Understanding with theoretical models

Petri Ylikoski and Emrah Aydinonat

Tweet

LiU LINKÖPING UNIVERSITY

Understanding with Theoretical Models

Petri Ylikoski
University of Helsinki
petri.ylikoski@helsinki.fi

N. Emrah Aydinonat
Bahçeşehir University
aydinonat@gmail.com

Abstract

This paper discusses the epistemic import of highly abstract and simplified theoretical models using Thomas Schelling's checkerboard model as an example. We argue that the epistemic contribution of theoretical models can be better understood in the context of a cluster of models relevant to the explanatory task at hand. The central claim of the paper is that theoretical models make better sense in the context of the menu of possible explanations. In order to justify this claim, we introduce a distinction between *causal scenarios* and *causal mechanism schemes*. These conceptual tools help us to articulate the basis for modelers' intuitive confidence that their models make an important epistemic contribution. By focusing on the role of the menu of possible explanations in the evaluation of explanatory hypotheses, it is possible to understand how a causal mechanism scheme can improve our explanatory understanding even in cases where it does not describe the actual cause of a particular phenomenon.

Highly abstract and simplified theoretical models have an important role in many sciences, for example, in evolutionary biology and economics. Although both scientists and philosophers have expressed doubts about the epistemic import of these idealized models, many scientists believe that they provide explanatory insight into real-world phenomena. Understanding the epistemic value of these abstract representations is one of the key challenges for philosophers of science who attempt to make sense of scientific modeling. In this paper we will examine Thomas Schelling's checkerboard model as an example of abstract theoretical models and articulate various ways in which it expands the social scientific understanding of segregation processes.

We argue that the epistemic contribution of theoretical models can only be fully understood in the context of a cluster of models relevant to the explanatory task at hand. To disambiguate this claim, we introduce a distinction between two different cluster claims. The first, the family of models thesis, states that the epistemic value of theoretical models is not well understood if they are treated as isolated representations. It makes more sense when considered in the context of as a family of related models. Our second cluster claim, the competing causal mechanisms thesis, says that the full epistemic contribution of theoretical models can only be understood in the context of competing explanations for the same phenomenon. In other words, theoretical models make better sense in the context of the menu of possible explanations.

The structure of the paper is as follows. In the first section, we start by making some observations about Schelling's famous checkerboard model and introduce our two cluster theses in the second section. In order to understand our second thesis, which brings up the contribution of the menu of possible explanations, it is necessary to understand the nature of how-possibly explanations (HPEs). Thus, in the third section, we introduce an enriched account of HPEs and a distinction between causal scenarios and causal mechanism schemes. This enriched account helps us make sense of the role of theoretical models in the construction of HPEs. The fourth section explicates the epistemic import of HPEs and introduces the possible ways in which highly abstract theoretical models such as Schelling's could contribute to our understanding. The fifth section

concludes the paper by summarizing our central claims and discussing the generalizability of our observations about the checkerboard model.

1. Schelling's checkerboard

Schelling proposed his self-forming neighborhood model in a series of articles (1969, 1971a, 1971b, 1972, see also 1978). The model, usually called the checkerboard model, was presented in the context of other models concerning segregation and tipping point processes that he considered to be examples of more general critical-mass models (Schelling, 1972: 157). The checkerboard model is by now one of the most discussed models in the philosophy of the social sciences (See, for example, Sugden 2000; Aydinonat 2008; Clark & Fossett 2008). Since it is a well-known model, we will briefly introduce and use it in order to highlight the defining lines of our argument (see Aydinonat 2007, 2008 for a detailed exposition of the model).

The checkerboard model focuses on discrimination and the effect of discriminatory preferences on the neighborhood structure. The assumption that the agents are able to discriminate is meant to signify that they are able to distinguish between their own group and other groups. The model assumes that there are two types of agents, As and Bs, and that all agents are able to identify which agent belongs to which group. However, "discrimination" does not imply that the agents dislike (or like) the other group. Nevertheless, the agents' preferences may be based on "discrimination." For example, As may prefer to live together with Bs, or Bs may prefer to live only in neighborhoods where there are no As. In what follows, we will call these "discriminatory preferences." Agents may have a range of discriminatory preferences. Let us distinguish between strong and mild discriminatory preferences. A strong discriminatory preference implies that the agent wishes to live in a neighborhood where its own type is the majority. For example, if As are racists, they will have a very strong preference for avoiding Bs. A mild discriminatory preference implies that the agent is happy to live in a neighborhood as a minority and could tolerate a minority status up to a certain point.

