readers:

> Now look at this list, and try to see it with the eyes of somebody coming to formal academic ethical theory and political philosophy for the first time. Forget, in other words, all the articles and monographs and introductory texts you have read over the years that may have socialized you into thinking this is how normative theory

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

> should be done. Perform an operation of Brechtian defamiliarization, estrangement, on your cognition. Wouldn't your spontaneous reaction be: How in God's name could anybody think that this is the appropriate way to do ethics? (Mills 2005, 169)

Notably, the term Mills uses here, "Brechtian defamiliarization" comes from Bertolt Brecht's 1936 essay "Alienation effects in Chinese acting," in which Brecht describes the way that certain theatrical techniques act on the audience. Brecht describes this effect as preventing members of the audience from identifying themselves with the characters in the play (Brecht 1977, 91).

Similarly, Mills's thought experiment here, speaking to political philosophers and ethicists as a whole, attempts to show us that if we are able to set aside our acculturation to the habits of philosophy, we will see that ideal theory also has this alienating effect. That is, the techniques of ideal theory prevent people—or, at least, anyone who is part of a marginalized social group—from identifying with the characters, so to speak, in the theory. Underscoring this, Mills writes: "It is no accident that historically subordinated groups have always been deeply skeptical of ideal theory, generally see its glittering ideals as remote and unhelpful, and are attracted to nonideal theory" (2005, 170). Here, Mills points to the ideal/nonideal distinction as a boundary between skepticism and attraction. As Hancox-Li (2017) and Táíwò (2023) point out, however, there is good reason to worry about the relevance—and even viability—of this distinction. Taking it in full generality threatens to render all such theorizing as ideal theory. But these critiques of the ideal/nonideal distinction do not cut against the sense of alienation that Mills emphasizes.

Moreover, focusing on alienation rather than idealization *per se* illuminates how fragile the ideal/nonideal distinction really is. Consider, for example, the critique of the Combahee River Collective. In their Statement of 1977, they argued that considering oppression only along the lines of race or gender, and ignoring the *sui generis* oppressions arising at the intersection of race, sexuality, class, and gender, imperils the capacity to actually address those oppressions. They write:

> The most general statement of our politics at the present time would be that we are actively committed to struggling against racial, sexual, heterosexual, and class oppression, and see as our particular task the development of integrated analysis and practice based upon the fact that the major systems of oppression are interlocking. The synthesis of these oppressions creates the conditions of our lives. (Combahee River Collective 1979, 362)

For the Collective, in other words, a purportedly nonideal model that includes oppression, and even includes separate axes of oppression, can remain alienating. It remains comparatively ideal because they cannot see themselves in the model. They are alienated from it. In this case, trying to parse the ideal/nonideal distinction sets up murky, shifting boundaries with little analytical value. The lens of alienation, by contrast, brings the experience the Collective describes into focus. The question becomes: Who is able to see themselves in the model, given the question at hand? This shift in focus prompts a reorientation, to which we now turn.

4. Reorientation

We have seen that Mills's and O'Neill's strategy is built around the strangest feature of modeling in moral and political theory: rampant, distorting idealization. They focus on

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

this feature in working to pinpoint the source of the problems they see in these corners of philosophy. It is trying to make this concern analytically tractable that gives rise to the tight analogy with modeling in classical mechanics. This, in turn, leads to the centrality of the ideal/nonideal distinction. As we have seen, however, this distinction is not robust (Hancox-Li 2017; Táíwò 2023). Moreover, as a tool for evaluating the normative adequacy of models, the ideal/nonideal distinction is unsuccessful because it facilitates only an indirect and sometimes inadequate analysis. As the Combahee River Collective's statement shows, idealization itself is a mere proxy for the sorts of harms that such critics of modeling practices identify.

In contrast, as I'll argue in this section, alienation provides a more targeted and more fundamental framework through which to understand these issues. Focusing on alienation reframes the problem in terms of the disconnect between models and the agents who create, use, or are represented by them. This reframing brings to the fore questions about modeling practices in their contexts and in terms of their actual (or plausible) effects: Do they misrepresent, silence, or otherwise alienate those they are meant to serve? And who ought such modeling practices serve to begin with? The ready availability of questions like these makes alienation a better lens through which to understand both the criticisms of modeling and the conditions under which it can be harmful.

Importantly, this shift in focus does not entail rejecting Mills's central insight in "'Ideal theory" as ideology' (Mills 2005). Mills is right to observe that there is something akin to ideology operating in the way that idealization is sometimes deployed, without awareness of or concern for its capacity to distort and thereby cause harm. But we do not need the distinction between ideal and nonideal theory to underwrite that claim. To the contrary, this claim is equally well supported by observing that the alienation felt by Mills—along with the women, people of color, poor, and working-class people he identifies as slighted through overzealous adherence to maximally idealized theorizing (Mills 2005, 172)—goes unnoticed by many who propose, build, and attempt to apply these models. Even as we shift to alienation as an analytical framework, the ideology claim still matters because such blindness allows alienation to flourish. And it is in such sites of alienation that we find apt targets for more fundamental, more precise ethical analysis than the ideal/nonideal distinction affords. So, when a modeling practice creates or reinforces a sense of estrangement, when the modeled cannot recognize themselves in the model, or when users are unable to engage it meaningfully, these are the places to focus ethical analysis.

Shifting our focus to alienation, therefore, reorients the discussion. Instead of asking whether a model is overly idealized or too formal, we may ask: Who is likely to be alienated by this model, and is that alienation justifiable given the model's purpose? These questions can be asked across the full range of modeling practices, from Rawlsian narrative constructions to technical agent-based simulations. Answering them requires coming to understand the normative weight of alienation. This perspective also furnishes tools by which we might distinguish between harmful and productive forms of alienation (e.g., Mills's proposal to engage in Brechtian defamiliarization)—a distinction the ideal/nonideal binary is poorly equipped to capture.

All of this, however, depends on the nature of alienation itself, to which we now turn.

