Modeling Minimal Conditions for Social Ills

Cailin O'Connor

Abstract:

There have been many worries within philosophy about the epistemic role of simplified mathematical models such as those used in economics and the other social sciences. In particular, many have wondered how and whether such models can illuminate complex phenomena related to human societies. This paper will discuss a class of idealized models that outline what I call "minimal conditions for social ills". But are such models useful to thinking about these complex factors? And, if so, how? Some such models will track actual causal factors that generate real world problems. Others may or may not. Whether or not these models do track these real-world factors is irrelevant to the role they play in showing that minimal, realistic factors are enough to generate social problems. In doing this they provide important counterfactual information. Investigations of social ills are often aimed at interventions to stop them. Given this it is important to know: if we intervene on the current causes of some social problem, what other common social factors might continue to contribute to it?

I: Introduction

This paper will discuss a class of idealized models that outline what I call "minimal conditions for social ills". These models can play an important role in reasoning that might not be initially obvious. Some such models will track actual causal factors that generate real world problems.[1] Others may or may not. Whether or not these models do track these real-world factors is irrelevant to the role they play in showing that minimal, realistic factors are enough to generate social problems. In doing so they provide important counterfactual information. Investigations of social ills are often aimed at interventions to stop them. Given this it is important to know: if we intervene on the current causes of some social problem, what other common social factors might continue to contribute to it? Notably, in these cases it is the fact that the models do not fit the world that makes them particularly epistemically useful. Empirical investigations cannot typically remove current causes of social problems to see what other, counterfactual causes may continue to act after intervention. Models make such investigations relatively easy.

The paper will begin by examining several, related models of the cultural evolution of inequity. This set of models looks at how conventions of bargaining and resource division emerge between social groups. Axtell et al. (2001) present an early model

---

[1] I will not say much in this paper about what is meant by causal factors or causal pathways, since this is orthogonal to the investigation at hand. The interventionist account from Woodward (2003) and others is a useful one in thinking about modeling, since causal pathways are typically identified in models by intervening on key features and comparing outcomes.

showing that under very bare bones conditions inequitable conventions of this sort can evolve. Philosophers and other social scientists have expanded this exploration to consider how other features like minority status, power, intersectional effects, and social network structure influence the emergence of such conventions. These models abstract away from many other complex psychological and structural details involved in the emergence of inequitable systems. In particular, they do not represent important causes of inequity such as implicit and explicit bias. As this paper will argue, though, they still give us important counterfactual information about how little is needed to generate inequity. And this information is critical to thinking about possible interventions, such as trainings aimed at reducing racial and gender biases.

There are a number of prominent accounts of the epistemic uses of simple models like those discussed here. For example, many previous authors have considered at length the epistemic usefulness of the Schelling checkerboard model of racial segregation, which is another minimal model of the emergence of a social problem that abstracts away many realistic causes of segregation (Gibbard and Varian, 1978; Sugden, 2000; Aydinonat, 2007; Grune-Yanoff, 2009; Ylikowski and Aydinonat, 2014). This paper is not meant to supplant previous accounts of this sort. As I will briefly illustrate, there is a sort of modeling pluralism not much discussed in philosophy of modeling, which is that the same model often plays multiple epistemic roles, even within the same project. The role outlined here is one that many models play, while also contributing to understanding, or promoting inquiry in other ways.

The paper will proceed as follows. Section two describes the "emergence of classes" model introduced by Axtell et al. (2001), and further extensions of this paradigm. Section three makes the main arguments of the paper---that despite certain failures to represent, and, surprisingly, sometimes because of these very failures, the models described nonetheless provide crucial information to those interested in social interventions aimed at decreasing inequity. Section four briefly concludes.

Before continuing, a few notes about what this paper will not do. First, there is a robust literature in the philosophy of modeling on whether and how simple models *explain*.[2] While this is a very important question, this paper will instead focus on the big-tent question of what "epistemic roles" the models in question play. I.e., how are they used in scientific investigation and the creation of knowledge more generally? This is because there are many ways models can be used epistemically outside the realm of explanation (Alexandrova, 2008; Rohwer and Rice 2016). And, as noted, the models investigated here support a picture where models can be understood as playing multiple epistemic roles, some perhaps explanatory and others perhaps not, even within one investigation.