The core of the checkerboard model is the interaction of agents who have mild discriminatory preferences. As the name suggests, the checkerboard model places the agents on a checkerboard-like city (NxM grid) where every agent has eight neighbors. The basic model assumes that agents only care about their immediate neighbors. Thus, their discriminatory preferences concern only their immediate neighbors. The model assumes that all agents have mild discriminatory preferences and that they cannot tolerate an extreme minority status. That is, agents will be happy to live in a neighborhood even if they have a minority status; however, they will be unhappy if they have an extreme minority status. In the original model Schelling assumes that agents will be content unless their own type is less than one third of the neighborhood (Schelling 1978: 148).

The dynamics of the model are based on the interactions of the agents. Every agent is free to move to another place if it is discontent with the composition of its immediate neighbors. If an agent moves out of its neighborhood, it changes both the composition of its past and future neighborhoods. For example, if an A moves out of a neighborhood, the number of As in this neighborhood decreases, and the number of As in its new neighborhood increases. This in turn may cause other agents to decide to leave one of these neighborhoods. The process ends when every agent is content with its neighbors.

The checkerboard model shows, 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 live as an extreme minority. Despite its highly unrealistic assumptions, the model is considered a paradigmatic example of an invisible-hand explanation (e.g., Aydinonat 2008), one of the predecessors of agent-based computer models (e.g., Epstein & Axtell 1996, Rosser 1999, and Casti 1989), an example of a good explanation in social sciences (e.g., Sugden 2000), and a classical exemplar of explaining using social mechanisms (e.g., Hedström & Ylikoski 2010).

It is important to recognize that the checkerboard model is not taken to provide a convincing

explanation of some empirical cases of residential segregation. Racial segregation is only an illustration of a more general idea that has a much broader application. For this reason, it does not really matter if 'mild discriminatory preferences' do not account for many real-world cases of ethnic segregation. As Schelling (1978: Chapter 4) himself quite aptly shows, the model could be applied to other types of segregation, such as segregation among friends at parties or the segregation of guests at a wedding reception. More generally, segregation can be defined as "the nonrandom allocation of people who belong to different groups into social positions and associated social and physical distances between groups" (Bruch & Mare 2009b: 270). The checkerboard model provides a generic template for thinking about a *possible way* in which segregation may emerge. Its general applicability is one of the reasons why social scientists find it appealing.

Like many other theoretical models in biology and economics, the checkerboard model is very abstract and contains strong idealizations. For example, it assumes homogenous preferences, ignores strong discriminatory preferences, and makes use of very specific assumptions concerning neighborhood structure. Most importantly, it ignores all well-known causes of segregation: the model does not take economic factors, welfare differences among groups, and organized discrimination into account (Aydinonat 2007, 2008).

These idealizations have often been presented as grounds for challenging the relevance and explanatory power of the checkerboard model. A common objection is that it is too abstract to be explanatory. First, the model is said to ignore more important causes of segregation, such as racism. Some even suggest that by ignoring the 'main' causes (such as racism), the checkerboard model trivializes them. As Lee puts it:

> In real life, racial segregation in American cities, such as Kansas City, was a
> result of deliberate housing policies by racist home construction companies
> and realtors and not the result of unintended consequences of individual

actions. […]. Schelling-type explanations of segregation trivialize racism and

hence endorse it. (Lee 2003: 77)

The model also ignores other obviously relevant factors such as the costs of moving, not to mention various psychological and sociological factors that influence neighborhood preferences and residential segregation. Furthermore, it focuses on a process (the interaction of individuals with mild discriminatory preferences) that is generally considered to be either unimportant or nonexistent from the perspective of residential segregation. These simplifications make the model highly unrealistic, and thus it is criticized as being irrelevant for the explanation of real-life residential segregation. For example, it is argued that the model may in principle explain segregation in highly simplified artificial worlds, but these results cannot be extrapolated to much more complex situations (Goering 2006, Cartwright 2009).