4.1. Alienation

As Gilabert (2020, 52) writes, alienation has "fallen out of fashion" as an analytical framework. But it remains a useful lens. Gilabert provides the following basic definition:

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

> At a minimum, "alienation" is used to refer to a state or process of separation or division—of lack of unity, harmony, or connectedness. This disunity occurs between a subject S and some object O in certain circumstances C. S is usually taken to be a person or an agent capable of self-knowledge and self-assessment. O, in turn, may be aspects of the natural world, other people, or the agent themselves. C, finally, may range over various material and social background factors. (2020, 53)

For our purposes, the object in question is a particular model or practice of modeling. The subject varies. Mills (2005), O'Neill (1987), and Nye (1990) are clearly alienated subjects themselves, but they also strive to speak for others: those purportedly, but not genuinely represented in the model, as well as those ignored and not even purportedly represented. They, too, are alienated subjects. The circumstances are the circumstances of use for the model or practice in question.

Alienation is also typically understood as having a normative dimension: the separation in question is a *problematic* separation.[9] This is because many instances of separation are perfectly acceptable. For example, trivial disinterestedness—for example, being indifferent to the finer points of seventeenth-century caseiculture—is a form of separation, but does not seem to induce anything that ought to be understood as alienation (except perhaps under very specific circumstances). As Leopold (2022, section. 1.2), puts it, "candidate separations have to frustrate or conflict with the proper harmony or connectedness between that subject and object." For present purposes, then, the "proper harmony" at issue will be the appropriate fit between modelers, models, and what is modeled—fit that reflects both the purpose of the model and the social and epistemic positions of those involved.[10]

With this understanding of alienation in mind, recall the practice-focused account of modeling, discussed above. On this view, modeling is a practice carried out by agents who intend to use a model to represent part of the world for a particular purpose. Questions of alienation, then, concern the relationship among these four elements—agent, model, target, and purpose. This understanding of modeling is useful because it illuminates the many junctures at which alienation might creep in. Alienation might be generated through a failure to understand the model itself, but it might also arise as a consequence of having different intentions from the model's creator, trying to use the model for different purposes, or, perhaps most straightforwardly, being a purportedly modeled subject of the model and failing to identify with the "characters in the play", to recall the Brechtian conception.[11] Alienation can also come in many degrees: being aligned in all but one of these elements is a way to experience alienation, but so is

[9]Notably, however, Gilabert (2020) is an exception to this.

[10]There is much more to be said about the nature of alienation, but a full analysis lies beyond the scope of this paper. My aim here is not to settle debates about the nature of alienation, but to show how this concept can serve as a fruitful framework for examining the problems that modeling introduces. Future work might explore how the account offered here intersects with other recent developments, such as the more specific accounts of alienation offered by Bousquet (2023), Wiersma (2026), and Jaeggi (2014). Later in the paper, I will rely on Jaeggi (2014) to fill in the details of how proper harmony and problematic separation might be understood, but other accounts will provide valuable analyses as well.

[11]Alienation may also be understood in a subjective or objective sense. The subjective sense is a matter of the agent's feeling the disunity or disharmony at issue, while the objective sense is a matter of there being a disharmony understood without essential reference to individuals' feelings or experiences. See Leopold (2022, sec. 3), Hardimon (1992), and Gilabert (2020) for further discussion. For the sake of space, I omit

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

misalignment between all of them, or any combination in-between. These disharmonies can also be generated or exacerbated by the circumstances in which the agent encounters the model or practice. To see this more clearly, we'll consider three forms of alienation that may arise in relation to formal modeling practices: alienation as inscrutability, as dissonance, and as exclusion.

We are already familiar with examples of *alienation as inscrutability*: the Lewis and Clark case with which we began and Nye's description of how illegible and impenetrable she found her introductory logic course. Instances of alienation as inscrutability are failures of "proper fit" between the formal object and the individual attempting to understand it. So, they often concern being unable to understand the model itself or the intentions about how it is meant to represent the world. As with each of the forms we will examine, whether a particular instance of this form *in fact* constitutes alienation depends on the particular account of alienation one adopts, as that account will serve to fill in the details of "proper fit."

In Nye's case, she neither understands the formalism itself nor how she is meant to use it. This latter part is evident in her exasperation around what she sees as archaic rules about which aspects of the situation matter to the representation.[12] In combination with the circumstances—the pressure of needing good grades in order to pass the logic course required for her philosophy degree—there is no room for generating harmony or connectedness in Nye's experience. The feeling of shifting from inscrutability to clarity can be powerfully rewarding, but likely not against the background conditions of a threat to one's felt purpose in life. Against the background of, for example, Jaeggi's (2014) account, Nye's experience will clearly constitute an instance of alienation. This is because Jaeggi argues that the problematic separation of alienation arises from disruption of one's capacity for "self-determination and being the author of one's own life" (2014, 39).[13] For Nye, running up against the wall of logical formalism generated exactly this sort difficulty: it was a disharmony that, given the role of logic in acquiring the credential necessary in achieving her goals, undermined her capacity to author her own life.[14]

The ease with which formal modeling generates inscrutability is also an important reason that alienation is such a common undercurrent in criticism of modeling. Formal

detailed discussion of this distinction except to highlight that both forms are relevant to the harms (and benefits) below.

[12]For example, characterizing her reaction to a homework problem in Quine's *Methods of logic*, Nye writes in third person, "'Jones away.' 'Smith not ill': the phrases jangle in her mind. Nonsense syllables. Was Jones often ill? Not ill today for once? And why? And Smith so often gone. Where?" (1990, p. 1).

[13]Thus, an "improper fit" between a modeling practice and an individual will be a relationship that impedes this ability.

