Patricia Marino June 2026

Idealized Models, Counterfactual Reasoning, and Inductive Risk in Social Science Contexts: The Case of Minimal Models for Inequity

- This project applies concepts of inductive risk to uses of simple, highly idealized models in social science contexts.
- Some simple models are considered useful even though they are highly idealized, lack direct empirical support, and are not embedded in a well-confirmed background theory.
- Such "theoretical" are said to be useful in a range of ways: generating how-possibly explanations (HPEs); adding to our "menu" of possible explanations and to help us engage in counter-factual reasoning; identifying the "minimal" conditions under which an effect will occur (Ylikoski and Aydinonat 2014; Verreault-Julien 2017; O'Connor 2019; Šešelja 2023).
- For example, bargaining models are used to show that inequality can arise even in the absence of intentional discrimination: e. g., when people bargain across groups, those in smaller groups end up worse off, leading to minority disadvantage.
- I focus here on one specific use for theoretical models: that of drawing counterfactual inferences about the results of interventions (intervention-prediction inferences).
- Epistemological evaluation of such inferences faces various challenges.
- Using the example of minimal conditions for inequity, I explore how consideration of inductive risk is relevant for counterfactual inferences, and I show complexities for evaluating support for such inferences.

Idealized models and counterfactual inferences: the case of minimal conditions for inequity

- "Theoretical models" are simple, abstract, highly idealized, and "autonomous" in the sense of not being embedded in a background theory.
- Equilibrium models used to establish the fundamental theorems of welfare economics: though assumptions are false, taken to establish a "mathematical HPE" (Verreault-Julien 2017).
- In analyses of scientific practices, agent-based models based on the bandit game are said to help show the counterfactual "minimal sets of conditions which need to hold for social diversity to be epistemically beneficial" (Šešelja 2023).
- The checkerboard model due to Sakoda and Schelling often seen as showing that racial residential segregation can arise in the absence of racism and institutional discrimination
- Focus here on bargaining models for inferences about counterfactual conditions for inequity.
- In the work of O'Connor, Mohseni, Bruner, Wu, Rubin and others, bargaining and evolutionary models are used to show that inequity can arise even in the absence of direct discrimination.
- O'Connor (2019) argues that modelling can show how social categories such as gender may form because they enable efficient responses to coordination problems, then also show how such categorizations may facilitate divisional inequity.
- The"cultural Red King" effect is that when people bargain across groups, those in smaller groups can end up worse off, simply because of group size (O'Connor 2017; Bruner 2019). This effect is based on games like the Nash Demand Game (NDG).
- In the actual world, a wide range of "thicker" factors including implicit bias, explicit bias, stereotype threat, and other causes related to direct oppression contribute to inequity.

- While thick factors may form the most important current factors explaining inequity, modelling shows that very minimal conditions related to bargaining can rise rise to divisional inequity in the absence of thicker factors (O'Connor 2017, 156).
- Can be used for HPEs in the sense of adding to our "menu." But focus here is intervention prediction inferences.
- We learn that "very bare bones assumptions" about interactions can lead to situations in which minority groups are disadvantaged; in the absence of "thicker" factors, inequity may persist.
- Useful in a range of ways, including helping to see the limits of pro-egalitarian interventions: e. g., if we intervene on implicit bias, we should not expect this to "fully solve our problem." Some changes may increase equality more effectively: collective action shocks, "moral preferences" and "other-regarding preferences," protest options such as using violence to disrupt normal social function (O'Connor 2019). But because of the factors indicated by the models, "random shifts can easily carry the population back to inequity."

2. Challenges for supporting counterfactual inferences about interventions

- Epistemological evaluation of intervention-prediction inferences is complicated.
- Insofar as the these models do not claim to model the most important actual causes of phenomena, it is not easy to test empirically correspondence to actual mechanisms.
- As Ylikoski and Aydinonat (2014) explain, one possibility is an "overdetermination" frame: the model may identify causes that are contributing causes or overridden. But determining how contributing or overridden cause functions in a complex array of intersecting interdependent causes typically requires some kind of controlled experiments or background theory.
- One mode of evaluation centers Sugden's (2000) idea of "credibility": 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.
- Sugden's analogy between credible models and realistic novels "emphasizes the role of the imagination (Sjölin Wirling and Grüne-Yanoff (2021).
- Another mode of evaluation focuses on modality: here, "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" (Sjölin Wirling and Grüne-Yanoff 2021). Grüne-Yanoff (2009) suggests that the conditional judgments used to evaluate HPEs are "driven by empathy, understanding, and intuition."
- Elements such as intuition, imagination, and empathy show that epistemic evaluation of idealized models may be subjective, variable, and uncertain (Marino 2025; see also Tan 2022).
- As Feiten (2023) points out, for minimal conditions, there are complexities assessing the structural similarity between the model and its target and in assessing "which human practices besides NDGs can be usefully modelled by NDG."
- When results are supported by empirical evidence such as behavior in laboratory settings, it is difficult to know the extent to which the laboratory mechanism plays a role in current practices and also in knowing what other factors and disturbing causes may complicate extrapolation to the complex settings of the actual world (see Nguyen 2020). Example of "moral preferences."
- Modellers are often careful to express their results in ways that acknowledge epistemic uncertainty, speaking of results that "may inform real world processes" (Wu 2023) and

