Preprint. Published in the *Journal of Economic Methodology*, 2025

Minimal Models, Feminist Epistemology, and Diversity[1]

By Patricia Marino, patriciamarino.org

Abstract:

This paper draws on feminist epistemology and epistemologies of ignorance to consider debates over "minimal" economic models and to showcase implications for diversity. Minimal models are highly idealized models put forward without specific empirical support. Criteria for evaluation include intuition, fit with background knowledge, and imagination. Minimal models may be interpreted modally, as giving us "how possibly" rather than "how actually" explanations; they are said to add to our "menu" of possible explanations. Feminist epistemology emphasizes that perspectival differences have epistemic consequences; epistemologists of ignorance show how social position influences what we do not know. Using the checkerboard model of segregation as an example, I argue 1) that because evaluation of minimal models rests on subjective criteria, their use gives us reasons to pursue diversity in the epistemic community and 2) that because of ignorance, adding to our menu of possible explanations can have epistemic risks.

Introduction

This paper explores recent debates over "minimal" models through a feminist epistemology lens to showcase implications for diversity. In the economic methodology context, minimal

[1] I am grateful to Luca Garzino Demo, Carla Fehr, Till Grüne-Yanoff, Sahar Heydari Fard, Alysha Kassam, Kareem Khalifa, Samuli Reijula, Alison Wylie, and two anonymous referees for this journal for suggestions, discussion, and feedback. Earlier versions of this work were presented at the 2023 meeting of the Canadian Society for this History and Philosophy of Science, the 2023 meeting of the European Network for the Philosophy of Social Science, the 2024 meeting of the Philosophy of Social Science Roundtable, and a workshop hosted by the editors for this special issue; my thanks to all the participants for their questions and comments.

models have been characterized as simple, highly idealized, theoretical models put forward without specific empirical support. Some proposals for evaluating minimal models focus on credibility: like a realistic novel, they provide a way the world could be. Others focus on the modal quality of minimal models: they may not tell us how things are, but they tell us how things could be, providing "how-possibly" explanations. In some interpretations, model pluralism -- the use of models in clusters -- is central, and minimal models are useful for adding to our menu of explanations potentially relevant to a situation.

Feminist standpoint theory has long emphasized that knowledge is situated: knowers' experiential and perspectival differences have epistemic consequences. Here I explore these consequences for the evaluation of minimal models. Using the checkerboard model of segregation as an illustrative example, I show that because evaluation of minimal models relies on factors such as background knowledge, intuition, and imagination, the use of minimal models shows the importance of situational diversity in the epistemic community. I then discuss the importance of apt evaluation of minimal models by explaining how the use of minimal models can have epistemic risks. In section one, I give an overview of various accounts of minimal models, showing how their evaluation rests on fit with background knowledge, intuition, and imagination, and identifying a few specific interpretative ways minimal models can be used. In section 2, I explain basic elements of feminist standpoint theory then use the checkerboard model to explore the ways that model evaluators from different social situations may come to different conclusions about minimal models. In section three, I consider potential harms of evaluating minimal models in ways that are not fully informed. While diversity of various kinds may be ultimately important for model evaluation, I focus in this paper on diversity of social situation.

As this paper engages the topic of how social situation informs perspective, it is appropriate to mention briefly my own social situation as a white woman originally from the US and now living in Canada. Content in this paper reflecting on how social positions may inform background knowledge, intuition and imagination are derived from my understanding of our social world.

1. Minimal models and their epistemic evaluation

In this section, I show how elements such as intuition, imagination, and fit with background knowledge are substantive elements in the epistemic evaluation of minimal models. After briefly characterizing such models and explaining the checkerboard as an example, I present the three most developed and influential accounts for minimal models might be evaluated and understood: a credibility account, a modal-conditional account, and a cluster-based account. For each, I show the ways that intuition, imagination, and fit with background knowledge come into play in evaluation. Minimal models can be used for various epistemic purposes, to draw different kinds of inferences, so I then proceed by drawing on these accounts to identify a few interpretations of minimal models, where an interpretation focuses on a particular inferential use.

The term "minimal model" is used in a range of ways, especially in different contexts. Is focus here on its use in recent debates in philosophy of economics and social science, where it is related to the concept of a "toy" model. Minimal models are highly simplified and idealized, and typically considered in the absence of specific empirical justification or evidence.[2] Toy models are understood as "extremely simple," as "highly idealized," and as representing "some target

[2] Grüne-Yanoff's characterization includes that minimal models lack similarity relations between the model and the world; since some others do not, we do not include this in our characterization.

system(s) in the world" (Reutlinger et al. 2018, Nguyen 2020). In other philosophy of science contexts, the term "minimal model" may be used for highly simplified or abstract models that are embedded in an established theory or that lead to specific new testable hypotheses (see, e. g., Reutlinger et al. 2018, Bokulich 2014). But the minimal models we consider here are not embedded; in fact, they are often used to challenge some part of our background understanding."[3] And these minimal models may not lead to specific new testable hypotheses.

The much-discussed checkerboard model of segregation originally developed independently by Sakoda and Schelling is often considered a minimal model in the relevant sense. In one version, dimes and pennies are arranged on a checkerboard then moved according to certain rules -- typically, that a coin is not "content," and wants to relocate, if a certain proportion (say, a third) of its immediate neighbors are of the other type. Because the preferences in the checkerboard model are mild, shared, and symmetrical, the model is taken to show, "somewhat counterintuitively," that "agents with mild discriminatory preferences who could live happily in integrated neighborhoods will end up living in segregated neighborhoods only because they do not want to be in an extreme minority" (Ylikoski and Aydinonat 2014). Other models interpreted as minimal models are Akerlof's market for lemons (Reutlinger et al., 2018), the Arrow-Debreu theorem of general equilibrium in economics (Grüne-Yanoff & Verreault-Julien 2021, Verreault-Julien 2017) and the hawk-dove model in evolutionary biology (Tan 2022). These are examples of what Reutlinger et al. (2018) call "autonomous" models -- that is, they are not embedded in a well-confirmed framework theory, and thus cannot derive justificatory support from the framework.

[3] The economist Rodrik says that economic models "open our eyes to counterintuitive possibilities and unexpected consequences" (2015, 46).

