Modal Knowledge and Explanatory Power in Social Science Modeling

Sean Huang

PHIL 177

Dr. Nadia Ruiz

December 15 2023

Word Count: 2783

1. Introduction

It has been argued that models in social science have explanatory power because they provide modal knowledge. Under this view, models are not idealizations of the actual world or abstract thought experiments. Instead, they describe phenomena in coherently possible scenarios. The results of these models then prove modal hypotheses of epistemic value. However, the previous literature has not given a good defense for why modal facts about possible worlds provide knowledge about the actual state of affairs. In this paper, I introduce the use of dominance claims, wherein the social scientist infers a necessity claim from a possibility claim established by a model. I argue that this methodology provides genuine explanatory knowledge.

In section 2, I provide a background overview of the issues surrounding the modal view of social-scientific modeling, focusing on criticisms leveled at Grüne-Yanoff's (2009) theory of minimal economic models. In section 3, I introduce dominance claims and demonstrate how they provide genuine explanatory information about the real world. In section 4, I illustrate the use of dominance claims by analyzing O'Connor and Weatherall's 2018 model of misinformation spread in scientific communities. In section 4, I address an obstacle to dominance claims in the work of Julien Ross (2013) and William Wimsatt (2007) and provide suggestions for surmounting these difficulties. Section 5 explores related strategies that may be useful in social-scientific modeling work. All in all, I defend the modal view of social-scientific models and thereby demonstrate how strictly false models can have explanatory power.

2. The Modal View of Social-Scientific Models

In this section, I provide an overview of the modal view of social-scientific models. I then address two issues with the account that the rest of the paper attempts to solve.

On the modal view, social-scientific models construct a scenario in a possible world, and the claims we establish by studying the model are only true in the possible world. The modal view stands in contrast to representative views, which claim that a model can be truthfully represent some aspect of a target phenomenon despite containing many properties that diverge from the target phenomenon. Representative views demand an explanation of how good models "link up" with their real-world counterparts. For example, Cartwright argues that models are meant to function as Galilean thought experiments—they replicate the behavior of real systems in such a way that the causal factor of interest operates "without impediment" (Cartwright 1999). Therefore, an epistemically valuable model must have assumptions that are based on confirmed capacities of the cause in question. Similarly, Mäki sees models as a partial representation of a real system (Mäki, 2004). For Mäki, we can learn from a model only if the model actually resembles relevant parts of the actual system, where "relevance" depends on the isolated aspect of a system we choose to study. Common to all representative views is the insistence that models can be explanatory only if their assumptions resemble or represent aspects of the real world.

In his 2009 paper, Grüne-Yanoff motivates the search for an alternative to representative views with the observation that many epistemically valuable models do not have any similarity, resemblance, or adherence to regularities to their target phenomenon. Instead, "When economic theorists describe the process of constructing a theoretical model they stress the role of creativity, playfulness, and imagination—but they do not mention the importance of well-confirmed theoretical principles, of data analysis, or of any other well-argued link to actual situations" (Grüne-Yanoff, 2009). Of course, many models inform their model-making with empirical evidence, as in the case of work in behavioral economics. However, highly abstract and minimal models with no relationship to the world still seem to generate knowledge.

Take Schelling's famous checkerboard model of neighborhood segregation. In Schelling's model, idealized agents occupy squares on a 2-dimensional grid. Each agent belongs to a group and is "happy" if a certain percentage of her neighbors are in the same group as her. (The exact percentage is adjusted by the modeler.) At each time-step, an agent randomly relocates to another square if they are unhappy. When constructing the model, Schelling makes no effort to justify calling these agents "families" or the grid a "neighborhood." Furthermore, there does not appear to be any way in which the model environment has any substantive similarity to the dynamics of a real-world neighborhood. Historically, Schelling decided on the parameters for his model simply because they led to interesting outcomes within the model (Schelling, 2006).