[14]Taking on board Shiffrin's (2012, p. 386) view of harm, on which harms "significantly impede one's capacity to achieve substantial congruity between one's will and one's life," Jaeggi's understanding of alienation entails that alienation is a harm to the alienated subject. Bradley (2012, p. 400), interprets Shiffrin as arguing that "an event harms someone iff it causes him to be in a state that he rationally wills not to be in." Here, too, restriction of alienation to problematic separation will render the experience of alienation harmful. Similarly, on a wellbeing-based account like the one offered by Purves (2019), alienation will—even on a more permissive account than Jaeggi's—constitute a harm, though perhaps not a major harm. Notably, a wellbeing-based account of harm makes harm fairly widespread. This underscores the importance of using a first-order moral theory to adjudicate individual cases, as described in section 4.1. One might prefer an analysis of harm on which only reductions in wellbeing over a certain threshold constitute harms. Even on such an account, the discomfort brought about by alienation marks an important site for ethical scrutiny because such analysis will be necessary to determine whether it does, indeed, constitute harm.

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

methods often require specialized training, which requires time and resources. Where those resources are unequally distributed in a way that marginalizes people, those marginalized people must work harder to acquire the training needed to avoid this form of alienation. Thus formal methods not only easily alienate, but are also easily bent toward alienation—toward creating work that is inaccessible to and exclusive of marginalized people and their circumstances.[15]

Alienation as dissonance is not a matter of failing to understand. Rather, it is a matter of discord between one's own views and the particular practice of modeling in question. Mills's (2005) imploring us to engage in Brechtian defamiliarization, quoted above, illustrates this experience of alienation as dissonance. The subject in this case—ourselves after sloughing off academic habits—is perfectly capable of understanding the model. What Mills wants us to see is that this model is disconnected and disharmonious with our untutored ethical reasoning. He writes:

> When we start from what is presumably the uncontroversial premise that the ultimate point of ethics is to guide our actions and make ourselves better people and the world a better place, then [ideal theory] will not only be unhelpful, but will in certain respects be deeply antithetical to the proper goal of theoretical ethics as an enterprise. (Mills 2005, 170)

Fragility of the ideal/nonideal distinction notwithstanding, "ideal theory" is meant to be a model for the purpose of ethical reasoning, but in Mills's view, it is ill-suited to that purpose. That dissonance arises not only from Mills himself, but also the circumstances—the actual, deeply nonideal world in which the model is deployed.

Finally, *alienation as exclusion*. This form of alienation concerns leaving out or mischaracterizing groups or individuals who the alienated subject realizes ought to be included. This exclusion might occur in the model itself, in the intentions of the agent who built the model, or in the purpose for which the model was designed.

Before turning to specific examples, it is worth noting that this form of alienation can feel particularly invasive. Recall that, under Giere's conception of modeling, part of the practice is employing the appropriate intentions. So, learning how to use a particular model requires coming to understand its creator's intentions and operating with them. When one encounters a model that is exclusionary, then, learning to use that model is learning how to think and act in that same exclusionary fashion.[16] This helps to illustrate why it is not the model itself, but the practice around the model that is often at issue.

This, I think, characterizes the alienation described in Collins's (2002) *Black feminist thought*. Chapter 11, "Black feminist epistemology," describes the state of sociology prior to the turn of the twentieth century, explaining how the dominant model of Black women as "welfare queens" affected not only the quality of the research, but also the capacity of Black women like her to engage with the field. This model was exclusive because the Black women it encompassed were fictions of the researchers' prejudice rather than individual people explaining the factors that impacted their choices. The

[15]In this context, alienation is both a harm in and of itself and a symptom of the systemic harms, such as the distributive inequalities that give rise to educational marginalization. Thank you to an anonymous referee for suggesting this point.

[16]This is not to suggest that one need *accept* the relevant intentions and assumptions, however, in the same way that learning to reason with phlogiston theory does not require you to accept phlogiston theory. Thank you to an anonymous referee for pressing this point.

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

representation of the world required to operate with that model is a deeply alienating one. Here, again, Jaeggi's (2014) conception of alienation may be useful in filling out the details of how this experience amounts to alienation: this dissonance between Collins's perspective and those of the community of researchers she wants to engage with limits her capacity to do so.[17]

At the level of the model itself, Plumwood's (1993) critique of the ubiquitous use of the classical negation operator picks out this form of alienation as well. Plumwood argues that we should understand our choice of a negation operator—deciding what the "not" of our model means—as creating "at a very abstract level, certain structures and principles for conceiving and treating otherness," (1993, 441). For Plumwood, the classical negation (the one from Nye's introductory logic class) does more than encode distinctions. It assumes a host of additional features that generates a pernicious hierarchical structure on top of any distinction we use negation to make. For example, a logic with classical negation will feature "radical exclusion" or "hyperseparation," which means that a category and its negation are strictly discontinuous (1993, 448–50).[18] There are no vague cases under classical logic. Of course, there are some realms of analysis, some subject matters, in which this is true. Plumwood's point is that this is not a matter of The One True Logic, it is a matter of the subject. To assume that classical logic is the default correct logic is to assume this is true everywhere. In the case of modeling, then, this assumption will be exclusive of, for example, certain genderqueer people who identify as borderline cases of "man" or "woman."[19] And this is where the alienation of exclusion may arise from within the model itself. While we need to be able to make distinctions in logic, we do not need every distinction to be one that fosters exclusionary thinking. By choosing a logical negation that does not embed these features *as a matter of logic*, we leave room for a less exclusionary model.

4.2. Two worries: ubiquity and turnabout

Given all of this, one might worry that changing our focus to alienation has simply brought us back around to Nye, with alienation lurking behind every p and q, portending a rejection of formal modeling altogether. This worry comes in two forms. First, that alienation is ubiquitous. How can we avoid creating models that alienate certain subjects when alienation requires little more than this disharmony and disconnectedness? Won't this simply be every model? The second worry, turnabout, puts a finer point on this: Won't those in power be able to say the same of feminist and anti-imperialist scholars whose models of oppression seem alienating to them?