My focus is epistemic evaluation. In a series of papers, Sugden addresses the question of how to interpret economic models that are "abstract and unrealistic," and "lead to no clearly testable hypotheses" (2000, 2; see also 2013). In presentations of these models, Sugden says, authors point to a real-world regularity R, create a model in which a set of casual factors F produce R in an extreme form, then conclude that the model gives us "some reason" to believe that F causes R in the real world. Sugden asks: what justifies this conclusion?

Sugden's answer is "credibility": such models are useful insofar as they are credible and thus provide a description of the way the world "could be" (2000, 2013). Models are credible when they are internally coherent, when their assumptions fit with background knowledge, and when the situation they present "could be" real, even if counter-factually. A useful model should describe "a state of affairs that is credible, given what we know (or think we know) about the general laws governing events in the real world" (2000, 25). For Sugden, similarity is central among the factors that license confidence that if F causes R in the model, then we have reason to believe that F causes R in the real world.

The checkerboard model is credible, Sugden says, because it creates an imaginary city that is "inhabited by people who are *like* real people" and in which segregation arises and then persists through various spatial changes, mirroring segregation real US cities (2000, 27). Further, he says, the assumption that people have mild segregationist preferences is "justified by psychological and sociological evidence" and "coheres with common intuition and experience (2000, 26)." There are therefore relevant similarities between the model world and the real world. Thus, he says, "we have been given some reason to think that segregation in real cities is caused by preferences for segregation, and that the extent of segregation is invariant to changes in the strength of those preferences. (2000, 24)."

Most relevantly here, Sugden says that judgments of similarity are "subjective," partly because people with different background knowledge may come to different judgments about credibility: "judgements of credibility can be affected by such factors as the judge's experiences, values, upbringing and education, and the theoretical preferences and history of the community of researchers to which she belongs" (Sugden 2013, 241-242). As Sjölin Wirling and Grüne-Yanoff point out, Sugden's analogy between credible models and realistic novels "emphasizes the role of the imagination, which plays a central role in most accounts of fiction (2021, 6)." Evaluation on the credibility account thus includes intuition, imagination, and fit with background knowledge.

Before we move on from Sugden and credibility to consider a second account, it is useful to note that the causes of segregation in the actual world are obviously complex. In this paper, I will focus on the context of US cities for specificity. While sociologists emphasize that we lack the kind of empirical data that would lead to precise conclusions about how a range of contributing factors may be in play, it is well-known that the central causes of US residential racial segregation are due to racism, including preferences shaped by racism as well as institutional discrimination. Preferences shaped by racism include racist preferences causing white people to move out of neighborhoods when Black people move in and non-racist preferences such as those causing people in marginalized groups to move to more integrated neighborhoods to reap the benefits of better schools or to move to more segregated areas to live further away from racist white people. Institutional discrimination includes racist policies such as redlining, where financial services were withheld from neighborhoods that with significant numbers of Black people, the greater likelihood of a Black person being poor than a white person due to systemic racism, racial discrimination at institutions like banks that grant loans and

mortgages, real estate "covenants," which encoded and formalized white's desires to sell only to other white families, and other manifestations of white supremacy (Ellen and Steil, 2019).

Causes of segregation traceable to racism are asymmetrical: as we know from theorists of oppression, racism has to do not with drawing racial distinctions but rather with hierarchies -- with the attitudes, actions, and social patterns sustaining white people's advantages over those of people of color (see, e. g., Lebron 2013, Mills 2003, Taylor 2013). By contrast, the "preferences" of the coins in the checkerboard model are symmetrical -- everyone has the same degree and kind of in-group preferences. The "preferences" in the model are often interpreted as innocuous -- Sugden describes them as "forgivable" -- and they are often seen as grounding an explanation of segregation that would be an alternative to causes related to racism. For convenience, I'll call mild, symmetrical, innocuous in-group preferences tracking race in the residential context "checkerboard preferences."

Even where the checkerboard model fails to capture the most relevant current factors causing actual residential segregation, it is still seen as useful, as it may tell us about possible causes or about actual causal factors that not the most relevant current factors. A second account of the epistemology of minimal models emphasizes the former: their modal contributions. Here, the checkerboard model is understood as giving us modal information because it shows that mild, symmetrical, in-group preferences *can* cause segregation in the absence of racism. Grüne-Yanoff says that before the model's introduction, "many people believed that segregation was necessarily a consequence of explicitly racist preferences" (2009, 96), but the model "forced people to change their confidence in the racism hypothesis" (2009, 96). In the absence of specific empirical justification, he points out, a minimal model may not license conclusions about real causal explanations, but it may still offer information about what is possible and thus provide a

source of learning. Minimal models present "relevant" possibilities (2009, 97). This approach links up with a strand of theorizing in which minimal models lead to "how possibly explanations" (HPEs) rather than how-actually explanations (HAEs) (Grüne-Yanoff and Verreault-Julien, 2021). Seen through a modal lens, the checkerboard model may not tell us about the actual causes of residential segregation in a particular city, but it can help us identify possible causes and understand possibility claims and if-then dependencies.

With respect to the modal aspects of models, it is crucial to distinguish different kinds of possibility. Grüne-Yanoff and Verreault-Julien distinguish between "epistemic" and merely "objective" possibility: the former the concerns possibility in our world; the latter includes "logical" possibility where a model represents a possible way a world could be but may invoke "an explanans that is known not to be actually true" (2021, 117). An epistemically possible how-possibly explanation (EpHPE) is one that "is not ruled out by an agent's (or epistemic community) knowledge about what is actually the case" (Grüne-Yanoff and Verreault-Julien 2021, 116) and may be developed on the way to finding HAEs. An explanation can, however, be objectively possible without being epistemically possible -- possible in some imaginary world, but not in ours -- leading to an ObHPE. Within the range between "epistemic" possibility and "logical" possibility, we may be interested in "physical, biological, political, or economical possibility" (Sjölin Wirling and Grüne-Yanoff 2021).