Even though these criticisms are common, many social scientists consider the checkerboard model as an insightful model that has explanatory import. Why is it taken to provide explanatory insight by these social scientists? How could it be explanatory despite its seeming dissimilarity to everything we know about the formation of ethnically segregated neighborhoods? Recently, a number of philosophers of economics have tried to answer these questions (Cartwright 2009; de Donato and Zamora-Bonilla 2009; Grüne-Yanoff 2009; Knuuttila 2009; Kuorikoski and Lehtinen 2009; Mäki 2009; Sugden 2009). These papers mostly focus on the relation between a model (e.g., *the* checkerboard model) and its supposed target. The focus on this relation had made it possible to consider how a single model represents its target and what kind of (approximately) true statements can be deduced from the model. However, the scientists also take the family of checkerboard models to provide a more general understanding that goes beyond the particular modeling targets. This more general understanding is left out when the focus is on the relation between the model and its target. For this reason, it still remains to be explained why social scientists find Schelling's simple and highly unrealistic model insightful.

In our view, the key to this puzzle is found in two aspects that have not received enough attention by philosophers. The first is the observation that *models come in clusters*. Understanding the epistemic import of highly abstract models requires more than analyzing the relation between a selected model and its target, it requires grasping how the model relates and contributes to an existing cluster of related models. The second is the role of models in the construction of HPEs. Understanding the epistemic import of a highly abstract model requires grasping how it contributes to an explanation, especially in the context of existing explanations. In what follows, we will introduce two important senses in which models come in clusters and then proceed to explicate the role of models in how-possibly reasoning about the phenomena. [1]

2. Two cluster theses

Family of models thesis

Although it is customary to talk about the checkerboard model as if it were a single model, the name 'checkerboard model' does not refer to a single model, but to a cluster of related models. This cluster consists of an expanding family of models that are variations of Schelling's original models. Thus, one cannot fully appreciate the epistemic import of such models by way of singling out one model from this cluster as 'the checkerboard model' and analyzing it in isolation. 'The checkerboard model' should be considered a family name: it refers to a whole cluster of models that are related to each other through genealogical origin and similarity. Secondly, 'the checkerboard model' is not a model of any specific empirical phenomenon – it is a theoretical model. It is a device for thinking about theoretical possibilities, not about a particular real (or possible) world phenomenon.

The implication of these observations is that a more appropriate unit of analysis is the ongoing research initiated by Schelling's papers. It would be arbitrary to focus only on early models in

---

[1] There is also a third sense in which 'models come in clusters' that we will not discuss in this paper. It is based on the observation that the same model can be given alternative interpretations that differ in their strength and assumptions (Godfrey-Smith 2006). Once there is a model, it is possible to have a whole cluster of ways to interpret it as a representation of a given target.

this tradition or to limit one's attention only to models that have been proposed by Schelling himself. The research concerning Schelling's original insights is extensive both in a temporal and a social sense: checkerboard models have been studied for over forty years by a multitude of scholars in various disciplines. The research has produced many surprising findings that should not be read back to the original models. It is by analyzing the whole continuum of such models that one can make better sense of the real epistemic contribution of this family of models. This is our first cluster thesis: abstract models make better sense when understood as a family of related models, not as isolated representations. We will call it the family of models thesis.

Recently some philosophers of science have argued – inspired by Richard Levins – that robustness analysis has an important role in working with models (Weisberg 2006; Kuorikoski, Lehtinen & Marchionni 2010; Kuorikoski & Lehtinen 2009). More specifically, scientists build models with slightly different auxiliary assumptions in order to test whether the key insights of the original model hold under different conditions. This is called derivational robustness analysis (Woodward 2006). The importance of robustness analysis derives from the possibility of finding which auxiliary assumptions are irrelevant to the conclusions of the model: if the result is robust with respect to the change, the researcher can infer that the given unrealistic assumption does not play a crucial role in the model and thus it is not an important explanatory variable in the model (Kuorikoski & Lehtinen 2009). Similar reasoning is at work also in the identification of the crucial explanatory dependencies underlying the model (cf. Kuorikoski, Lehtinen & Marchionni 2010). The point in this case is to find those assumptions that do affect the modeling outcomes: to study exactly what difference those assumptions make to the outcome. Here, it is important that the assumptions make a difference within a realistic range of variation as only these assumptions allow inferences to real-world systems.