The first of these worries is not without merit: There will, indeed, often be someone who feels alienated from a particular modeling practice in at least some circumstance. In part, this is a consequence of abstraction. We are individuals, and abstraction must abstract away from the features that make us individuals, at least to some extent. (This,

[17]Though, as she describes, she is able to do so and eventually flourish, she did so under hardship she ought not have borne and many in similar positions were unable to do the same (Collins, 2002).

[18]Plumwood (1993, 447–52) enumerates several other pernicious features of classical negation: Backgrounding, Relational Definition, Instrumentalism, and Homogenization.

[19]Not all non-binary or genderqueer people identify as such. In fact, there are many versions of genderqueer identity that will be best approached non-classically. For example, people who simply identify as "neither" may also wish for it turn out false they are a man *and* false that they are not a man. This will require non-classical truth tables for the standard connectives.

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

again, is why the close analogy with scientific modeling in physics can obfuscate the role of alienation: There is little lost in abstracting away from the individuality of a particular electron.) As a result, the second worry gains some traction, too: If we suppose that every instance of alienation is enough to scuttle the development, consideration, or application of a model, this complaint will ring from every corner. But it is not the case that every instance of alienation is occasion for abandonment.

Instead, analysis in terms of alienation offers a path toward resolving these worries: Because alienation influences our wellbeing, it invites treatment within the framework of ethical theory.[20] Analysis based on the ideal/nonideal distinction is grounded only in features of the model itself: Does this formal object exhibit the features of an ideal-as-idealized model? Taking alienation as the basis of our critique, however, it becomes clear that the question of whether a particular modeling practice is acceptable cannot be resolved simply by looking at the model itself. Consider, for example, the stress and pain in Nye's experience of alienation, which comes through quite clearly in her description of her introductory logic course. It is *this* that constitutes a reason to evaluate the permissibility of the practice, not the formality of the practice itself.

From this perspective, we can also see how to carry the analysis forward. Continuing with Nye's case, for example: Given that logic serves a range of valuable purposes and must be taught somewhere in order to do so, Nye's sense of alienation in her introductory logic course might not justify forgoing the teaching of logic altogether. Even so, from the standpoint of, for example, care ethics, which emphasizes the moral significance of relationships, trust, dependence, and the like, the deterioration of Nye's relationship with her professor and the broader philosophical community she hoped to join may well constitute grounds for reconsidering the role and teaching of logic in philosophy curricula.[21] Analyzing cases like Nye's through the concept of alienation helps isolate the morally relevant features of the situation and identify them as proper targets of ethical inquiry. First-order ethical theories like care ethics provide the normative resources for carrying out that inquiry.

Turning to the exclusionary alienation case from Collins, we see alienation sewing wariness and distrust. There, Collins describes a deep sense of alienation brought about by the models of US Black women, especially single mothers, developed by white sociologists. Such wariness makes collaboration, much less redress, difficult.[22] Moreover, such distrust may provoke further negative consequences, such as testimonial quieting or smothering (Dotson 2011). These are cases in which improper fit between model, modeler, and modeled causes harm, giving rise to both ethical and practical reasons against the permissibility of these particular practices. Importantly, it is clear that

[20]This may be so even where an instance of alienation is not felt by the alienated subject, at least where objective alienation is a possibility. In such cases, alienation is still understood as problematic and ethically significant. For broader discussion of the ethical import of alienation, see Leopold (2022) and Jaeggi (2014). See n. 11 with respect to the subjective/objective distinction.

[21]More generally, care ethics likely provides a particularly apt framework for understanding the harms of alienation when it comes to modeling because lack of adequate concern for the relationships between modeler and modeled will often be a significant causal factor in the generation of that alienation. See Baier (1987, 1994), Kittay (1999), and Tronto (1993) for articulations of care ethics as an ethical framework. For extended discussion of feminist logic pedagogy, Saint-Croix and Cook (2024), section 5.

[22]Even so, because that distrust is brought about by an apt assessment of alienation, and that alienation arises from an unwillingness to engage with Black women like her unless they are willing to "work within taken-for-granted assumptions of Black female inferiority" (Collins 2002, 254), it is wariness well-placed. Being able to sense and respond to alienation is an important survival mechanism.

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

addressing these harms will not be accomplished by simply de-idealizing. Instead, these are cases in which building trust and familiarity—processes that serve to remediate alienation—are far more likely to be successful.

This is not to suggest that alienation is always harmful in the long run, however. Recall Mills asking readers to engage in an act of Brechtian defamiliarization; producing a sense of alienation in oneself may be uncomfortable or even painful, but nevertheless intellectually productive. So, too, in the situation posed by the turnabout objection, supposing that it counts as an instance of alienation to begin with. The person who feels alienated by a model of oppression (perhaps one that implicates them) may feel pained by that experience, but whether that feeling of discomfort and disconnection amounts to the sort of failure of "proper harmony" that constitutes alienation is another matter. This determination will depend on the details of one's account of alienation. Supposing that the model of oppression in question is an accurate one, it is plausible that this sort of dissonance—between an accurate account of a morally salient aspect of the world and one's own views or self-narrative—cannot *by itself* constitute alienation on a given account of alienation. Even if the turnabout case does meet that threshold, however, a further question arises: What follows? Why think that this particular sort of alienation is cause for curtailing the modeling practice?