In an overview of the epistemology of modal models, Sjölin Wirling and Grüne-Yanoff say that in addition to internal coherence, in an apt modal model "development in the imagined world -- what 'happens' in the fiction/model -- must be judged to be plausible conditional on the background information provided about for example preferences, environment, and so on" (2021, 6). They add that in judging conditional relationships, "only the assessment of a

competent user of the model will do -- someone with the appropriate background knowledge and experience, that is." Grüne-Yanoff suggests that the conditional judgments used to evaluate HPEs are "driven by empathy, understanding, and intuition" (2009, 95). Because epistemic evaluation of modal models rests on conditional judgements, I'll call this mode of evaluation "modal-conditional." As this discussion shows, evaluation in this mode involves intuition and imagination; to evaluate whether an HPE is also an EpHPE requires also fit with background knowledge.

A third account of the epistemology of highly idealized models emphasizes the role models play in developing a useful diversity or "cluster" of models and the ways models can identify potential causal factors that are not the main or most important causal factors in the actual world. Ylikoski and Aydinonat (2014) argue that models with HPEs should not be evaluated in isolation for their reliability or fit; instead, theoretical models are used in clusters and add to our menu of possible explanations. Instead of asking about the relationship between a model and the world, we should focus on the ways that models used in clusters provide explanations of "causal mechanism schemes," not specific scenarios. These schemes are "simplified theoretical explananda," and finding models for them can help show that some "scenario candidates should be dropped from consideration ..." (2014, 27; see also Verreault-Julien 2019, 7). That is, "theoretical models modify the menu of possible mechanisms" (p. 28).

For example, in the checkerboard context, Ylikoski and Aydinonat say that the cluster approach explains why the model is useful despite the fact that it ignores "all well-known causes of segregation" including "economic factors, welfare differences among groups, and organized discrimination" (2014, 21). The model should not be interpreted as providing explanations of "empirical cases" of residential racial segregation, they say. Instead, it provides a "generic

template" for thinking about a *possible* way in which general segregation emerges, where general segregation is understood broadly to include non-random allocation into groups in various contexts and settings -- that is, a range of in-group sorting effects. The checkerboard model provides one of a "family" of "competing explanations" for similar kinds of in-group sorting effects.

In this framing, given that we have evidence that the main causes of segregation in actual cities are not checkerboard preferences but are instead factors related to racism, there are a few possibilities: 1) checkerboard preferences may function as contributing causes alongside these other causal factors; 2) these other causal factors may "preempt" the effect of checkerboard preferences by not allowing them "room to operate"; and 3) checkerboard preferences may be relevant counterfactually, telling us that residential segregation would result in actual cities even if other factors -- such as racist preferences or institutional discrimination -- were eliminated.[4] In these cases, we might see the HPEs in the checkerboard model as potential HAEs -- and thus EpHPEs -- for contributing, preempted, or counterfactually relevant causes. Work in mathematical sociology drawing on theoretical models uses the concept of "overdetermination": it is suggested that actual residential segregation may be the consequences of "multiple sufficient causes" (Fossett 2006, 194-195; see Macy and Van De Rijt 2006, 276), so that discriminatory practices "may be sufficient to produce segregation, but eliminating these factors may have minimal impact on ethnic residential distributions due to the persisting effects of ethnic in-group

[4] In some places Ylikoski and Aydinonat (2014) seem to take preferences shaped by racism (and thus strong discriminatory preferences) a variation on checkerboard preferences and thus represented by models in the checkerboard's cluster. For simplicity, in this paper I focus on framings in which the causal factors in the checkerboard model are seen as alternatives to causes due to racism, and I interpret preferences shaped by racism as being in the same category as institutional discrimination.

preferences (Macy and Van De Rijt 2006, 276)."[5] Overall, Ylikoski and Aydinonat say that simply by introducing a possible mechanism for general in-group sorting, the checkerboard model expanded the menu of possible causes of residential segregation and thus changed the "evidential landscape," making the requirements for an acceptable explanation of a segregation process "much more stringent" (2014, 31).

In this cluster account, it is not obvious how minimal models are to be evaluated epistemically (Mireles-Flores 2018, 100); Ylikoski and Aydinonat do not present their account with an associated distinct proposal. Of course, models need epistemic evaluation if we use them to draw conclusions about overdetermination, contributing or preempted causes, and associated counterfactual inferences in the actual world. But notice that even to be useful for additions to our "menu," models must be epistemically evaluated. This is because we do not want to add to the menu just any logically possible explanation: without the right criteria, we would have an unworkably wide array of models -- an "embarrassment of riches" (Grüne-Yanoff and Marchionni 2018, 273). From the discussion above, I suggest we may use either credibility or modal-conditional evaluation when determining whether a model is useful for expanding our list of possible explanations; from this it follows that adding appropriately to our menu requires evaluation based on elements such as fit with intuition, imagination, and fit with background knowledge. In support of this conclusion, in defending the practice of working with a plurality of models, economist Rodrik says that "intuitiveness" and "plausibility" are among the constraints we use in deciding whether to add to our menu: like a fable, a model added to the menu should

[5] A complexity here is that they are contrasting preferences (of any kind) and institutional factors rather than racist causes with non-racist causes, and they use ethnicity, not race as the central concept. See previous note.

provide a narrative "whose storyline revolves around clear cause-and-effect, if-then relationships" (Rodrik 2015, 19; see Grüne-Yanoff and Marchionni 2018, 266).

Discussion of these three accounts highlights the complexity of the relationship between epistemic evaluation of minimal models and how the models are used. We've seen two distinct proposals for evaluation: credibility and modal-conditional evaluation, and we've seen a range of potential uses to a variety of conclusions. For clarity, I will outline and focus on a few specific interpretations -- ways minimal models may be understood so they are suited to specific epistemic uses.

First, there are two candidate interpretations -- where a model suggests how-possibly explanations (HPEs) that could, potentially, turn out to be how-actually explanations (HAEs). There is one main-cause candidate interpretation in which a model suggests an explanation that is a candidate for a main or important current explanation in the actual world: here, the model is taken to identify a current, highly relevant, causal factor. For the checkerboard model, this interpretation would say of one or more specific actual cities that we might learn that segregation there has been caused mainly by checkerboard preferences rather than racism and institutional discrimination, and that the model gives us "some reason" to believe that it has. Then there is second candidate interpretation in which a model provides an explanation based on a causal factor that is a just a contributing or preempted causal factor in the actual world as in the "overdetermination" frame. For the checkerboard model, this would be saying of a range of specific actual cities that segregation there could have arisen through a combination of sufficient causes, including checkerboard preferences but also including other causes such as institutional discrimination and other factors related to racism. Either candidate interpretation could fit what is meant by "adding to our menu" and either could fit Sugden's idea that a credible model gives

us "some reason" to believe the causal factors cause the given regularity in the actual world. Epistemic evaluation for either could be via the modes of either credibility or modal-conditional evaluation for status as EpHPE.