Research concerning the checkerboard segregation models is a perfect illustration of this kind of reasoning. If one takes a look at the recent literature on 'the' checkerboard model (e.g., Benard & Willer 2007; Bruch & Mare 2006; Clark & Fossett 2008; Fossett 2006, 2011; Macy & van de Rijt

8

2006; Pancs & Vriend 2007; Muldoon, Smith & Weisberg 2012; see Aydinonat 2008 for a review), one finds that the main focus of research is on the exploration of a series of what-if questions.

It turns out that the segregation equilibrium seems to be a robust outcome in many of the modifications to the model's assumptions (Aydinonat 2007; Muldoon, Smith & Weisberg 2012). Most changes in the size of the neighborhood, individual preferences, or the availability of apartments do not change the segregated outcome. Similarly, a small random move can lead a non-segregated area (where everybody is satisfied) towards a path that ends in complete segregation. The robustness of the model speaks to the wide relevance of the mechanism: it can work under a number of different circumstances (Schelling 1978; Follett 2006). However, changing the neighborhood structure, the probability of random errors, or the type of preference functions may lead to non-segregation outcomes (e.g., see Bruch and Mare 2006, 2009a; van de Rijt, Siegel, and Macy 2009).

The robustness of the segregation outcome depends on a relatively unspecified and qualitative notion of segregation. If one focuses on how long reaching the equilibrium takes on average, or on what kind of segregation the typical outcome is, the results are much less robust. However, it is important to recognize that non-robust results are also interesting. They provide valuable information from the perspective of explanatory understanding as they tell us that we need to invoke other explanatory factors or that we have to pay more attention to specific details of the model. Knowing more about robust and non-robust cases gives us better tools for answering a broader set of what-if questions. What may seem a discomforting non-robust result from the point of view of an individual model may be a crucial step in the road to better understanding once it is located within a larger context of related models.

It is important to note that explorations of 'the' checkerboard model are not limited to derivational robustness analysis. For example, with the help of agent-based computer

simulations, social scientists are asking questions such as (Ylikoski 2013):

· What would happen if we change the size of the population?

· What would happen if we change the size of the neighborhood that the agents observe?

· What would happen if we change the number of groups (or their relative sizes)?

· What would happen if we change the ratio of vacancies (or their accessibility)?

· What would happen if we change the nonlinear choice function to a linear one?

· What would happen if we change the symmetric preferences to asymmetric ones?

· What would happen if we allow errors in individual choices (and observations)?

· What would happen if we introduce other factors (wealth, social status, etc.)?

· What would happen if we assume that some agents have a preference for integration?

These what-if questions represent modifications of the original idea presented by Schelling, and their point is to see how the changes influence the segregation outcomes and the working of the mechanism. All of these modifications cannot be categorized as derivational robustness analysis for at least two reasons. Firstly, such modifications do not only involve modifications to background assumptions. Asking different *what-if* questions involve modifications of core assumptions too. Secondly, such modifications are not always carried out as a means of derivational robustness tests, rather some modifications involve the addition of new elements to the model, and/or changing a group of assumptions altogether. Thus derivational robustness analysis explains only a subset of the modifications carried out in actual research.[2] The cluster conception of models not only emphasizes the importance of derivational robustness analysis, but also provides a wider perspective within which the epistemic import of each model in the cluster

---

[2] In fact, it is hard to list and classify all of the ways in which models are tested. Lacking a better term we use 'modifications' and 'explorations' interchangeably in order to indicate the general idea. Aydinonat (2008) discusses the specific ways in which models of emergence of money and segregation modify the original models that they are based on.

becomes meaningful.