So, the turnabout objection faces two hurdles. First, it must be that the turnabout case genuinely constitutes alienation. Second, even if it does, it must warrant a drastic enough response to curtail the modeling practice in question. To see how this might go, consider again Jaeggi's (2014) account of alienation. As we saw, Nye's experience clearly constitutes alienation under this account. But the same is not obviously true of the turnabout case. The mere fact of being presented with a model that impugns one's moral standing does not intrude on one's agency in the same way.[23] If that's right, the turnabout case will not constitute an instance of alienation and, therefore, not trigger the same sort of ethical analysis that Nye's experience does—the subject's psychological distress in the turnabout case may constitute harm, but that experience does not wrong them, at least on Jaeggi's account. But suppose we do grant that the turnabout case constitutes an instance of alienation. What, then, should follow? Here, the answer depends not on the theory of alienation alone, but also on the first-order ethical theory against which we analyze the situation. The discussion of Nye's case in terms of care ethics is one example of how this sort of analysis might be carried out. Continuing with that framework, if the turnabout case really is an instance of alienation, it might nonetheless be morally valuable because it serves as impetus to try to resolve that feeling. Resolving that feeling, in turn, might provoke a better understanding of the relationships of oppression and domination in which the turnabout subject is implicated, and thereby promote their developing more perceptive, more apt relationships of care.[24]

The point is not that identifying alienation yields a verdict on the permissibility of a particular modeling practice. Rather, the point is that this framework directs our attention to the features of the practice that demand further scrutiny. The cause, context, and significance of an instance of alienation all matter to the ethical evaluation of that alienation and the practice giving rise to it. Thus, this lens offers a vocabulary for raising

[23]It may, for example, threaten one's self-narrative as a good person, but much more context would be necessary for such threats to amount to a diminution of agency.

[24]On this analysis, however, the person presenting the model may have some obligation to facilitate such a resolution (as opposed to less positive outcomes, such as the alienated party ignoring the model and shunning feminists forever!), though this will depend on details of the context.

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

questions that the ideal/nonideal distinction tends to obscure: How might this modeling practice—in its creation, use, representation, or purpose, to return to Giere (2010)'s elements of a modeling practice—prompt alienation in others? When, and to what extent, is that risk permissible?[25]

The ubiquity and turnabout objections press a shared worry: that appeals to alienation are too indiscriminate to be useful. As the account developed here demonstrates, however, these objections can be met. Alienation is neither a moral trump card nor a mere synonym for psychological discomfort. Instead, alienation functions as a diagnostic category, illuminating experiences that warrant ethical examination and, potentially, redress. Paired with a first-order normative theory, this enables more responsible modeling practices. Both retrospectively, as in the cases presented by Nye, Collins, and the Combahee River Collective, and prospectively, we can scrutinize those practices not only in terms of their truth and utility, but also with respect to the harms they may engender.

5. Why model?

With this reorientation in mind, I want to close by considering one final worry: Why engage in modeling—especially formal modeling, with its greater risk of alienation—at all? Why should scholars of marginalization and resistance engage in with these practices, having identified a morally salient risk that comes with the practice? To answer this, I suggest that there are at least three functions of formal modeling that are both valuable to scholars of marginalization and resistance and otherwise difficult to achieve. These functions are (1) exposing and excavating assumptions, (2) avoiding obfuscatory accommodation, and (3) being heard beyond the bounds of our academic or political silos. Though these functions do not provide reason for all scholars in these areas to adopt such practices, they are nevertheless important enough to warrant careful use of formal modeling practices. We turn now to examining these functions.

5.1. Exposing and excavating assumptions

Formal models are uniquely adept at exposing and excavating assumptions, especially when it comes to addressing marginalization and demonstrating the need for resistance. Bruner and O'Connor (2017), one of the central cases Hancox-Li (2017) and Táíwò (2023) consider, provides a clear case in point. Brunner and O'Connor (2017) use game theoretical modeling to explore the ways that power relations influence the distribution of benefits arising from academic collaboration. The mechanistic form of game theoretic modeling requires clearly setting out a wide array of essential assumptions. As Hancox-Li (2017, 337) points out, one of those assumptions is that agents interact selfishly, always attempting to adopt a maximally beneficial strategy. This is an important assumption to draw out, because failing to do so might lead to researchers to very different conclusions about the same population and circumstances. Assuming, for example, that agents prefer altruistic strategies would engender starkly different dynamics. Similarly, consider efforts to formalize degrees of social inequality. As Eliazar

[25]Hancox-Li (2017, 343) also argues that we ought to consider the ethical ramifications of different models, but instead focuses on how different idealizations can facilitate ethically important work. This approach also opens connections with existing literature on inductive risk in assessing the adequacy of models for their purpose (Parker 2020) and representational risk (Kukla 2024).

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

(2018) points out, there are many proposed inequality indices, such as the Gini index, the Pietra index, the vertical-diameter index, the horizontal-diameter index, the poverty index, and others. Each of these articulates a different set of assumptions about what matters to measuring inequality and, because they are rendered in mathematically precise terms, these assumptions can serve as a point of comparison and critical feedback. In this way, formal modeling is a service not only to ourselves, but also to our critics.

Formal modeling aids in excavating assumptions as well. While explicit disavowals of sexist, racist, and imperialist tendencies are common, denying a commitment is not the same as innocence from it. And both of those states differ from building a theory that actually avoids it. In either case, the problem is that we very seldom get to the foundations of our own views. But formal modeling requires that we do exactly that. Formal modeling requires that we interrogate, formulate, and state the variables and assumptions that give rise to the situation we are interested in modeling. Thus, formal modeling of this kind has the capacity to expose and allow us to excavate difficult-to-identify assumptions.

Whether we are exploring gender concepts, argumentation norms, or our understanding of the metaphysics of modality, each requires, to varying degrees, that we understand ourselves and our assumptions, especially where they are so fundamental as to be invisible. Creating a model can, in such cases, allow us to hold a mirror up to our own ideas.

This is not to say that formalizing our ideas is a panacea—it is not. But, in creating an object separate from ourselves, we create something easier to study and understand than our own complex, opaque ideation. Whether it is we or our critics who are able to see the assumptions our models build in, those models aid in exposing and excavating them.