Yet when we see how the agents in Schelling's model tend to separate into segregated communities solely because they prefer to live in proximity to people of the same group, we have learned *something*. That such a minimal model is informative in this way implies that the purpose of modeling is not to represent the world as accurately as possible.

Therefore, Grüne-Yanoff argues that the model aims to represent a parallel world that is nonetheless conceivable and credible. By "credibility," Grüne-Yanoff means that it depicts a real and meaningful possibility. How does knowledge about a parallel world contain epistemic value? By demonstrating a relevant possibility, the result of a model can disprove necessity claims: to prove that possibly $\neg\phi$, after all, is equivalent to proving that it is not necessary that ϕ. For example, by demonstrating that segregation can occur without systemic racism, the Checkerboard model disproves the claim that segregation is necessarily caused by institutional racism.

However, two issues arise from this account. First, it is unclear why a fact about a possible state of affairs can inform our understanding of the actual state of affairs. On Schelling's checkerboard model, Julien Ross comments, "This hypothesis at best shows us that actual segregation can result otherwise, not that it does so, even in a single case" (Ross 2013). In Ross's view, although models allow us to learn about the world by becoming aware of relevant possibilities, these possibilities do not help us explain actual social-scientific phenomena.

There is also the related issue of what counts as a credible world. Grüne-Yanoff argues that credibility does not depend on whether we think the state of affairs in a model could possibly obtain in the real world. The models under discussion are "objective possibilities:" the assumptions that define the model are clearly false (Grüne-Yanoff and Verreault-Julien, 2021). Instead, he argues that we judge models to be credible like we judge fictional novels to be credible—we ask whether the agents' actions are reasonable given the model environment. In this way, "conditional credibility judgments are driven by empathy, understanding, and intuition" (Grüne-Yanoff, 2009). Empathy, understanding, and intuition may be sufficient for evaluating novels, but they are not sufficient for justifying the epistemic value of models. If the role of a model is merely to tell a believable story, then fictional stories would be just be as epistemically valuable as social-scientific models. There must be some feature whereby the world that Schelling's model describes is more conducive to explaining real phenomena than the world of Harry Potter or Sherlock Holmes, and we ought to be able to evaluate the epistemic merit of a model by judging to what extent it exhibits such a feature.

3. Dominance Claims

In this section, I introduce dominance claims and argue that they address the above issues. At the highest level, dominance claims are propositions of the form " if possibly ϕ and ψ, then necessarily q," where ϕ is the claim in question and ψ are conditions of the model. Given a dominance claim and a model result, we can infer:

$$\begin{aligned} \text{Model Result:} &\quad \Diamond(\phi \wedge \psi) \\ \text{Dominance Claim:} &\quad \Diamond(\phi \wedge \psi) \implies \Box\phi \\ \text{Therefore,} &\quad \Box\phi. \end{aligned}$$

Usually, ϕ is a ubiquitous phenomenon that we already observe and wish to explain. By establishing a necessity claim, we have proved a law-like tendency that is the hallmark of scientific explanation.

How do we establish results that follow from dominance claims? A straightforward way is to prove that some variable r is not the only cause of a phenomenon. Suppose that through empirical evidence, we know that a parameter robustly correlates with a ubiquitous phenomenon. For example, r could be the wealth of a democratic country, and p the homicide rate. Suppose we create a model in which the wealth of an ideal democratic country is unlimited and discover that in the model, the homicide rate is still substantial. We then infer from the fact that in all possible democracies, including our actual democracies, wealth is finite to the fact that necessarily, the prevalence of homicide cannot be solely caused by wealth, but instead in conjunction with some other factor that is present both in actual democracies and the possible model democracy. The empirical evidence licenses our judgment that the modal claim from the model implies a necessity claim about all democracies.