From the philosophical point of view, the important thing is that the exploration of what-if questions in robustness analysis only makes sense if one takes the cluster conception of models seriously. By focusing on how scientists systematically study and manipulate a whole family of related models is it possible to understand the activity that they are engaged in and how it is possible to make inferences about real-world phenomena based on these very abstract models.

Competing causal mechanisms thesis

Our second cluster claim about models is based on the observation that theoretical models (such as the checkerboard model) are usually discussed and evaluated in the context of competing theories. Sometimes these alternative explanations have not yet been articulated into explicit models (as in the case of alternative sociological explanations of segregation), but this does not diminish their relevance. They provide the necessary context for making sense of the full epistemic import of the checkerboard model. This is our second cluster thesis: models make better sense in the context of competing explanations for similar kinds of effects. We will call this the competing causal mechanisms thesis.

The observation that segregation is a robust result of even mild discriminatory preferences expands the *menu of possible causes* of segregation – it adds one more explanatory factor to the ones suggested by earlier studies of ethnically segregated neighborhoods. The model shows how the individual preferences (avoiding being in a minority) produce a surprising collective outcome (segregation). As there is discontinuity between individual preferences and the collective outcome, the possibility of this causal mechanism is only visible when it is explicitly modeled. Outside the checkerboard model, among possible causes of residential segregation are:

· Strong discriminatory preferences (e.g., straightforward racism)

· Economic inequalities between groups (combined with price differences in housing)

· Institutional discrimination (redlining, racial steering by real-estate agents, etc.)

- Structural features associated with group membership (the location of employment, channels of information flow about available housing, etc.)

The checkerboard model adds one theoretically important explanatory mechanism to the list. Importantly, it is not a mere addition to this list. These mechanisms can both compete with the Schelling mechanism and combine with it.

While Schelling originally suggested using the checkerboard model to describe a sufficient cause of ethnic residential segregation, he did not imply that mild discriminatory preferences are necessary for the explanation of segregation. Although the model isolates away other causes, such as racism or welfare differences, it does not exclude them as possible causes of segregation (also see Aydinonat 2007).

In other words, Schelling suggested a how-possibly[3] explanation (HPE) (Brandon 1990: 176-184, Forber 2010). How-possibly explanations refer to causes that could bring about the effect to be explained. In contrast, the actual explanation tells us what actually brought about the phenomenon. The actual explanation is one of the possible explanations. Quite often we do not know the actual explanation, so we have to operate with possible explanations and try to find evidence that would discriminate between them. Some possible explanations are incompatible: the actual cause may make it impossible for any of the alternative causes to be involved in the production of the phenomenon. In such cases, the other causes could be causes only in the absence of the actual cause. However, in many cases it is possible that more than one actual cause is involved in the production of the *explanandum* phenomenon. The relations between these actual causes might vary: sometimes they constitute cases of overdetermination, in some other cases their relation is that of pre-emption (more on this later). It turns out that possible

---

[3] The notion of the how-possibly explanation was originally introduced by Dray (1957). His concern was with apparently surprising *explananda* that challenge our beliefs concerning what is possible. The answers to these explanation-seeking questions would tell us how the phenomenon is possible at all. Such explanations provide the necessary conditions for something to happen and they might be called how-possible explanations (von Wright 1971). In contrast, the how-possibly explanations that are the focus of this paper are about various ways in which something could have happened. There is no assumption that the *explanandum* is in any way surprising.

explanations of residential segregation belong to this latter, more complicated, class of how-possibly explanations.

3. Causal scenarios and causal mechanism schemes

In order to properly understand HPEs and how they contribute to our understanding, we have to recognize the difference between two different kinds of *explananda*. First, a HPE might address some particular empirical fact. We will call these empirical how-possibly explanations possible causal scenarios. A second kind of HPE does not address any *specified* empirical fact. It is concerned with what is possible in principle: what kinds of processes can bring about certain kinds of effects. We will call these theoretical how-possibly explanations causal mechanism schemes. Let us take a closer look at their differences.