5.2. Avoiding obfuscatory accommodation

Formal modeling is also valuable because having an independent artifact with clearly stated assumptions helps to avoid what we might call *obfuscatory accommodation*. This is the tendency of well-meaning parties to a discussion—academics included—to assent to the use of a technical term while knowingly or unknowingly having a significantly different, but unvoiced understanding thereof. The development of feminist standpoint epistemology provides an example of this.[26] To wit, Hekman (1997, 346) argues that Smith (1987) offers an understanding of the standpoint of women on which the knowledge and privilege of women's standpoint is an automatic consequence of their social location.[27] By contrast, most standpoint theorists require that standpoints be achieved by their occupants. Similarly, some standpoint theorists are relativists (Ashton 2019), while most are not (Harding 2009). And some standpoint theorists take the purported advantage of standpoint occupancy to be better access to the truth (Hekman 1997) while others deny that truth, at least in the sense of correspondence theory and similarly realist views, is of much import to standpoint theory at all (Hartsock 1997). Deepening divides, some standpoint theorists take the experience of marginalization to

[26]Roughly speaking, this is the idea that individuals who occupy or relate to particular social locations (in a particular way) have an epistemic advantage over those who do not occupy such positions when it comes to propositions concerning that social position.

[27]Smith (1997) vehemently denies this characterization of her work.

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

be necessary for occupying a standpoint while others do not.[28] While there are many articles, collections, and volumes dedicated to disentangling these confusions, the need for them arose from the evocative, but somewhat ambiguous foundational works.[29]

Each of these choice points results in a different theory with different consequences. Absent a model laying out the assumptions of the theory in distinct, recognizable form, it is easy to let the contours of "standpoint epistemology" shift with the conversation in which it is embedded. In a muddy theoretical space, we may apply the assumptions of one understanding of the theory to a different one without realizing it. At times, this mutability is welcome and helpful. For example, in the context of conversation—even one carried out across journals, volumes, collections, and conferences—in which we are working together to build and understand a concept, grasping at its edges and stumbling in the dark, it is generally good to be able to change our assumptions fluidly, learning as we map the territory. But when we look to share those ideas with others, this mutability becomes a hindrance. Clearly setting out our assumptions and their presumed interactions—modeling them in more or less formal ways—provides a shared point from which to move forward; an epistemic trailhead from which we can share next steps.

5.3. Being heard

Finally, a squarely pragmatic function. The purpose of activism, whether within academia or beyond it, is change.[30] But change is hard. Within scholarships of resistance and marginalization, it is especially hard because it is those in power struggle, willfully or unwittingly, to understand the world in the ways they are being asked to. This is another way that modeling is useful. Returning to Nye (1990), much of her discussion revolves around the observation that models are powerful because they are a language that those in power respect.

In this, Nye's criticism echoes the sentiment expressed by Sandra Harding and Merrill Hintikka in the introduction to their 1983 collection of essays, *Discovering reality*. They describe the goal of their volume as follows: "We must root out sexist distortion and perversions in epistemology, metaphysics, methodology and philosophy of science—in the 'hard core' of abstract reasoning thought most immune to infiltration by social values" (p. ix). While Nye takes this exclusionary reverence for formally inclined work as a reason to bury such tools, Harding and Hintikka take it as a challenge.

This is because the "hard core's" gleam of verisimilitude has its value. Collins (2002, ch. 11), for example, observes that epistemic practices that are not sanctioned by the mainstream—such as the knowledge validation processes of Black women—are easily dismissed by the mainstream. This is surely reason to improve the mainstream. But it is not reason for turnabout—we ought not reject the mainstream methods as epistemically unfounded in retribution. Doing so is a way of being complicit in the intellectual siloing that such dismissal makes possible. Taking up those methods makes such dismissal—dismissal founded on culture and vogue rather than content—untenable. For his part, Mills (2005) takes this possibility to be a crucial reason for saving at least *nonideal* theory, writing: "In certain respects, it engages with mainstream ethics on what are

[28]See Tilton (2024) for extended discussion of this dispute.

[29]Of particular interest in this regard are *Signs: Journal of Women in Culture and Society* 22 (2) (Allen and Howard 1997) and Harding (2001). More recently, Wylie (2013), and Toole (2021) have also set about this task.

[30]Collins (1997) makes this point especially clear with respect to feminist standpoint epistemology.

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

nominally its own terms, thereby (at least in theory) making it somewhat harder to ignore and marginalize" (p. 166). When scholars of marginalization and resistance render their observations in formal terms,[31] we can then share them with scholars and policymakers who do not necessarily understand or care to understand these perspectives on their own terms.[32]

6. Conclusion: modeling as an ethical practice

Critiques of formal modeling practices have often focused on the formal properties of the models in question—how abstract they are, and in what manner. Understood purely in terms of the ideal/nonideal distinction, the most influential of those critiques are, as Táíwò (2023) and Hancox-Li (2017) argue, unconvincing. But this is not to suggest that they can be set aside, as they (along with the very different critiques offered by Nye) are unified by a more subtle undercurrent: alienation. Formal models, modeling practices, and the institutions developed around them can alienate not just by idealizing away from the concrete, but also by shaping practices that disconnect individuals from communities, institutions, and their own agency. Reframing our analysis in terms of alienation rather than the ideal/nonideal binary allows for a more fundamental, more precise approach to assessing particular modeling practices. This reframing, alongside a suitably pluralistic concept of modeling, clarifies the ways that modeling may inflict harm.

On this account, the possibility of alienation operates not as a trump card against a modeling practice, but as a diagnostic tool: it flags moments of tension that call for further ethical evaluation. But the result of that evaluation depends on our broader normative commitments. Through engagement with a first-order ethical theory—whether grounded in care, justice, utility, or something else—we may determine whether that alienation (of the risk thereof) is permissible. As cases like Mills's exercise in Brechtian defamiliarization remind us, alienation may be productive, even if painful. Thus, rather than providing a verdict, focusing on alienation reframes our questions. It shifts attention from structural features of models alone to the broader context of their development, use, ways of representing, and purpose.

Understanding the role and moral weight of alienation facilitates modeling, even highly formalized, abstract modeling, in a way that is conscious of and responsive to the harm that can be done—it facilitates treating modeling as an ethical practice. Much work remains to be done in understanding modeling this way, and I leave that work to future efforts. Even so, with that caution in mind, we can retain a practice with real liberatory value: exposing, exploring, and excavating the assumptions that underlie our theoretical practices.