Second, this paper will not attempt to offer a general account of the workings of simple models. Instead, it will focus on one class of models and outlining one

---

[2] See, for example, Alexandrovna (2008) and Alexandrovna and Northcott (2013) who argue that simple economic models do not explain, Aydinonat (2007) who offers an account of "partial potential" explanations in models, Sugden (2000,2013) who makes the case for real explanation by models, and Bokulich (2014) who disambiguates types of explanation in simple models.

important epistemic role they can play. This is because the categories of "simple models" or "idealized models" or "economic models" are too complex and variable to be treated by one (small) account such as this one. Hopefully, though, the case outlined here will be useful to philosophers of modeling developing more general accounts.

II. Models of the Emergence of Inequitable Norms

Empirical literature has revealed that women and people of color tend to get less on average in scenarios of resource division than men and white people (Ayres and Siegelman 1995; Steinpreis et al.,1999; Bertrand and Mullainathan 2004). Axtell et al. (2001) provide an early model addressing this sort of phenomenon via the emergence of what they label "discriminatory norms".[3] Their model involves agents of two types – potentially representing social categories such as race or gender – who interact in a bargaining scenario called a Nash demand game. This involves dividing a resource (let's say of size 100), by requesting a low, medium, or high amount of it (say 30, 50, or 70).   If the demands are compatible in that they add up to 100 or less, each agent gets what they request. Otherwise, they get a poor outcome because they are jointly too aggressive to reach an agreement.

In their model, each agent remembers their last *n* interactions with the two different types of individuals, and uses these memories to decide which demand to pick. In particular, they choose the demand that would have done best against those they met in the past. Suppose the two types are "blues" and "greens". Consider an agent who recently met five greens each demanding 70, and also recently met five blues each demanding 50. The next time that agent meets a green they will demand 30, which would have done best against their memories, and the next time they meet a blue they will demand 50.

What the authors find is that simulations of this model – where agents meet each other randomly, and update their memories each time – often arrive at stable patterns where each type demands 50 of those in their own group, but between groups one side always demands 70 and the other 30. In other words, everyone treats their in-group fairly, and treats the out-group differently to the detriment of one out-group. This pattern emerges commonly despite the fact that a fair outcome is possible between the two groups, and that the two groups are completely symmetric. In other words, in explaining the emergence of inequity, there is nothing about the properties of the two groups to appeal to.

Furthermore, they find that models with a single group most often evolve to fair states where everyone demands 50 of their interactive partners.[4] In other words, the simple addition of types or social categories changes the model from one where fairness is expected to one where inequity tends to emerge.

---

[3] In designing their model, they draw on previous work by Young (1993).

[4] This is consonant with other work on the evolution of bargaining. See, for example Young (1993) and Skyrms (2014).

Philosophers and social scientists have used this sort of model, and many variations of it, as a framework for understanding the emergence of inequitable conventions, and for exploring the conditions under which one social group tends to get more than another.[5] For example, Bruner (2017) and later O'Connor (2017) show that under many circumstances the bare fact of minority status increases the likelihood that a social group will end up being discriminated against in these models. They call this the cultural Red King effect after an analogous effect in biology where a slow evolving species can gain an advantage in mutualisms.[6] (See (Bergstrom and Lachmann 2003) for the biological version of the effect.) Bruner and O'Connor (2017) and LaCroix and O'Connor (2020) also explore the effects of power on the emergence of such conventions, arguing that various sorts of empowerment can increase the likelihood that a social group will end up at an advantaged norm in terms of resource division. In O'Connor, Bright, and Bruner (2019) the authors use these effects to show that those at the intersection of two minority or disempowered groups can be especially disadvantaged in these models. In Poza et al. (2010), Gallo (2014), and Rubin and O'Connor (2018), authors explore the emergence of such conventions on networks.