Possible causal scenarios

HPEs that address particular empirical facts are historical scenarios that tell us how the *explanandum* fact could have come about. They are causal histories of that particular fact that are at least partially incompatible with each other. One could call them alternative causal histories. To count as separate scenarios, there must be some difference in their crucial explanatory details. If two scenarios only have differences in the background factors that do not make a difference to the outcome, they do not count as competing scenarios. The causal scenarios might be mere sketches of causal histories, or they might be highly detailed. What matters is that they contain the pivotal elements of the causal chain leading to the particular fact to be explained.

What makes a causal scenario possible or impossible? To answer this question, we must consider two dimensions concerning "explanatory possibility." First, there is the question: if the scenario had taken place, would it have produced the effect to be explained? Let us call this *causal possibility*. Judgments about causal possibility are made on the basis of general causal knowledge. Thus, in order for a causal scenario to be deemed a possible scenario its depiction of causal factors and their interaction should not contradict our general causal knowledge. The

second dimension concerns *factual possibility*. The crucial thing here is the compatibility with facts known about that particular causal history. A causal scenario could only be considered a factually possible scenario if it is consistent with the facts we know concerning the particular *explanandum* phenomenon at hand. In sum, a possible causal scenario is such that it does not contradict the known facts or generally accepted causal principles.[4] Thus, because a possible causal scenario is causally and factually possible, it could be considered a possible explanation of the particular fact at hand. If we had complete knowledge of causal possibilities and we knew all the facts about the causal history, there would be no need to talk about possible causal scenarios. However, often the available evidence underdetermines a set of competing causal scenarios. Accordingly, in this context we are talking about what is possible in the light of existing evidence.

The purpose of gathering evidence in the context of competing causal scenarios is to narrow the space of possibilities, that is, to eliminate as many of the competing scenarios as possible. The possible causal scenarios play an important role in the search for new evidence; they tell what to look for and what may be relevant. This role determines a key virtue of a good causal scenario: it should be specific enough to suggest what kind of evidence would speak for or against it. Moreover, as the aim is to discriminate between the scenarios, the description of the causal scenario should be particularly clear about those aspects that make it different from its competitors. The ultimate aim of this process of elimination is a situation where only one possible or plausible causal scenario is left standing. One does not usually achieve this situation due to the limitations of the evidence. This should not be taken as an argument against this account of evidential reasoning. Rather, it is a fact of scientific life that we have to learn to live with. The idea of excluding all explanations but one still serves as a regulative ideal.

One more question needs to be settled before going further: Is it possible to consider mechanisms

---

[4] Forber's distinction between global and local how-possibly explanations is roughly equivalent with our distinction between causal and factual possibility. This distinction should not be confused with the distinction between causal scenarios and causal mechanism schemes.

depicted by abstract theoretical models as causal scenarios? For instance, it seems to be possible to take any instance of the checkerboard model as a very simple or sketchy causal scenario. The checkerboard model may *in principle* show, for example, how ethnically segregated neighborhoods emerge in some real cities. However, it is our claim that the main epistemic import of the checkerboard model does not rest on this possibility. First, a highly stylized model is not a very good causal scenario, particularly because it abstracts from those particular details that make up a good causal scenario. Schelling's checkerboard model is not based on observational data, it typically considers only one kind of preference, and the model completely ignores other mechanisms that might contribute to segregation. Like the majority of abstract theoretical models, it focuses on a limited set of explanatory factors (under a restrictive set of conditions) that commonly cannot alone explain real-world facts. Second, the interpretation of the checkerboard model as a causal scenario does not do justice to the generality of the model. As we have already pointed out, residential segregation merely serves as an illustration of the more general idea. It provides a template with an open application domain. Only future research will show how widely it can be employed. In sum, mechanisms depicted by abstract theoretical models can only serve as rough skeletons for causal scenarios, hence their contribution is better understood in relation to *causal mechanism schemes*.

Causal mechanism schemes

Working with possible causal scenarios presupposes that one has a list of things that could in principle generate the fact to be explained. This knowledge is not *a priori,* so theoretical work can influence the set of possible causal scenarios. Theoretical arguments can suggest a novel way to bring about the phenomenon of interest. Similarly, they can sometimes show that some causal scenario candidates should be dropped from consideration as they are not in fact able to bring about the phenomenon.