Second, there is an interpretation in which the model gives us some reason to believe that under certain changes or interventions to the actual world, a certain effect would result -- I call these "predictive counterfactuals." For example, the checkerboard model may be used to support inferences about what would happen in actual cities if causes due to racism were to disappear. For this kind of interpretation, to infer what would happen in a possible future for the actual world could be through the overdetermination lens evaluated via credibility: insofar as the model gives us reason to believe that checkerboard preferences form contributing or pre-empted causes in actual cities, we could conclude that in those cities, if we eliminated other causes, segregation would still persist. Note that this way of reasoning requires seeing the model as yielding more than candidate explanations: we would have to have reason to believe the relevant causal factors are factors in the actual world which would be unaffected by the changes indicated by the relevant counterfactual. Or to use modal-conditional evaluation, epistemic support would come from reflecting on whether what "happens" in the model world is judged to be plausible "conditional on the background information"; for the checkerboard, this might mean imagining possible residential choices in a future of the actual world in which factors such as racist preferences and institutional discrimination have disappeared.

In a third and most abstract interpretation, a model can be used to understand possible worlds that may not be the actual world, even though the inferences may be relevant to understanding the actual world. In the checkerboard case, the model could tell us that in possible, imaginary cities where we observe segregation, this may not be due to racism. In this

interpretation, inferences take the form of conditional dependencies featuring conditions that may not obtain in the actual world. In the checkerboard case, such an inference might be "if people make residential choices based only on checkerboard preferences, then segregation will occur" -- unlike the predictive counterfactuals above, the validity of this conclusion does not depend on whether we have checkerboard preferences in the actual world, it simply says what would happen if we did. Models in this interpretation could be used for adding to our menu of possible explanations for the actual world, in which case we are back to the "candidate" interpretations above. But they could also be used for assessing inferences about what is possible in different modalities. For example, when we say "If individuals had not strong discriminatory preferences, then residential segregation could still occur" (Verreault-Julien 2019, 12), we may consider whether the "could" indicates merely logical possibility or something stronger such as sociological possibility. To say that the explanations suggested by the checkerboard model are ObHPEs might be to say that it's logically possible for cities to exist where segregation is caused only by checkerboard preferences or to say in a logically possible future of the actual world, checkerboard preferences alone cause residential segregation. To say the explanations suggested by the checkerboard model are sociologically possible might be to say that there are possible worlds sharing our sociological causal generalities but where various contingencies are such that segregation there is caused by checkerboard preferences only. As this interpretation concerns possible worlds, the mode of evaluation most relevant here is modal-conditional.

Thus, we have two modes of epistemic evaluation (credibility and modal-conditional evaluation) and we have several interpretations: two candidate interpretations (relevant to actual main causes or actual contributing/pre-empted causes) a predictive counterfactual interpretation (relevant to inferences that in the actual world, under certain changes or interventions, certain

effects would arise) and a possible-world interpretation (relevant to casual connections in possible worlds that may not be the actual world).

2. Evaluating minimal models: social situation and diversity

In this section, I briefly explain standpoint theory, apply it to the above interpretations of the checkerboard model to showcase implications for diversity, then draw more general conclusions about how situational diversity may be relevant for evaluating minimal models.

The strand of feminist epistemology known as standpoint theory focuses on the complex and interwoven ways that what a person is in a position to know can depend on factors beyond the epistemological, including those related to the person's social position (Collins 1990, Grasswick 2018, Mills 2014). The most relevant aspect of standpoint theory here is "situated knowledge": the way that "contingent histories, social context and relations, inevitably affect what epistemic agents know" (Wylie 2012). In particular, those in socially marginalized positions may have access to knowledge that the comparatively privileged do not, because the latter do not have the relevant experiences and background. In the sociology context, Collins (1991) argues that the questions asked, information gathered, and methods used reflect researchers' positions, often dominated by white men; she draws on her experience growing up in a working-class Black US family to challenge prevailing assumptions and framings.

Relatedly, scholars of the epistemologies of ignorance emphasize that social position influences not only what we know but also what we do not know (see, e. g., Mills 2014 and Medina 2016). For example, people racialized in various ways occupy different positions with respect to understanding racism, with Black people experiencing first-hand phenomena like over-policing, discrimination, and explicit and implicit racist attitudes. Crucially, white people are

often ignorant of how racism works in their society. This ignorance is produced and regenerated by various forces (see Sullivan and Tuana 2012, especially Mills 2007). White people may benefit from their own ignorance about racism as it blocks the burden of having to address it, so there can be benefits to not knowing; those experiencing oppression may have a better epistemic position. As Wylie puts it, "entrenched, systemic inequalities in our material and social conditions of life can be epistemically enabling" (2012, 55). Furthermore, Harding (1991) argues that we gain "strong objectivity" only when we have knowledge from different perspectives (see Rolin 2006). In this view, since knowledge is perspectival, a full understanding of the world can only result from incorporating a representational range of perspectives; by seeing phenomena from different points of view we gain understanding that is less "partial" and hence "more objective" (Grasswick 2018).

An epistemic community may have situational and/or epistemic diversity. As Fehr (2011) explains, a community is "situationally diverse" when "its membership consists of individuals with different social and material locations (gender, race, class, sexuality, etc.)," while a community is "epistemically diverse" when it includes members with "a range of different background assumptions" who hold various "theoretical and methodological perspectives." Situational diversity can lead to epistemic diversity, as when a person's social situation leads them to have particular background knowledge, experience, and perspective, and when these inform the community's practices and conclusions.[6]

While various forms of epistemic diversity are likely relevant to assessing minimal models, I focus in this paper on situational diversity. I argue that the checkerboard model provides a vivid

[6] Fehr (2011) emphasizes that for situational diversity to lead to beneficial epistemic diversity, scientific communities must cultivate dissenting voices, nurture epistemic diversity workers, and avoid diversity "free-riders."

case study showing the possibility that people from different social situations may come to different conclusions about whether a minimal model is apt for a given use, with evaluation from those in marginalized communities possibly epistemically superior. Because evaluation of minimal models relies on subjective and variable elements such as background knowledge, intuition, and imagination, the socially situated aspect of epistemological judgment will be particularly salient.