A number of authors use similar models to draw similar conclusions about a wider range of interactions. For instance, authors use variations of coordination games, or hawk-dove, or other models of discrimination to show how learning groups easily move towards patterns where one group gets more and the other less (Hoffmann, 2006, Stewart, 2010, Hwang et al., 2016).

Taken as a set the models described here fit well with a picture outlined in Aydinonat (2007) and Ylikowski and Aydinonat (2014). They point out that understanding often emerges from families of models, where researchers develop many variations on some model to test robustness and explore outcomes. This set of models of inequity constitutes such a family. They all instantiate a short list of core causal factors. These are that 1) actors learn to do what is best for them, 2) actors condition their choice of behavior on the social identity of their interactive partners, and 3) actors regularly engage in bargaining scenarios. Together, then, the models show that even under many variations with respect to population structure, learning rules, etc. this small set of causal factors is enough to reliably generate inequity. Individual models in the family are used to many other epistemic purposes, but we will focus here on the general picture where inequity is shown to emerge from these highly minimal preconditions.

---

[5] For those who care about such distinctions, many of the results I now mention are from population models using the replicator dynamics or best response dynamics rather than the specific agent-based models used by Axtell et al. (2001). The work all involves the use of Nash demand games and social categories within a population to see how patterns of resource division emerge that disadvantage those in one social group.

[6] This occurs because minorities meet their out-group very commonly, while majorities meet them only rarely. As a result, minorities tend to learn more quickly how to interact with their out-group, which, in a bargaining scenario, often involves making low, safe, accommodating demands (30 rather than 50 or 70). The majority group can then slowly learn to take advantage of this accommodation.

II: Modeling Minimal Conditions for Inequity

As I will now argue, these, and similar, models play an important epistemic role in outlining minimal, realistic conditions that are sufficient to generate some social ill. To play this role, models must, 1) reproduce some social ill (like inequitable conventions) and 2) show that this outcome results from causal variables that are realistic and minimal (like conditions 1-3.) In doing this, these models show how little is needed to generate something surprisingly bad. This is especially important in thinking about social problems, where invention is a typical goal. Such models can give counterfactual information about what might happen once interventions occur, and, in particular, about why interventions might not work as planned. The rest of this section will be devoted to further elaborating how this epistemic role works and how it relates to other accounts of the workings of simple models.

First, what is meant here by calling the conditions explored by these models "minimal" and why does "minimality" matter? The models described in the last section are all high idealized, by which I mean that they ignore and alter real world features of the systems they represent.[7] No model can embody every possible modeling virtue (for example, by being maximally simple, causally transparent, and perfectly accurate), so modelers typically choose to elevate some virtues over others (Weisberg 2012). These particular models opt for causal transparency and simplicity over complexity and fit to real phenomena.

Philosophers of science have described some such models as "minimal" or "minimalistic", though what is meant by that varies across accounts. [8] One influential account comes Weisberg (2007) who describes minimalist idealization as paring away irrelevant details from real world processes to arrive at a few key, causal factors for the target phenomenon. (See also Potochnik (2007) and Strevens (2008).) Such models explain or illuminate the phenomenon at hand by virtue of properly representing the corresponding causal factors in the world, even if they do not represent other, irrelevant aspects of the real-world system.[9] Notice that models that explore minimal conditions for social ills may sometimes capture the most relevant causal factors at play in the world, and when they do they can act as

---

[7] We might draw a distinction here between models that are abstract, meaning that they ignore some features of the world, and models that are idealized, meaning that they alter these features in some way. These models are actually both.

[8] Weisberg (2007, 2012) outlines minimalist idealization as a strategy of paring models down to only the most relevant causal factors. On the other hand, Batterman (2002) and Batterman and Rice (2014) argue that "minimal" models explain by virtue of belonging to a universality class that also includes real, complex systems. On this story, the representation relation between model and system is not supposed to do the work in allowing the model to tell us something about real systems. In yet another use of the term Grüne-Yanoff (2009), describes minimal models as lacking, "any similarity, isomorphism or resemblance relation to the world" and "unconstrained by natural laws or structural identity" (83). Such models improve our understanding of the world via proofs of impossibility or necessity. (See also Knuuttila (2009).) While the terms are the same, the epistemic roles of these various models are very different.