Dominance claims also provide criteria for choosing a model that goes beyond intuition and empathy. We ought to choose a model that describes an *extreme* world in order to *stack the deck* against the target phenomenon. In the example above, we stack the deck against the target phenomenon by setting our hypothesized cause (wealth) to the extreme (infinity). An explanandum is a good candidate for a model-based approach if the extreme world that would establish a dominance claim is adequately describable with the social scientist's modeling tools. In the next section, I analyze one such phenomenon to demonstrate how the dominance claim approach can be applied with great epistemic awards.

4. Dominance Claims in the Sociology of Science

Misinformation is not just an issue that plagues the general public. Throughout history, scientists have become misinformed and polarized about many facts, even though they and the institutions that aggregate and evaluate their work were more epistemically rational than the general population. Suppose we want to explain why misinformation is so ubiquitous, even in the scientific community. O'Connor and Weatherall (2018) suggest that misinformation can arise on sufficiently uncertain claims from the intrinsic structure of a social community. To support their claim, they construct and study the behavior of a connected social network graph that represents an ideal scientific community. In the graph, agents are represented as nodes, and their social connections are represented as edges. At each time step, agents "run an experiment" on a random variable p, sampling it once and observing the outcome. Furthermore, at each time step agents receive information about the credences of the agents to whom they are connected. Agents are perfect Bayesians; At each time step, agents update their credences on a proposition based on 1) the experiment they run on each step and 2) the credences of agents to whom they are connected. (In later versions of the model, O'Connor and Weatherall consider agents that follow Jeffery's rule, a modified form of Bayes theorem that allows agents to discount the weight of agents with sufficiently different credences.) O'Connor and Weatherall discover that in circumstances where 1) the number of experiments that an agent can run at each step is limited, 2) the social network is organized into cliches where intra-cliche connectivity is high but inter-clique connectivity is low, and 3) agents discount the weight of the evidence from other agents with sufficiently different credences, the social network can become polarized; for instance, one group of scientists may believe that the frequency of p is 1, while one group of scientists may believe that the frequency of p is 0, regardless of the true frequency of p.

Suppose we have evidence that humans diverge from Bayesian reasoning in predictable ways that make them more likely to have inaccurate credences. Indeed, it is mathematically impossible for any agent to have more accurate credences than a Bayesian agent. Furthermore, suppose that in the social network, connected agents influence each other's beliefs more rationally and clearly than in the real world. We then have the informal argument that

1. The simulated ideal agents can fall victim to misinformation.
2. If ideal agents fall victim to misinformation, then non-ideal agents are even more susceptible to misinformation.
3. Necessarily, all people, including scientists, are non-ideal agents.
4. Necessarily, people are susceptible to misinformation.

More precisely, suppose that we can quantify rationality in a state of affairs as a rational number $p : [0, 1]$, where communities are perfectly rational when $p = 1$. Suppose also that the less rational agents are, the more prone they are to misinformation. Let ϕ be the proposition that agents in a network can become misinformed and r be the level of rationality in a community. We then have

$$\begin{aligned} \text{Model Result:} \quad & \Diamond(\phi \wedge (p = 1)) \\ \text{Dominance Claim:} \quad & (\exists n \in [0, 1] \ : \ r = n \wedge \Diamond(r \geq n \wedge \phi)) \rightarrow \phi \\ \text{Quantifiability Assumption:} \quad & \Box(0 \leq p \leq 1). \\ \text{Therefore,} \quad & \Box\phi. \end{aligned}$$ [1]

By building a model that stacks the deck against the hypothesis that scientists can become misinformed, O'Connor and Weatherall license an inference from the possible world to the actual world. Through their demonstration that even ideal scientists in ideal networks often polarize, they demystify and explain an initially puzzling observation about collective scientific irrationality. Misinformation is not solely caused by individual or group irrationality; it can arise out of the inherent structure of a social community.

5. Establishing Domaince Claims

The main obstacle to this social-scientific methodology is the robustness of the dominance claim. In social systems and other complex phenomena, a multitude of causes interact in unexpected ways. The inability of models to account for all the causes that influence the behavior of the target phenomena may lead to a *misdescription of the interactions* of the variables that are included, producing apparent interactions where there are none, or apparent independence where there are interactions

[1] Strictly speaking, this deduction is only valid in some systems of modal logic. Thankfully, these technicalities are irrelevant to this paper's main argument.