Consider first the "candidate" for a main or important explanation interpretation, where we consider the model as identifying causal factors that are candidates for the most central or important causes in actual world. In the checkerboard case, we would have to gauge how credible it is that checkerboard preferences are a central or important cause for actual segregation. A person racialized as Black who has experienced segregation in particular US cities may form judgements about its causes grounded in racism and thus in asymmetrical forces importantly different from the mild, shared, symmetrical preferences in the checkerboard model. On this basis, they may judge the checkerboard not credible for this interpretation: it is unintuitive and does not fit with background knowledge. But to a white person who has experienced shared, mild, symmetrical in-group preferences in contexts like social events, and has not experienced the effects of racism first-hand, the model may seem credible, including when applied to racial segregation in US cities. This is especially so given white ignorance about racism and its effects. In the modal-conditional mode, where the model suggests epistemically possible HPEs (EpHPEs), fit with background knowledge is central; again, insofar as white people are less likely to have knowledge of how racism pervades their society, white model evaluators may judge it epistemically possible that checkerboard preferences are a good explanation for residential racial segregation in actual cities. For both credibility and modal-

conditional evaluation, conclusions may be different, but those from white people may be less accurate: if white scholars lack relevant background knowledge, they may judge a segregation model credible or epistemically possible when they should not.

Suppose that instead of seeing the model as offering a candidate for a main or central explanation we see it as offering a candidate for explanations based on contributing or pre-empted causes, as in the overdetermination frame. Recall that to add usefully to our menu, we need at least some reason to believe that the causal factors identified by the model are plausible candidates for causal factors in the actual world. In the context of overdetermination, gaining evidence from direct observation is challenging, as the effects are mixed with or pre-empted by effects from other causal factors. In the checkerboard case, we would need some way of gauging the likelihood that underlying checkerboard preferences are a causal factor influencing residential decisions -- that is, we would need to gauge not only whether people have in-group preferences in general, but also whether symmetrical in-group preferences specifically tracking race play a role in residential decisions in such a way that we can appropriately view them as a separable and distinct causal factor.

With respect to evaluating this likelihood, consider fit with background knowledge and intuition. We've seen Sugden's suggestion that the existence of checkerboard preferences coheres with "psychological and sociological evidence" and "common intuition and experience." Among the latter, we might find people reflecting on their own residential preferenecs: how likely would they be to move to an area where they are with people of different races to various degrees? Among the "sociological evidence" may be information gained from surveys, where people are asked to consider: would they like to live in an area that is 10 percent people of another race or ethnicity? What about 30 percent? (see e. g. Clark 1991).

However, a complexity in interpreting such information is that preferences and choices are themselves shaped by multiple contextual social factors -- including factors related to racism. It is therefore unclear to what extent information about stated preferences supports seeing innocuous symmetrical in-group preferences as a distinct separable causal factor from preferences shaped by racism and other cultural influences. For example, if a person is reluctant to move to a neighborhood with neighbors of a different race because they expect to be the target of racism there, they may answer survey questions based on this reluctance, but this would not be correctly interpreted as evidence of checkerboard preferences. Likewise, if a white person is reluctant to move to a neighborhood with neighbors of a different race because of racist beliefs about Black people, they may answer survey questions based on this reluctance, but this would not be correctly interpreted as evidence of checkerboard preferences. If a Black person wants to move to an integrated neighborhood because doing so is the only way to access social or material resources, they may answer survey questions based on this desire, but this would not be correctly interpreted as a preference for living in a more integrated neighborhood per se.

In light of this complexity, social situation may affect how people assess support from intuition and fit with background knowledge. A person with experience of racism may interpret sociological evidence about residential preferences in light of that experience, inferring that the preferences expressed in survey results are informed by the effects of racism and other cultural factors rather than underlying checkerboard preferences. In that case, background knowledge and intuition may provide little support to the plausibility of checkerboard preferences as distinct causal factors and little support for the usefulness of identifying checkerboard preferences as such a factor. A person less knowledgeable about racism may interpret sociological evidence about residential preferences as reflecting expression of mild in-group preferences similar to

ones they have observed in other contexts. This shows how fit with background knowledge may be variously interpreted, informing different intuitions, where background knowledge is relevant to both credibility and modal-conditional evaluation for EpHPEs. Insofar as marginalized people are more likely to make correct inferences about whether data reflecting their preferences fits the checkerboard template, their perspective will be important to an accurate overall assessment. This discussion illustrates how social situation may inform the conclusions we draw from using intuition, imagination, and fit with background knowledge when using minimal models for candidate explanations.

Let's turn now to the second, "predictive counterfactuals" interpretation. Here we take the model to tell us what would happen under certain changes or interventions to the actual world: e. g., if racist preferences and institutional discrimination were to disappear from actual cities, segregation would persist. To evaluate the model in this interpretation, we would have to consider what it would be like to live in a society where other causes of segregation have been eliminated and to consider whether checkerboard preferences would cause segregation in such circumstances. We may evaluate such counterfactual reasoning either through the lens of overdetermination via credibility or through a modal-conditional lens.

For the lens of overdetermination, we would need to use credibility to gauge not only 1) whether underlying checkerboard preferences form a distinct separable causal factor (as in the previous discussion) but also 2) whether checkerboard preferences would influence our residential choices in the actual world in radically changed future circumstances. For modal-conditional evaluation, we would use our imagination and background knowledge to imagine what it would be like to live in a world that is like the actual world except that some factors have been removed; here, to know whether checkerboard preferences would cause segregation in such

circumstances we would need (again, as with overdetermination) to evaluate 2) whether checkerboard preferences would influence our residential choices in the actual world in radically changed future circumstances.

We've seen already the complexity and social contextuality of 1) in the discussion of underlying causal factors above. I argue that 2) may also elicit judgments varying with social situation. To evaluate 2) requires assessing what kind of in-group residential preferences might persist or arise in a highly changed world -- one in which causal factors related to racism have disappeared. Here we may wonder: in such circumstances, would in-group features track race? Or might they shift to other factors such as socio-economic status, political affiliation, or culture? It is obviously unclear what is meant by saying "causal factors related to racism have disappeared," but even setting that difficulty aside, the conditions under which in-group residential preferences would continue to track race in changed circumstances is difficult to determine and is a matter about which people may come to different judgments contextualized by social situation.