[9] Unsurprisingly, because of the idealizations inherent in building minimal models, some have argued that they cannot, in fact, explain, or that their capacity to do so is limited in various ways, as in Fumagalli (2015, 2016), though this discussion is beyond the scope of this paper.

minimalist models in the sense of Weisberg (2007). However, they need not play this role to be epistemically useful.

This is an important observation, because many of these models fail represent current, important causes of the social ills they produce. With regards to the models of inequity from the last section, one of the most glaring representational lacunas has to do with psychological factors involved in discrimination and inequity. Implicit bias, explicit bias, stereotype threat, and confirmation bias for example, are arguably important causes of the emergence of inequity across social groups.[10]   In the models described, however, the emergence of inequity does not involve these psychological elements. But, as noted, this does not prevent them from providing counterfactual information about what might happen if current causes of inequity are intervened on.

But, we might ask, why does minimality matter? How does it allow the relevant models to play this role? First, because it is relevant for causal transparency, as noted above. Causal transparency is key in providing information about counterfactual causes. Additionally, there is another reason why minimality matters at least to some degree. To see this, it is useful compare these models to one description of the workings of "how-possibly modeling" (or "how-possibly explanation"). These terms are used in many ways in the philosophy of modeling literature (Bokulich, 2014). Sometimes "how-possibly modeling" describes models which show that some phenomenon can in principle (possibly) be generated from a set of starting conditions, often in respond to an extant impossibility claim. Let's call this sort of epistemic role HP1. To give an example, Brian Skyrms's work on the evolution of signaling refuted claims by natural language skeptics like W.V.O. Quine that linguistic meaning could not possibly emerge on its own (Skyrms 2010; W. V. O. Quine 1960). Skryms' models showed that, indeed, language can emerge spontaneously among learners. Notice that to play this sort of HP1 role, a model need not be particularly simple, and the relevant preconditions need not be minimal. The goal is just to show possibility in the face of an impossibility claim.

But if the models here showed that inequity could result, in principle, from elaborate, highly complex conditions, they would not be of much interest, because such conditions are typically unlikely to hold in the real world. Unlike natural language, it is usually not in doubt that social ills can in principle emerge on cultural evolutionary timescales as a result of different causal factors. The goal is not to show that they can. The conditions causing social problems in the models in question must be few enough that they might really be able to act as causes of social ills.[11]

---

[10] See, for example, Saul (2013) for the effects of implicit bias and stereotype threat in academic philosophy.

[11] Note that, of course, there might be some model illuminating a set of highly complex, numerous preconditions for a social ill, but where all of these are very likely to hold in the real world. Such a model could still play the role outlined in this paper. The point is that typically such a model is unlikely to be able to do so.

Related to this is the requirement that these conditions must be "realistic". Brett Calcott has described what he calls "how-MacGyvery" models, which show how from extremely minimal conditions some surprising phenomenon can arise.[12] An extra requirement that goes beyond how-MacGyvery modeling is that these minimal conditions are ones we think could actually be instantiated in the world. For instance, if Axtell et al. (2001), and the many extensions discussed in section II, showed that inequitable norms could emerge among people with the surprisingly minimal condition that they dislike Nicholas Cage movies, these models would be completely irrelevant. In this the models here play something similar to the "how-possibly" roles described by authors like Brandon (1990), Forber (2010), and Bokulich (2014) who argue that how-possibly explanations outline processes that might really occur in the world (even if we are not certain they do). (We can call this HP2.)[13]

This realism requirement raises a question, though: what sorts of conditions count as realistic in models that illuminate minimal conditions for social ills? Sugden (2000) tackles this same question in thinking about how economic models explain. As he points out, realism of this sort is usually grounded in similarity judgments. The models are judged to be similar, in relevant ways, to the real-world systems they represent. As he points out, saying just what this means can be a bit tricky. But he ultimately settles on a notion of credibility as important, which approximately tracks, "the extent to which we can understand the relevant model as a description of how the world *could* be" (24). Discussing whether and how models can be successfully compared to real-world systems in this way goes well beyond the scope of this paper. I follow others, like Sugden, who argue that they can.