Acknowledgments. I wish to acknowledge Roy T. Cook, Robin Dembroff, Corey Dethier, Will Fleisher, Liam Kofi Bright, and Jingyi Wu for comments, conversation, and companionship that greatly improved this paper. Thank you. I am also grateful to the Foundations Interest Group at the Minnesota Center for Philosophy of Science, the members of Prof. Dembroff's fall 2023 "Feminist Philosophy" seminar at Yale, the organizers and audience of the "Feminist Philosophy and Formal Logic" session at the 2023 Central Division

[31]For example, by adopting agent-based modeling frameworks (Wu 2022) or formal logic (Kosten 2025).

[32]This is not to suggest that those in power are always persuaded or assuaged by the adoption of formal approaches, however. Rather, this tactic responds to one among many forms of dismissal faced by scholars of marginalization and resistance.

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

meeting of the American Philosophical Association, and the anonymous referees whose feedback contributed to this work.

References

Allen, Carolyn, and Judith Howard. 1997. Editor's note. *Signs: Journal of Women in Culture and Society* 22 (2): 366.

Ashton, Natalie Alana. 2019. Relativising epistemic advantage. In *The Routledge handbook of philosophy of relativism*, ed. Martin Kusch. New York: Routledge.

Baier, A. 1987. Hume, the women's moral theorist. In *Women and moral theory*, ed. Diana T. Meyers. Totowa, NJ: Rowman & Littlefield Publishers.

Baier, Annette. 1994. *Moral prejudices: Essays on ethics*. Cambridge, MA: Harvard University Press.

Binnema, Theodore. 2001. How does a map mean? Old Swan's map of 1801 and the Blackfoot world. In *From Rupert's Land to Canada: Essays in honour of John E. Foster*, ed. Theodore Binnema, Gerhard J. Ens, and R. C. Macleod. Alberta: University of Alberta Press.

Bousquet, Chris. 2023. Work and social alienation. *Philosophical Studies* 180 (1): 133–58.

Bowman, Melanie, and Roy T. Cook. forthcoming. The logic of knowing-what-it-is-like. In *Feminist philosophy and formal logic, Minnesota studies in the philosophy of science*, ed. Roy T. Cook and Audrey Yap. Minneapolis: University of Minnesota Press.

Bradley, Ben. 2012. Doing away with harm. *Philosophy and Phenomenological Research* 85 (2): 390–412.

Brecht, Bertolt. 1977. Alienation effects in Chinese acting (1936). In *Brecht on theatre: The development of an aesthetic*, ed. Steve Giles, Marc Silberman, Tom Kuhn, and John Willett. New York: Hill & Wang.

Bruner, Justin, and Cailin O'Connor. 2017. Power, bargaining, and collaboration. In *Scientific collaboration and collective knowledge*, ed. Thomas Boyer-Kassem, Conor Mayo-Wilson, and Michael Weisberg. New York: Oxford University Press.

Callender, Craig, and Jonathan Cohen. 2006. There is no special problem about scientific representation. *Theoria: Revista de Teoría, Historia y Fundamentos de la Ciencia* 21 (1): 67–85.

Collins, Patricia Hill. 1997. Comment on Hekman's "Truth and method: Feminist standpoint theory revisited". *Signs: Journal of Women in Culture and Society* 22 (2): 375–81.

Collins, Patricia Hill. 2002. *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. New York: Routledge.

Combahee River Collective. 1979. A Black feminist statement. In *Capitalist patriarchy and the case for socialist feminism*, ed. Zillah Eisenstein. New York: Monthly Review Press.

Cook, Roy T., and Audrey Yap. forthcoming. Introduction. In *Feminist philosophy and formal logic, Minnesota Studies in the Philosophy of Science*, ed. Roy T. Cook and Audrey Yap. Minneapolis: University of Minnesota Press.

Crary, Alice. 2018. The methodological is political. *Radical Philosophy* 2 (2): 47–60.

Dotson, Kristie. 2011. Tracking epistemic violence, tracking practices of silencing. *Hypatia* 26 (2): 236–57.

Eliazar, Iddo. 2018. A tour of inequality. *Annals of Physics* 389: 306–32.

Frigg, Roman, and Stephan Hartmann. 2024. Models in science. In *The Stanford encyclopedia of philosophy*, Fall 2024 edn., ed. Edward N. Zalta and Uri Nodelman. Stanford, CA: Metaphysics Research Lab, Stanford University.

Frigg, Roman, and James Nguyen. 2017. Models and representation. In *Springer handbook of model-based science*, ed. Lorenzi Magnani and Tommaso Bertolotti. New York: Springer.

Giere, Ronald N. 2010. An agent-based conception of models and scientific representation. *Synthese* 172 (2): 269–81.

Gilabert, Pablo. 2020. Alienation, freedom, and dignity. *Philosophical Topics* 48 (2): 51–80.

Hancox-Li, Leif. 2017. Idealization and abstraction in models of injustice. *Hypatia* 32 (2): 329–46.

Hardimon, Michael O. 1992. The project of reconciliation: Hegel's social philosophy. *Philosophy and Public Affairs* 21 (2): 165–95.

Harding, Sandra. 2009. Standpoint theories: Productively controversial. *Hypatia* 24 (4): 192–200.

Harding, Sandra G. 2001. *The feminist standpoint theory reader: Intellectual and political controversies*. New York: Routledge.