For example, to those who see racial distinctions and attitudes as relatively stable facts, it may seem likely that in-group preferences would continue to track race in a variety of circumstances, so the model may seem credible for predictive counterfactuals. But to those who understand the existence of racial categories as emerging from the socio-historical imperialism of European countries, there may be reason for skepticism: if our sensitivity to racial categories is a by-product of racism, then the relevance of racial categories as in-group sorting categories for residential choices is contextual and contingent, so in-group preferences may shift to other features. In yet another perspective, a person with experience of Black culture may perceive race and culture as deeply entangled. Drawing on the writings of Du Bois, Jeffers (2013) suggests that

one reason for Black preferences for self-segregation is the fostering and protection of Black culture; insofar as self-segregation preferences are tied to community and culture in this way, such preferences may persist and continue to track race even in different social circumstances -- not because of symmetrical in-group checkerboard preferences, but for other reasons.

More generally, the example shows that using minimal models to make counterfactual and "what-if" judgements about the future of the actual world requires using intuition, imagination, and background knowledge in ways that may be influenced by social situation and theoretical perspectives. For predictive counterfactuals, the centrality of imagination shows the relevance of social situation, as people with different experiences, etc. may imagine circumstances and conditional dependencies in possible worlds differently.

The third interpretation is one in which the model is seen as informing us about a possible world that is not the actual world. Here, by contrast to the previous "predictive counterfactual" interpretation, instead of imagining changes to the actual world we imagine abstract possibilities that may not be our world. We do not need to gauge whether causal factors such as checkerboard preferences are present in the actual world: the inference is that if they were, certain effects would result.

In this possible world interpretation, one use of the model may be generating candidate explanations to add to our menu for the actual world, in which case we are back to the discussion of candidates and EpHPEs above. But a second use is interpreting the modality of "could" in conclusions like "If individuals had not strong discriminatory preferences, then residential segregation could still occur (Verreault-Julien 2019, 12). In this second use, I claim we may come to different judgments about what is possible in what modalities and whether the explanations in the model are only logically possible or possible in some more substantive way.

A model's internal coherence may suffice for it to present a logical possibility, so that social situation plays a less obvious role in evaluation. If we are only inferring that in a logically possible world with possible beings who may not even be human, segregation of some kind might be correctly explained as due to checkerboard-like preferences, our intuitions and background knowledge about the actual world may not play much of a role, as there are many ways of imagining that will instantiate logically possibility. But logical possibility alone tells us little about whether and how the model may be useful. If the "could" is merely logical possibility for the checkerboard model applied, this is compatible with highly unlikely causal factors such as randomness or even paranormal factors leading to segregation in the actual world.

To be informative, we may appeal to some other form of possibility. For sociological possibility, the inference would be that residential segregation would be aptly explained as due to checkerboard preferences rather than racism in a world that is sociologically possible for US cities. To use modal-conditional evaluation here, we may ask ourselves whether there are possible worlds sharing our sociological generalities but in which various contingencies are such that segregation there is caused by checkerboard preferences only. For example, perhaps in such a possible world, a social generality is that people have checkerboard preferences, but historically, some contingent factors leading to white supremacy never happened the way they did in the actual world, so racism as we know it never emerged. Is such a world sociologically possible? Here, as above, a person with experience and knowledge of racism may judge differently from others whether shared in-group preferences tracking race are more like sociological generalities or whether they arise from past and current contingencies in our world. Thus, social situation may inform whether we see the explanations in the model as merely logically possible or possible in a more substantive way. Here, Sjölin Wirling and Grüne-

Yanoff's caution that only "competent users of the model will do" entails the necessity of including people with knowledge and experience of racism among model evaluators. The example illustrates that using the model abstractly to evaluate what could happen in various circumstances especially depends on using our imagination in ways that may vary with experience and perspective.

In this section, I have used the checkerboard model to show how evaluation of minimal models may vary with social situation. The analogies with fiction and fables from section one help support the claim that social situation may affect model evaluation: what counts as realistic in fiction is obviously subjective; what seems realistic to one person might not seem so to another; and social situation is likely to inform what seems realistic and what does not. Because evaluation is sensitive to social situation, a fuller picture arises when there is situational diversity in the epistemic community. Further, if social situation affects what we know and do not know, and if people experiencing oppression have knowledge others lack, then using minimal models in contexts related to causes and effects of oppression, discrimination, and racism especially requires representation from people in marginalized communities.

Discussion of the example raises the question of whether evaluation of all minimal models benefits from situational diversity and where such diversity is relevant. A full discussion is beyond the scope of this paper, but a central part of the answer is that when minimal models are applied to contexts of social phenomena, different perspectives on those social phenomena are relevant to their evaluation. It is worth noting in this context that while minimal models have been largely analyzed as part of economics methodology, economic methods are now applied to a wide range of social phenomena including family life, crime, discrimination, etc. Where intuitions, imagination, and background knowledge have to do with such social phenomena,

diversity of social situation will be relevant. Given the wide-ranging effects of oppression and discrimination, perspectives from marginalized people are likely to be relevant to many modelling contexts.

3. The importance of aptly evaluating minimal models

In this section, I explore the importance of using minimal models in epistemically responsible ways. Such responsibility is especially important in contexts as complex as residential racial segregation, where there are not only causal complexities but also a variety of normative perspectives. Normatively, it's notable that while some urge the positive value of racial integration, others highlight the reasons people of color may be better off self-segregating and the potential costs for them of integration. For example, Anderson's (2010) views on the imperative of integration highlight the desirability of racial integration, arguing that it brings benefits both to white people and to people of color. In response, Jeffers (2013) argues that insofar as self-segregation may help foster Black culture, any need for integration to gain resources and end stigma-based discrimination should be seen as "tragic" rather than constructively hopeful, and James (2013) urges attention to the negative "burdens of integration" -- the "bulk of which" would fall on Black people.