(Wimsatt 2007). O'Connor and Weatherall's model may suffer from the second case. For example, it is conceivable that the ways that other factors not included in the model cause scientists in the actual world to counteract the tendency towards misinformation better than the ideal social network. It would then be the case that irrationality was the root cause of misinformation in actual scientific communities after all. A related worry is that we must extrapolate existing empirical evidence to justify the dominance claim. In the democracy case in section 2, for example, our existing evidence can only ascertain the negative correlation between wealth and the homicide rate up to a certain level of wealth. We must extrapolate from existing data points to the hypothesis that this trend continues to the limit, and the process of extrapolation necessarily introduces uncertainty into our inferences. However, only with this extrapolated correlation can we justifiably claim that our model has the effect of "factoring out" the effect of wealth.

Nonetheless, we can ameliorate these uncertainties and make correlations more robust through further empirical work. For example, in the context of scientific communities, we can run experiments and perform observations to determine whether the correlation between susceptibility to misinformation and divergence from ideal rationality is robust enough to warrant belief in the dominance claim. By robustness, I mean the invariance of the phenomenon in different scenarios and variable ranges. If we have overwhelming evidence that irrational scientific communities are more likely to polarize than more rational scientific communities in most or all situations, no matter the degree to which each community is irrational or rational, we would have good evidence for the claim that polarization and rationality are always inversely related from $p = 0$ to $p = 1$. This fact would then allow us to gainfully employ O'Connor and Weatherall's model to infer our desired proposition.

6. Conclusion

If dominance claims are a coherent approach to making epistemic progress in the social sciences, we establish the thesis that social-scientific models can be genuinely explanatory even if they are strictly unrepresentative of the real world. Furthermore, my framework provides a methodology for selecting explananda that allows for a model-based approach and selecting the types of models that would generate knowledge.

A further point of exploration would be to determine more methods for establishing dominance claims. In this paper, I have focused on O'Connor's and Weatherall's methodology, but there may be other imaginative arguments that would justify an inference from a proposition about a possibility to a necessary truth that holds in the actual world.

References

[1] Cartwright, N. (1999). *Capacities*. In J. B. Davis, D.W. Hands, & U. Mäki (eds.), *The Handbook of Economic Methodology* (pp. 45-48). Cheltenham: Edward Elgar.

[2] Grüne-Yanoff, Till. (2009). *Learning from Minimal Economic Models*. *Erkenntnis (1975-)*, *70*(1), 81–99. JSTOR, http://www.jstor.org/stable/40267409.

[3] Grüne-Yanoff, Till, & Philippe Verreault-Julien (2021). *How-possibly explanations in economics: anything goes?*. *Journal of Economic Methodology*, *28*(1), 114-123. DOI: 10.1080/1350178X.2020.1868779.

[4] Mäki, U. (2004). *Realism and the nature of theory: A lesson from J.H. von Thunen for economists and geographers*. *Environment and Planning A*, *36*, 1719-1736.

[5] O'Connor, Cailin, and James Owen Weatherall. (2019). *The Misinformation Age: How False Beliefs Spread*. Yale University Press. JSTOR, https://doi.org/10.2307/j.ctv8jp0hk. Accessed 4 Dec. 2023.

[6] Priest, G. (2008). *An Introduction to Non-Classical Logic: From If to Is* (2nd ed.). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511801174.

[7] Reiss, Julian. (2013). *Philosophy of Economics: A Contemporary Introduction*. New York, NY: Routledge.

[8] Schelling, T. (2006). *Strategies of Commitment and Other Essays*. Cambridge, MA: Harvard University Press.

[9] Wimsatt, W. (2007). *Re-engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Boston, MA: Harvard University Press.