Theoretical HPEs are not primarily concerned with compatibility with the evidence about a

particular causal history. Their central concern is the generative sufficiency of the candidate scenarios. That is they are concerned with whether the depicted causal mechanisms are sufficient to produce the specified effect. Theoretical reasoning attempts to explore the limits of causal possibility and to test the explanatory sufficiency of the candidate scenarios. Theoretical work such as this abstracts away from the evidential constraints of a particular case, and consequently (but not necessarily), its products (often) have a greater scope of application. To clearly distinguish them from particular causal scenarios, we will call them *causal mechanism schemes*.

The *explananda* of theoretical HPEs are often highly abstract and stylized. For example, Schelling's model does not explicate any particular case of residential segregation. If one tried to directly use this model in explaining something, it would only explain how segregation emerges in the abstract checkerboard city under the conditions specified by the model. Thus, its *explanandum* would be the type of segregation that could emerge in a NxM grid which is occupied by two types of agents who have mild discriminatory preferences concerning their immediate (eight) neighbors given the other assumptions of the model. It is true that the model ultimately wishes to shed light on actual cases of segregation. However, its *explanandum* is very abstract. Another way to put this is that abstract theoretical models are highly selective representations (more on this in the next section). Nevertheless, only by abstracting away from the details of particular applications is it possible to study the generative sufficiency of a particular mechanism. This generality has, however, a cost. Causal mechanism schemes do not directly explain any particular empirical fact. They address only simplified theoretical *explananda*. For this reason, the knowledge of possible causal mechanisms influences particular explanations mainly indirectly. First, theoretical considerations of explanatory sufficiency modify the menu of possible explanations for a given particular fact. Second, articulations of mechanisms help to construe causal scenarios by providing skeletons of causal processes that play a crucial role in the causal scenario. They tell which parts of the causal process are important for the explanatory purposes at hand and how they could affect the outcome of the process.

This line of reasoning leads to the conclusion that it would be a mistake to assume that all theoretical models are models of some empirical phenomena. For this reason, focusing on the relation between an isolated theoretical model and its assumed particular target does not help us in understanding the epistemic import of these models. The route from the theoretical model to a particular empirical application is commonly more complicated. The main contribution of theoretical models to the explanation of particular facts is indirect: theoretical models modify the menu of possible mechanisms. This is accomplished by providing rough skeletons for causal scenarios and studying what the specified mechanisms can achieve under certain conditions.

While causal mechanism schemes can serve as building blocks of causal scenarios, there is no direct route from the causal mechanism scheme to a full-blown causal scenario that would explain some particular facts. First, often a single causal scenario may involve a number of different theoretical causal mechanisms. These mechanisms could combine to produce the outcome, or they might even influence the process in opposing directions. For this reason, it is fruitful to think of theoretical causal mechanism schemes as building blocks of causal scenarios (Ylikoski 2011). Second, theoretical causal mechanism schemes can be adapted to particular cases in a number of different ways. The skeleton provided by a causal mechanism scheme allows many, often incompatible, ways of building a representation of a particular causal scenario.

When causal scenarios combine multiple mechanisms, the respective explanatory roles of the mechanisms can vary. For example, when considering mild and strong discriminatory preferences, both should be considered as special cases of discriminatory preferences. Both can be subsumed under a generalization that says: "above the (specified) threshold (of tolerance), segregation will follow." In contrast, when considering the relation between 'mild preferences' and institutional discrimination, it seems that the latter pre-empts the influence of the former. When institutional discrimination is at work, the Schelling mechanism may not have room to operate, although some individuals might have mildly discriminatory preferences. So it is

possible to accept the claim presented by many critics that the Schelling mechanism is not the most prominent cause of racial segregation in real-world cities. Usually, other factors provide an explanation for the emergence of segregation and its resilience. While the Schelling mechanism has potential relevance as a contributing cause, in most real-world cities other causal factors pre-empt its causal influence.