There are many reasons it's important to use minimal models responsibly; here I focus just on the possibility that wrongly decreasing confidence in hypotheses we have good evidence for may lead to harmful ignorance. "Agnotology" can refer to the production of ignorance through the sowing of doubt about established theories, undermining useful knowledge (Proctor and Schiebinger 2008). For example, tobacco companies were accused of agnotology when they insisted that despite overwhelming evidence that smoking causes cancer, the conclusion wasn't

certain, because there were other possible explanations and it was a case of needing more evidence. Similarly, climate change deniers, including those in the fossil fuel industry, are seen as using agnotology when they argue that we don't have sufficient evidence that human activity is causing climate change.

I use the checkerboard model to illustrate that the use of minimal models has the potential to create or perpetuate ignorance. Recall from section 2 that epistemologists of ignorance stress that social situation can inform not only what we know but also what we do not know, where white ignorance of racism is a central example. Characterizing white ignorance in terms of patterns of ignorance that "systematically emerge from our social practices and are importantly related to the persistence of racial inequality," Martín (2021) argues that white ignorance is "of great practical and moral concern because it has bad consequences: it plays a role in sustaining racial injustice (865-866)."

One way that things can go wrong in our use of minimal models is that a model may be evaluated as apt for a particular interpretation when it shouldn't be. For example, suppose a model is wrongly seen as giving candidates for main causal factor explanations when it only gives candidates for contributing causes or possible world inferences. If we have good evidence for the main causal factors in the actual world, then decreasing our confidence in those explanations by misusing the model could add to our ignorance. In the checkerboard case, we know that the main causes of residential segregation are related to racism and white supremacy, so if we mistakenly take the model to show that checkerboard preferences should be considered as an alternative explanation for a main cause, this would wrongly undermine confidence in our evidence for racism's effects in our world. The result would foster ignorance about racism, thus, as Martín says, helping to sustain racial injustice. Insofar as including diverse perspectives and

those of marginalized people will make epistemological evaluation of minimal models more accurate, the less likely it is that using minimal models will lead to this kind of harm.

There is also a risk of harm from miscommunication about a model's correct interpretation. While scholars using the checkerboard model do not typically interpret it in light of offering candidates for main causes in the actual world, many statements about the model could be understood this way, which would cause harmful ignorance. We've seen Grüne-Yanoff's suggestion that before the checkerboard model's introduction, "many people believed that segregation was necessarily a consequence of explicitly racist preferences" (2009, 96), so the model "forced people to change their confidence in the racism hypothesis" (Grüne-Yanoff 2009, 96)." In context, this passage refers to other "plausible settings;" it is thus a modal claim, and not a claim about segregation in the actual world or what is epistemically possible as a candidate explanation. However, taken out of context it could be interpreted as a claim about what is a candidate explanation for the main causes of actual segregation. Such a conclusion would wrongly undermine knowledge we have about racism in actual cities, thus producing ignorance and harm. Likewise, consider the conclusion that in the absence of racism, "segregation could still occur." Those who draw this conclusion may mean that in some modalities, it is possible for segregation to occur in certain circumstances. But if such a statement were interpreted as a predictive counterfactual, it could be seen as supporting the claim that checkerboard preferences are among the the important causes of segregation in the actual world, which could be harmful in undermining knowledge about racism and its effects. In this context, epistemic participation of those alert to the potential harms of miscommunication could help to make those harms less likely. The discussion above in section one shows the complexity of the relationship between epistemic evaluation of minimal models and how the models are used. This complexity suggests

avoiding miscommunication may require careful attention. Insofar as social situation affects what we know and do not know and those experiencing oppression have knowledge others lack, a diverse epistemic community may be better equipped to ensure apt communication about a model's appropriate interpretation.

Finally, consider that in the cluster approach, Ylikoski and Ayindonat (2014) argue that absent empirical evidence, the importance of models for schemes is that they change the "evidential landscape" by giving us new possibilities that have to be ruled out. An explanation is only well-supported when all other explanations have been examined and ruled out. So in the checkerboard case, the model "made the requirements for an acceptable explanation of a segregation process much more stringent." After Schelling," they say, "many social scientists are no longer satisfied with purely verbal explanations of segregation. Whatever the suggested mechanism, social scientists inspired by Schelling demand that a detailed dynamic model with explicitly articulated underlying individual mechanisms is provided (2014, 31)."

If the explanations suggested by the modal are only logically possible (ObHPEs), these conclusions about actual segregation would not follow. As Fumagalli says, in the absence of similarity or other relation to the actual world, the checkerboard models lead only to justified changes in confidence in hypotheses about processes in the model worlds, not the actual world (2015, 805). This conclusion fits with Ylikoski and Aydinonat's presentation of how clusters of models answer "what-if" questions, as their questions concern what would happen in checkerboard-like models in which we vary input such as population size and sizes of neighborhoods. However, if the models in a cluster were wrongly interpreted as relevant to the actual world when they suggest only logically possible explanations, any decrease in confidence in the explanations we have would be mistaken and could lead to harmful ignorance. Here, again,

to know whether the explanations are more than logically possible requires the evaluations described above, with all the implications for diversity.[7]

Overall, since the use of minimal models can entail to epistemic risks, especially when their credibility, modality, or similarity to the actual world is misjudged, diversity in the community of epistemic knowers is important to minimize the possibility of such harm.

Conclusion

I've used the checkerboard model to explore the ways that social situation may inform minimal model evaluation. Situational diversity in epistemic communities may thus lead to a useful range of perspectives. Furthermore, as those from marginalized groups have knowledge others lack, their perspective may be necessary, especially in contexts such as that of segregation where relevant causal factors are related to oppression. I've also argued that since the use of minimal models may have epistemic risks, responsible evaluation is important. While I have suggested some reason to believe situational diversity would be broadly relevant, analysis of the general conditions under which diversity is necessary for minimal model evaluation is a subject for future research.

[7] When models are used as argumentative devices, these epistemic risks will have further implications for the broader context (see Aydinonat, Reijula, and Ylikoski 2021). For more on how the introduction of alternative explanations can generate harmful ignorance, see Fernández Pinto and Leuschner (2021 and 2022).

References

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

Aydinonat, N. Emrah, Samuli Reijula, and Petri Ylikoski. 2021. "Argumentative Landscapes: The Function of Models in Social Epistemology." *Synthese* 199, no. 1-2: 369-395.