When the right sorts of realism and minimalism are present in these models, they arguably can give some counterfactual information about the world. Whether or not the factors in the model *are* at play in the real world, these factors are plausible candidates to act as causes in the real world. And this fact makes them relevant to thinking about interventions on inequitable systems. One thing we learn from the models of inequitable norms in section II is that removing implicit bias, stereotype threat, and confirmation bias might not be enough to fix the problem. In other words, the sorts of interventions being implemented by businesses, universities, etc. to improve inequity do not impact a set of conditions which occur in these organizations, and should be expected to generate inequity even without various biases at play.

Notice that when models outline minimal conditions for inequity, it is the very fact that they do not capture many aspects of the real world that generates their epistemic power. In other words, both their realism and their lack of realism is important. In particular, in stripping away realistic causes of inequity, they provide

---

[12] This term comes from public talks, rather than published work. For an example, see Soriano et al. (2015). This refers to the 80s and 90s TV show MacGyver, where the title character routinely built technological apparatuses from random everyday objects.

[13] This also fits well with the account of Sugden (2000) who emphasizes the "credibility" of simple models in grounding their epistemic import.

valuable information about counterfactual causes that would be hard to study otherwise. Implicit bias, or more pernicious racism, could be added to these models, and that would, in some ways, improve their fit to the world. But doing so would obscure the observation that conditions 1-3 can yield inequity on their own. In this way, the models discussed in this paper fit well with some accounts of simple models that recognize the importance of abstracting away from realistic details when identifying sufficient causes for a phenomenon (Ylikowski and Aydinonat 2014).[14]

Notably, identifying this sort of role makes clear why models, and not other forms of investigation, are so useful in this sort of case. Empirical investigations are often limited in that they cannot remove causal factors from the world. For instance, in studying the emergence of inequitable conventions in real organizations and societies, we cannot easily remove or alter aspects of human psychology in these societies. Because models are not grounded in reality, they provide a relatively cheap, easy way to investigate potentially relevant causal pathways in advance of difficult, costly interventions.

Before concluding, I'd like return to one final point. In the introduction, I argued that the epistemic role outlined in this paper is not necessarily in conflict with previous analyses of simple models. Philosophers have often recognized one kind of modeling plurality – that models come in many forms, and play many sorts of epistemic roles (Downes 1992; O'Connor and Weatherall 2016; Rohwer and Rice 2016). But there has been less attention to a second sort of model plurality – that the same model can play multiple epistemic roles even with one project. The models described in this paper illuminate this very point. Consider models of the cultural Red King effect, for instance, which show how minority status alone may lead to bargaining inequity. We can identify at least four separate, epistemic roles these models play within the same project. First, they illuminate a causal pathway that may be at play in the real world, thus acting as minimalist models in the Weisberg (2007) sense. Second, they can direct future empirical investigation on this topic. Mohseni et al. (2019) investigate the cultural Red King effect in the lab. As they describe, without the formal models, there would be no reason to test whether minority status confers a bargaining disadvantage. But they do, in fact, find such a disadvantage among experimental subjects. Third, the models add a new causal scenario to the set of possible explanations for minority disadvantage. This is a role for simple models described by Ylikowski and Aydinonat (2014).[15] And fourth they outline extremely minimal conditions for a minority group to end up disadvantaged, and in doing so provide counterfactual information in the way described here. Thus this paper, in showing that models of social ills provide

---

[14] They also are similar to models outlined by Batterman (2002) and Batterman and Rice (2014) who argue how a lack of fit to the world, or representative power, can sometimes enhance a model's ability to explain.

[15] These authors also support the point here, by identifying a laundry list of useful epistemic roles that the Schelling checkerboard model can play.

important counterfactual information relevant to interventions, does not negate or preclude other analyses of the epistemic import of these, or similar, models.