concluding "it is plausible that the processes observed in these models can tell us something about what really happens when groups learn to divide resources" (O'Connor 2019).

3. Inductive risk applied to the context of minimal models for inequity

- I argue that these epistemic challenges underscore the importance of inductive risk assessments in evaluating uses of theoretical models.
- Inductive risk refers to the idea that deciding whether our evidence is strong enough for a given conclusion may depend on evaluation of the consequences of getting it wrong in one way or the other (Douglas 2000). We may ask: what are the potential consequences of under- and over-confidence in drawing the above counterfactual inferences using theoretical models?
- On the one hand, knowing the minimal conditions for an unwelcome phenomenon is useful: as O'Connor (2019) argues, if "the forces of cultural evolution can pull populations toward inequity" then it is useful to learn that "combatting these forces requires constant vigilance." It would follow that even if our pro-egalitarian interventions help, "equity is a state we must keep seeking in an ever-evolving process that naturally generates inequities" (O'Connor 2019).
- Based on this, we see that the disadvantages of under-confidence are substantive.
- We may usefully learn 1) that particular forms of change are essential and 2) that we cannot plan to achieve equity through interventions and then go back to our standard practices.
- A further considerations is that 3) we may learn not to wrongly classify some interventions as failures when they do not bring about equity, because we fail to realize they can be carried out effectively yet remain limited in their effects.
- However, I argue that there are also potential disadvantages to overconfidence. Discussing the work of Piketty and others, historian Eli Cook (2020) says that to naturalize inequality is to see it as "mostly determined by natural or quasi-natural laws" and to emphasize its "inevitability."
- If we view the conditions of the model as realistic, we must be committed to thinking that people regularly engage in bargaining along the lines above; then to say that divisional inequality arises in the models is to frame it with a kind of inevitability.
- O'Connor's discussion emphasizes the ways minimality is linked to seeing the forces toward inequality as powerful forces, difficult to counteract. She says the models show how inequity emerges from "processes driven by the basic structures of our social situation -- structures that are themselves hard to do away with" so that our fixes will be generally "temporary."
- So we may interpret over-confidence in the given use of models as a form of overly naturalizing inequality.
- Cook traces various connections between recent framings of inequality, the way those framings see inequality as natural, inevitable, or difficult to resist, and "economic fatalism" in which counteracting inequality is seen as useless, pointless, or too costly.
- Cook cites a 2014 special issue of *Science* on "The Science of Inequality" featuring work by a range of economists, psychologists, archaeologists, and physicists -- rest on or are inspired by Piketty's research, and especially his well-known conclusion that $r > g$ -- that is, the rate of return on capital is greater than economic growth, ensuring that very high incomes will eventually perpetuate into very large fortunes, making economic inequality inevitable.

- Cook: "If most Americans are eventually led to believe that the enormous disparities of wealth in their society are the product of natural and inexorable laws, the chances that they will demand social, institutional, or political change will likely decline dramatically."
- Walter Scheidel on his own book *The Great Leveler:* "mass violence and catastrophes [are] the only forces that can seriously decrease economic inequality." Review in the NY *Times*: "Such is the political message that natural inequality creates: there is no point pushing for a Democratic Party that is far more progressive and egalitarian than Barack Obama's fairly neoliberal administration. The die has been cast" - Eduardo Porter.
- How we evaluate these risk for an inductive risk assessment depends on a range of epistemic factors and risk estimates, but also on our ethical and social values.
- Epistemically, we may be uncertain how likely any of the effects in question will occur or be relevant to our particular situation.
- Ethical and political values will also inform how we judge the various risks associated with over and under confidence: a Nozikian libertarian who sees ethical value in the kind of "voluntary exchanges" libertarians center may regard the economic fatalism of this section as neutral or even good: if inequality arises out of a set-up deemed to be itself appropriate, then the sooner we give up on efforts toward greater equality, the better for everyone.
- How we judge these consequences depends on our risk estimates and our ethical and political values, showing further epistemic complexities of using highly idealized, theoretical models.
- Overall then, for intervention-prediction inferences from models for inequity, we face 1) the challenge of figuring out how much epistemic support we have for the inferences in question and also 2) the challenge of figuring out how much epistemic support would be needed to form a strong enough reason in the relevant context, given the complexity of non-epistemic values and inductive risk.
- Among uses of theoretical models, drawing intervention-prediction inferences is especially affected by the factors above as these inferences are public communication (see John 2015).
- "Representational risk" is the risk associated with making representational decisions, especially in the context of modelling (Harvard and Winsburg 2022).
- Theoretical models are often used in a wide range of ways not only by modellers but also by *users* of models: does the burden of representational risk for theoretical models lie more with users than with modellers?