Bokulich, Alisa. 2014. How the Tiger Bush Got its Stripes: 'How Possibly' vs 'How Actually' Model Explanations." *The Monist* 97, no. 3: 321-338.

Clark, William. 1991. "Residential Preferences and Neighborhood Racial Segregation: A Test of the Schelling Segregation Model." *Demography* 28: 1-19.

Collins, Patricia. 1990. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment.* New York: Routledge.

Collins, Patricia Hill. 1991. "Learning from the Outsider Within." In *Beyond Methodology: Feminist Scholarship as Lived Research*, edited by Mary Margaret Fonow and Judith A. Cook, 35-39. Bloomington, IN: Indiana University Press.

Ellen, Ingrid, and Justin Steil, eds. 2019. *The Dream Revisited: Contemporary Debates about Housing, Segregation, and Opportunity.* Columbia University Press.

Fehr, Carla. "What Is in It for Me? The Benefits of Diversity in Scientific Communities." *Feminist Epistemology and Philosophy of Science: Power in Knowledge* (2011): 133-155.

Fumagalli, Roberto. 2015. "No Learning from Minimal Models." *Philosophy of Science* 82, no. 5: 798-809.

Grasswick, Heidi. 2018. "Feminist Social Epistemology", *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2018/entries/feminist-social-epistemology/>.

Grüne-Yanoff, Till. 2009. "Learning from Minimal Economic Models." *Erkenntnis*, 70(1), 81–99.

Grüne-Yanoff, Till, and Caterina Marchionni. 2018. "Modeling Model Selection in Model Pluralism." *Journal of Economic Methodology* 25, no. 3 (2018): 265-275.

Grüne-Yanoff, Till, and Philippe Verreault-Julien. 2021. "How-Possibly Explanations in Economics: Anything Goes?" *Journal of Economic Methodology* 28, no. 1: 114-123.

James, V. Denise. 2013. "The Burdens of Integration": in Fall 2013 Symposium on Gender, Race and Philosophy: Commentaries on Elizabeth S. Anderson, *The Imperative of Integration*. At https://sgrponline.com/symposia/#f13

Jeffers, Chike. 2013. "Anderson on Multiculturalism and Blackness: A Du Boisian Response." in Fall 2013 Symposium on Gender, Race and Philosophy: Commentaries on Elizabeth S. Anderson, *The Imperative of Integration*. At https://sgrponline.com/symposia/#f13

Lebron Chris. 2013. *The Color of Our Shame: Race and Justice in Our Time.* Oxford: Oxford University Press.

Leuschner, Anna, and Manuela Fernández Pinto. "How Dissent on Gender Bias in Academia Affects Science and Society: Learning from the Case of Climate Change Denial." *Philosophy of Science* 88, no. 4 (2021): 573-593.

Leuschner, Anna, and Manuela Fernandez Pinto. "Exploring the Limits of Dissent: the Case of Shooting Bias." *Synthese* 200, no. 4 (2022): 326.

Martín, Annette. 2021. "What is White Ignorance?" *The Philosophical Quarterly* 71, no. 4.s

Medina, José, 2016. "Ignorance and Racial Insensitivity," in *The Epistemic Dimensions of Ignorance*, Rik Peels and Martijn Blaauw (ed.), Cambridge: Cambridge University Press, 178-201.

Mills, Charles. 2007. "White Ignorance." In Sullivan and Tuana eds., *Race and Epistemologies of Ignorance,* 26-31.

Mills, Charles W. 2014. *The Racial Contract*. Cornell University Press.

Mireles-Flores, Luis. 2018. "Recent Trends in Economic Methodology: A Literature Review." In *Research in the History of Economic Thought and Methodology: Including a Symposium on Bruce Caldwell's Beyond Positivism after 35 Years*, pp. 93-126. Emerald Publishing Limited.

Nguyen, James. 2020. "It's Not a Game: Accurate Representation with Toy Models." *The British Journal for the Philosophy of Science*.

Proctor, Robert N., and Londa Schiebinger. 2008. *Agnotology: The Making and Unmaking of Ignorance*. Stanford University Press.

Reutlinger, Alexander, Dominik Hangleiter, and Stephan Hartmann. 2018. "Understanding (with) Toy Models." *The British Journal for the Philosophy of Science* (2018).

Rodrik, Dani. *Economics Rules: The Rights and Wrongs of the Dismal Science*. WW Norton & Company.

Rolin, Kristina. 2006. "The Bias Paradox in Feminist Standpoint Epistemology." *Episteme* 3, no. 1-2: 125-136.

Sjölin Wirling, Ylwa and Grüne-Yanoff, Till. 2021. "The Epistemology of Modal Modeling. *Philosophy Compass*, 16(10), e12775.

Sjölin Wirling, Ylwa. 2021. "Is Credibility a Guide to Possibility? A Challenge for Toy Models in Science." *Analysis* 81, no. 3: 470-478.

Sugden, Robert. 2000. "Credible Worlds: The Status of Theoretical Models in Economics." *Journal of Economic Methodology* 7, no. 1: 1-31.

Sugden, Robert. 2013. "How Fictional Accounts Can Explain." *Journal of Economic Methodology* 20, no. 3: 237-243.

Sullivan, Shannon, and Nancy Tuana, eds. 2007. *Race and Epistemologies of Ignorance.* SUNY Press.

Tan, Peter. 2022. "Two Epistemological Challenges Regarding Hypothetical Modeling." *Synthese*, *200*(6), 448.

Verreault-Julien, Philippe. 2017. "Non-Causal Understanding with Economic Models: the Case of General Equilibrium." *Journal of Economic Methodology*, *24*(3), 297-317.

Verreault-Julien, Philippe. 2019. "Understanding Does not Depend on (Causal) Explanation." *European Journal for Philosophy of Science* 9, no. 2.

Verreault-Julien, Philippe. 2023. "Toy Models, Dispositions, and the Power to Explain." *Synthese* 201, no. 5.

Wylie, Alison. 2012. "Feminist Philosophy of Science: Standpoint Matters." Presidential Address delivered at the Eighty-Sixth Annual Meeting of the Pacific Division of the American Philosophical Association.

Ylikoski, Petri, and Emrah Aydinonat. 2014. "Understanding with Theoretical Models." *Journal of Economic Methodology*, 21 (1), 19–36.