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

Harding, Sandra, and Merrill B. Hintikka. 1983. *Discovering reality: Feminist perspectives on epistemology, metaphysics, methodology and philosophy of science*. Boston: D. Riedel.
Hartsock, Nancy. 1997. Comment on Hekman's "Truth and method: Feminist standpoint theory revisited": Truth or justice? *Signs: Journal of Women in Culture and Society* 22 (2): 367–74.
Hekman, Susan. 1997. Truth and method: Feminist standpoint theory revisited. *Signs: Journal of Women in Culture and Society* 22 (2): 341–65.
Hudson's Bay Company Archives, Archives of Manitoba, Peter Fidler fonds, Peter Fidler journals of exploration and survey, 1792–1806, Sketch map "Drawn by the Feathers or ak ko mock ki a Black foot Chief 1802,", redrawn by Peter Fidler. HBCA H1-30-3 (E.3/2), fo. 104.
Hunt, Dallas, and Shaun A. Stevenson. 2017. Decolonizing geographies of power: Indigenous digital counter-mapping practices on Turtle Island. *Settler Colonial Studies* 7 (3): 372–92.
Jaeggi, Rahel. 2014. *Alienation*. New York: Columbia University Press.
Kittay, Eva Feder. 1999. *Love's labor: Essays on women, equality, and dependency*. London: Routledge.
Kosten, Rebecca. 2025. A transformative intuitionist logic for examining negation in identity-thinking. *Australasian Journal of Logic*. 22 (4): 438–74.
Kukla, Quill. 2024. Maps and the epistemic risks of visual representation. *Journal for the Philosophy of Language, Mind and the Arts* 5 (1): 39–60.
Leopold, David. 2022. Alienation. In *The Stanford encyclopedia of philosophy*, Winter 2022 edn, ed. Edward N. Zalta and Uri Nodelman. Stanford, CA: Metaphysics Research Lab, Stanford University.
Lowan-Trudeau, Greg. 2021. Mapping (as) resistance: Decolonizing ↔ indigenizing journalistic cartography. Media+environment 3 (1): 1–22.
MacPherson, Brian. 1999. Three misrepresentations of logic. *Informal Logic* 19 (2): 185–99.
Makarenko, Jay. 2009. Fair opportunity to participate: The pharter and the regulation of electoral speech. *Canadian Political Science Review* 3 (2): 38–51.
Mills, Charles W. 2005. "Ideal theory" as ideology. *Hypatia* 20 (3): 165–83.
Nye, Andrea. 1990. *Words of power: A feminist reading of the history of logic*. New York: Routledge.
O'Neill, Onora. 1987. Abstraction, idealization and ideology in ethics. *Royal Institute of Philosophy Lectures* 22: 55–69.
Parker, Wendy S. 2020. Model evaluation: An adequacy-for-purpose view. *Philosophy of Science* 87 (3): 457–77.
Plumwood, Val. 1993. The politics of reason: Towards a feminist logic. *Australasian Journal of Philosophy* 71 (4): 436–62.
Purves, Duncan. 2019. Harming as making worse off. *Philosophical Studies* 176 (10): 2629–56.
Rawls, John. 1999. *A theory of justice: Revised edition*. Cambridge, MA: Harvard University Press.
Russell, Gillian. 2024. From anti-exceptionalism to feminist logic. *Hypatia* 39 (1): 79–96.
Saint-Croix, Catharine. 2020. Privilege and position: Formal tools for standpoint epistemology. *Res Philosophica* 97 (4): 489–524.
Saint-Croix, Catharine, and Roy T. Cook. 2024. (What) is feminist logic? (What) do we want it to be? *History and Philosophy of Logic* 45 (1): 20–45.
Shiffrin, Seana. 2012. Harm and its moral significance. *Legal Theory* 18 (3): 357–98.
Smith, Dorothy. 1987. *The everyday world as problematic: A feminist sociology*. Northeastern series in feminist theory. Toronto: University of Toronto Press.
Smith, Dorothy. 1997. Comment on Hekman's "Truth and method: Feminist standpoint theory revisited". *Signs: Journal of Women in Culture and Society* 22 (2): 392–98.
Táíwò, Olúfémi O. 2023. Uncommon features: Defending ideal theory with model-to-world inference. In *Routledge handbook on African political philosophy*, ed. Uchenna Okeja. London: Routledge.
The MappingBack Network. 2022. About us—Mappingback.org. https://web.archive.org/web/20230322021651/http://mappingback.org/about-us/
Tilton, Emily C. R. 2024. "That's above my paygrade": Woke excuses for ignorance. *Philosophers' iIprint* 24 (8): 1–19.
Toole, Briana. 2021. Recent work in standpoint epistemology. *Analysis* 81 (2): 338–50.
Tronto, J. C. 1993. *Moral boundaries: A political argument for an ethic of care*. London: Routledge.
Valentini, Laura. 2012. Ideal vs. nonideal theory: A conceptual map. *Philosophy Compass* 7 (9): 654–64.

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press

Walkerdine, Valerie. 1988. *The mastery of reason: Cognitive development and the production of rationality*. London: Routledge.
Weiner, Joan. 1994. Review: Andrea Nye, Words of power. *A feminist reading of the history of logic, Journal of Symbolic Logic* 59 (2): 678–81.
Wiersma, Wouter. 2026. Alienation as a social pathology: Evaluating Jaeggi's concept of alienation. *Journal of Social Philosophy* 1–11.
Wu, Jingyi. 2022. Epistemic advantage on the margin: A network standpoint epistemology. *Philosophy and Phenomenological Research* 106 (3): 755–77.
Wylie, Alison. 2013. Why standpoint matters. In *Science and other cultures*. London: Routledge.

Catharine Saint-Croix is an Assistant Professor of Philosophy and Stephen R. Setterberg, M.D., Faculty Fellow in Philosophy at the University of Minnesota–Twin Cities. They are also a Research Associate with the African Centre for Epistemology and Philosophy of Science (ACEPS) at the University of Johannesburg. Their research interests include formal and social epistemology, the epistemology of attention, and feminist philosophy.

Cite this article: Saint-Croix C 2026. "Alienation, Idealization, and Modeling." *Hypatia* 1–22. https://doi.org/10.1017/hyp.2026.10060

https://doi.org/10.1017/hyp.2026.10060 Published online by Cambridge University Press