IV: Conclusion

The goal of this paper was to pull out in detail an epistemic role played by a class of models that focus on inequity. These models outline minimal conditions for inequity, while idealizing away from causal factors that are at play in the real world. Nonetheless, in identifying minimal, realistic conditions that can generate patterns of inequity, they provide important counterfactual information about the target system that is especially useful for planned interventions. In particular, they tell us that some interventions on real causes of inequity might not be successful, since there are other, realistic, counterfactual causes that might continue to generate inequity. Since much investigation into inequity focuses on intervention, this is an important epistemic contribution.

Although this paper focused on one family of models to illustrate this point, there are other notable candidates, such as the Schelling checkerboard model (Schelling, 1969; 1971). There have been a number of in-depth accounts of how the Schelling model is explanatory. (See for example, Sugden (2000), Aydinonat (2007), Grune-Yanoff (2009), Ylikowski and Aydinonat (2014).) This paper adds to these. The Schelling model is also one that plays the epistemic role laid out here. It outlines minimal conditions (preferences not to be in a small racial minority, a tendency to move when one is) for a social ill (racial segregation). It is widely recognized to ignore important causal factors that drive racial segregation in the real world (Farley, Fielding, and Krysan 1997; Galster and Godfrey 2005; Denton 2006). But it gives counterfactual information that realistic, minimal factors can cause segregation, even if we intervene on more pernicious factors like housing discrimination.[16] Relatedly, as Ylikowski and Aydinonat (2014) point out, "even if the Schelling mechanism does not bring about segregation, its presence has important counterfactual consequences, as it might cause segregation in the absence of other causes" (28).

Another set of candidate models outline what we might call minimal conditions for polarization. Polarization is often recognized as a social ill in that it prevents effective democratic functioning. A number of models have shown how simply adding to a model that agents ignore those with different beliefs tends to generate polarization (Hegselmann & Krause, 2002; Macy et al., 2003; Olsson, 2013). Likely the real world causes of polarization are much more variable, ranging from innate in-group biases, to partisan news sources, to conformity bias. But these models, again, illustrate how little is needed to generate polarization, and thus why interventions aimed at, say, in-group bias, might not be enough to eliminate it.

---

[16] Though one might respond that this model does not fall under the analysis here, because racial segregation simpliciter is not necessarily a social ill. It is only the pernicious racism involved in current segregation that we should try to intervene on. Thank you to (REMOVED FOR REVIEW) for this point.

One broader take-away from this entire responds to claims that highly idealized models cannot inform us about the world. In fact, the epistemic roles that models play are promiscuous and varied (Downes 1992; O'Connor and Weatherall 2016; Rohwer and Rice 2016). The investigation here outlines one more way that simple models can play important roles in reasoning, and contribute to epistemic progress, even if they do not perfectly match the target systems they address.

Works Cited:

Alexandrova, A. (2008). Making models count. *Philosophy of Science*, *75*(3), 383-404.

Alexandrova, A., & Northcott, R. (2013). It's just a feeling: why economic models do not explain. *Journal of Economic Methodology*, *20*(3), 262-267.

Axtell, Robert, Joshua Epstein, and H. Peyton Young. (2001). The Emergence of Classes in a Multiagent Bargaining Model. In *Social Dynamics*, 191–211. Eds. Steven Durlauf and H. Peyton Young. Cambridge, Massachusetts: MIT Press.

Aydinonat, N. E. (2007). Models, conjectures and exploration: An analysis of Schelling's checkerboard model of residential segregation. *Journal of Economic Methodology*, *14*(4), 429-454.

Ayres, Ian, and Peter Siegelman. (1995). Race and Gender Discrimination in Bargaining for a New Car. *The American Economic Review* 85 (3): 304–21.

Batterman, Robert W. (2002). Asymptotics and the Role of Minimal Models. *The British Journal for the Philosophy of Science* 53 (1): 21–38. doi:10.1093/bjps/53.1.21.

Batterman, Robert W., and Collin C. Rice. (2014). Minimal Model Explanations. *Philosophy of Science* 81 (3): 349–76. doi:10.1086/676677.