Conclusions

- Overall, I've argued that for minimal models for inequity, complexities of inductive risk make it unclear when having some epistemic support for inferences about interventions is enough.
- My discussion points toward the potential for further study applying concepts of representational risk to the different uses and users of theoretical models.

References

Bruner, Justin P. 2019. "Minority (Dis) Advantage in Population Games." *Synthese* 196: 413-427.

Cho, Adrian. 2014. "Physicists Say It's Simple." *Science* 344: 828.

Cook, Eli. 2020. "Naturalizing Inequality: the Problem of Economic Fatalism in the Age of Piketty." *Capitalism: A Journal of History and Economics* 1: 338-378.

Douglas, Heather E. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press.

Feiten, Tim Elmo. 2023. "The Map/Territory Relationship in Game-Theoretic Modeling of Cultural Evolution." *Philosophy of Science* 90: 1427-1436.

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

Grüne-Yanoff, Till, and Caterina Marchionni. 2026. "Multiple Routes to Progress in Model-Based Economics." *Philosophy of Science* 93: 65-82.

Harvard, Stephanie, and Eric Winsberg. 2022. "The Epistemic Risk in Representation." *Kennedy Institute of Ethics Journal* 32: 1-31.

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.* https://sgrponline.com/symposia/#f13

John, Stephen. 2015. "Inductive Risk and the Contexts of Communication." *Synthese* 192: 79-96.

Marino, Patricia. 2025. "Minimal Models, Feminist Epistemology, and Diversity." *Journal of Economic Methodology*: 1-15.

Mohseni, Aydin, Cailin O'Connor, and Hannah Rubin. 2021. "On the Emergence of Minority Disadvantage: Testing the Cultural Red King Hypothesis." *Synthese* 198: 5599-5621.

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

Northcott, Robert, and Anna Alexandrova. "Prisoner's Dilemma Doesn't Explain Much." *The Prisoner's Dilemma* (2015): 64-84.

O'Connor, Cailin. 2017. "The Cultural Red King Effect." *The Journal of Mathematical Sociology* 41, no. 3: 155-171.

O'Connor, Cailin. 2019. *The Origins of Unfairness: Social Categories and Cultural Evolution.* Oxford University Press (UK), 2019.

Parker, Wendy S. 2020. "Model Evaluation: An Adequacy-For-Purpose View." *Philosophy of Science* 87, no. 3 (2020): 457-477.

Porter, Eduardo. 2016. "A Dilemma for Humanity: Stark Inequality or Total War." NewYork *Times* (December 6, 2016).

Reutlinger, Alexander, Dominik Hangleiter, and Stephan Hartmann. 2018."Understanding (With) Toy Models." *The British Journal for the Philosophy of Science* 69: 1069-1099.

Rubin, Hannah, and Cailin O'Connor. "Discrimination and Collaboration in Science." *Philosophy of Science* 85, no. 3 (2018): 380-402.

Scheidel, Walter. 2017. *The Great Leveler: Violence and the History of Inequality from the Stone Age to the Twenty-First Century*, Princeton: Princeton University Press, 2017.

Šešelja, Dunja. 2023. "Agent-Based Modeling in the Philosophy of Science", *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/win2023/entries/agent-modeling-philscience/>. https://plato.stanford.edu/entries/agent-modeling-philscience/#EpisAgenBaseMode

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

Sugden, Robert. 2000. "Credible Worlds: The Status of Theoretical Models in Economics." *Journal of Economic Methodology* 7: 1-31.

Sugden, Robert. 2013. "How Fictional Accounts Can Explain." *Journal of Economic Methodology* 20: 237-243.

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.

Wu, Jingyi. 2023. "Epistemic Advantage on the Margin: A Network Standpoint Epistemology." *Philosophy and Phenomenological Research* 106: 755-777.

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