Bergstrom, Carl T., and Michael Lachmann. (2003). The Red King Effect: When the Slowest Runner Wins the Coevolutionary Race. *Proceedings of the National Academy of Sciences* 100 (2): 593–98. doi:10.1073/pnas.0134966100.

Bertrand, Marianne, and Sendhil Mullainathan. (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *The American Economic Review* 94 (4): 991–1013.

Bokulich, A. (2014). How the tiger bush got its stripes: 'how possibly' vs. 'how actually' model explanations. *The Monist*, *97*(3), 321-338.

Brandon, R. N. (1990). *Adaptation and environment*. Princeton University Press.

Bruner, Justin P. (2017). Minority (Dis)Advantage in Population Games. *Synthese*, July, 1–15. doi:10.1007/s11229-017-1487-8.

Bruner, Justin P., and Cailin O'Connor. (2017). Power, Bargaining, and Collaboration. In *Scientific Collaboration and Collective Knowledge*. Vol. Eds. T. Boyer, C. Mayo-Wilson, and M. Weisberg. Oxford, UK: Oxford University Press.

Clark, W. A. (1991). Residential preferences and neighborhood racial segregation: A test of the Schelling segregation model. *Demography*, *28*(1), 1-19.

Denton, N. A. (2006). Segregation and Discrimination in Housing. In *A Right to Housing: Foundation for a New Social Agenda*, 61–81. Eds. Rachel G. Bratt, Michael E. Stone, and Chester Hartman. Philadelphia: Temple University Press.

Downes, Stephen M. (1992). The Importance of Models in Theorizing: A Deflationary Semantic View. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1992 (1): 142–53. doi:10.1086/psaprocbienmeetp.1992.1.192750.

Farley, Reynolds, Elaine L. Fielding, and Maria Krysan. (1997). The Residential Preferences of Blacks and Whites: A Four-metropolis Analysis. *Housing Policy Debate* 8 (4): 763–800. doi:10.1080/10511482.1997.9521278.

Forber, P. (2010). Confirmation and explaining how possible. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *41*(1), 32-40.

Fumagalli, R. (2015). No Learning from Minimal Models. *Philosophy of Science* 82 (5): 798–809. doi:10.1086/683281.

———. (2016). Why We Cannot Learn from Minimal Models. *Erkenntnis* 81 (3): 433–55. doi:10.1007/s10670-015-9749-7.

Galster, George, and Erin Godfrey. (2005). By Words and Deeds: Racial Steering by Real Estate Agents in the U.S. in 2000. *Journal of the American Planning Association* 71 (3): 251–68. doi:10.1080/01944360508976697.

Gibbard, A., & Varian, H. R. (1978). Economic models. *The Journal of Philosophy*, *75*(11), 664-677.

Grüne-Yanoff, Till. (2009). Learning from Minimal Economic Models. *Erkenntnis* 70 (1): 81–99. doi:10.1007/s10670-008-9138-6.

Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of artificial societies and social simulation*, *5*(3).

Hoffmann, R. (2006). The cognitive origins of social stratification. *Computational Economics*, *28*(3), 233-249.

Hwang, S. H., Naidu, S., & Bowles, S. (2016). Social conflict and the evolution of unequal conventions. In *Santa Fe Institute Working Paper*.

Jhun, Jen, Patricia Palacios, and James O. Weatherall. (2017). Market Crashes as Critical Phenomena? Explanation, Idealization, and Universality in Econophysics. *Synthese*.

Knuuttila, Tarja. (2009). Isolating Representations Versus Credible Constructions? Economic Modelling in Theory and Practice. *Erkenntnis* 70 (1): 59–80. doi:10.1007/s10670-008-9137-7.

LaCroix, T., & O'Connor, C. (2020). Power by association. *Ergo*.

Macy, M. W., Kitts, J. A., Flache, A., & Benard, S. (2003). Polarization in dynamic networks: A Hopfield model of emergent structure.

Mohseni, A., O'Connor, C., & Rubin, H. (2019). On the emergence of minority disadvantage: testing the cultural Red King hypothesis. *Synthese*, 1-23.

O'Connor, Cailin. (2017). The Cultural Red King Effect. *The Journal of Mathematical Sociology* 0 (0): 1–17. doi:10.1080/0022250X.2017.1335723.

O'Connor, C., Bright, L. K., & Bruner, J. P. (2019). The emergence of intersectional disadvantage. *Social Epistemology*, *33*(1), 23-41.

O'Connor, C., & Bruner, J. (2019). Dynamics and diversity in epistemic communities. *Erkenntnis*, *84*(1), 101-119.

O'Connor, Cailin, and James Owen Weatherall. (2016). Black Holes, Black-Scholes, and Prairie Voles: An Essay Review of Simulation and Similarity, by Michael Weisberg. *Philosophy of Science* 83 (4): 613–26. doi:10.1086/687265.

Olsson, E. J. (2013). A Bayesian simulation model of group deliberation and polarization. In *Bayesian argumentation* (pp. 113-133). Springer, Dordrecht.

Potochnik, Angela. (2007). Optimality Modeling and Explanatory Generality. *Philosophy of Science* 74 (5): 680–91. doi:10.1086/525613.

Poza, D., Galán, J. M., Santos, J. I., & López-Paredes, A. (2010). An agent based model of the Nash demand game in regular lattices. In *International Conference on Information Technology for Balanced Automation Systems* (pp. 243-250). Springer, Berlin, Heidelberg.

Quine, Willard Van Orman. 1960. *Word and Object*. Cambridge, Massachusetts: MIT Press.

Rohwer, Y., & Rice, C. (2016). How are models and explanations related?. *Erkenntnis*, *81*(5), 1127-1148.

Rubin, H., & O'Connor, C. (2018). Discrimination and collaboration in science. *Philosophy of Science*, *85*(3), 380-402.

Saul, Jennifer. 2013. "Implicit Bias, Stereotype Threat, and Women in Philosophy." In *Women in Philosophy: What Needs to Change*, eds. Katrina Hutchinson and Fiona Jenkins, 39–60. Oxford, UK: Oxford University Press.

Schelling, Thomas C. (1969). Models of Segregation. *The American Economic Review* 59 (2): 488–93.

———. (1971). Dynamic Models of Segregation. *The Journal of Mathematical Sociology* 1 (2): 143–86. doi:10.1080/0022250X.1971.9989794.

Skyrms, Brian. 2010. *Signals: Evolution, Learning, and Information*. Oxford University Press.

Skyrms, B. (2014). *Evolution of the social contract*. Cambridge University Press.

Soriano, Miguel C., Daniel Brunner, Miguel Escalona-Morán, Claudio R. Mirasso, and Ingo Fischer. (2015). Minimal Approach to Neuro-Inspired Information Processing. *Frontiers in Computational Neuroscience* 9. doi:10.3389/fncom.2015.00068.

Steinpreis, Rhea E., Katie A. Anders, and Dawn Ritzke. (1999). The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study. *Sex Roles* 41 (7–8): 509–28. doi:10.1023/A:1018839203698.

Strevens, Michael. (2008). *Depth: An Account of Scientific Explanation*. Harvard University Press.

Sugden, R. (2000). Credible worlds: the status of theoretical models in economics. *Journal of economic methodology*, *7*(1), 1-31.

Sugden, R. (2013). How fictional accounts can explain. *Journal of Economic Methodology*, *20*(3), 237-243.

Weisberg, Michael. (2007). Three Kinds of Idealization. *The Journal of Philosophy* 104 (12): 639–59.

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

Woodward, J. (2004). *Making things happen: A theory of causal explanation*. Oxford university press.

Ylikoski, P., & Aydinonat, N. E. (2014). Understanding with theoretical models. *Journal of Economic Methodology*, *21*(1), 19-36.

Young, H. P. (1993). An Evolutionary Model of Bargaining. *Journal of Economic Theory* 59 (1): 145–68. doi:10.1006/jeth.1993.1009.