Nevertheless, even if another factor such as institutional discrimination provides a satisfactory explanation of the emergence and resilience of segregation, in some cases the Schelling mechanism may help in providing a more complete, better explanation. For example, if institutional discrimination is not wide ranging, mildly discriminatory preferences may work as a contributing cause, e.g., by way of increasing the speed and range of segregation caused by institutional discrimination. In such cases a more complete explanation is provided by showing how the Schelling mechanism works as a contributing cause.

Similarly, when we consider the relation between 'mild preferences' and economic inequalities, the most plausible way to think of them is as complementary causes. These two mechanisms may work independently of each other, so their combination can be analyzed as causal overdetermination. Thus, understanding how the Schelling mechanism works could provide us with a more complete understanding of the segregation process.

Moreover, 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. For example, we learn from Schelling models (and also other segregation models) that the removal of institutional segregation will not necessarily remove the segregation. Without this theoretical insight, we would probably infer that the elimination of the factors that have caused the segregation and/or currently sustain it would also eliminate the segregation (see Aydinonat 2008, 65). However, segregation models allow us to avoid this fallacy by showing that it is possible that the segregation would remain (or re-emerge) even when other causes are absent.

The conceptual tools introduced by Schelling make it possible to systematically study the conditions in which this would happen. In these cases the Schelling mechanism can be regarded as a backup cause.

Our key claim is that the mechanism suggested by Schelling is not considered important because it is assumed that mild discriminatory preferences are the most important or the commonest cause of residential segregation. Rather, the importance of Schelling's mechanism is based on the fact that it changes the menu of possible causes of residential segregation. After Schelling, serious attempts to explain the emergence or persistence of segregation have to take his mechanism seriously. In the next section we will consider the epistemic consequences that this change in the menu of possible explanations has.

4. Making sense of epistemic import

As the goal of this paper is to analyze the contribution of abstract theoretical models to scientific understanding, we need an account of understanding. We suggest that understanding should be regarded as a kind of ability – an ability to make correct *what-if inferences* about a phenomenon (Ylikoski 2009 & Ylikoski & Kuorikoski 2010). When one understands something, one can make various correct inferences about it. This ability is based on knowledge about dependences. The fundamental criterion of understanding is the ability to make inferences about counterfactual situations, e.g., the ability to answer contrastive *what if* questions by relating possible values of the *explanans* variables to possible values of the *explanandum* variable. Explanations aim to track objective relations of dependence. These dependencies are modal: explanation is not about subsumption under empirical regularities, but about counterfactual dependence (Woodward 2003; Ylikoski & Kuorikoski 2010).

This notion of understanding helps us to make sense of the epistemic goals of modeling and gives us the means to articulate the various ways in which a model could contribute to scientific knowledge. Explanatory knowledge consists of knowledge about dependencies that characterize

the phenomenon at stake. The point of (explanatory) models is to represent these dependencies in a cognitively salient manner. Explanatory understanding is based on the ability to use a family of models to make counterfactual inferences about the phenomenon. The explanatory use of models does not necessarily require that models are isomorphic with target systems. The degree of similarity between the model and the target may indeed be very low. What is required, however, is that the model should represent relevant dependencies given the explanatory task at hand. That is, it is only required that the family of models makes it possible to make correct *what if* inferences about the intended aspect of the modeled phenomenon (Ylikoski & Kuorikoski 2010). Abstract theoretical models are highly selective representations of the relevant explanatory dependencies. The modeler's judgment that a model increases her understanding is a judgment that the model increases her ability to make *what if* inferences about the phenomenon. However, as we are arguing in this paper, there is more than one way to contribute to the ability to make correct *what if* inferences about the phenomenon.

Here we see the importance of the family of models thesis: the contribution of an individual model to our ability to make counterfactual inferences about an object is very limited, as an isolated model can only answer a small number of explanation-seeking questions about a specific aspect of the phenomenon in question. Moreover, such an isolated model will seldom give a satisfactory explanatory insight about any particular instance of the phenomenon. For a systematic understanding of the explanatory dependencies, one needs a cluster of models that provides answers to a series of *what if* questions concerning a variety of aspects of the phenomenon at stake. An ability to make reliable inferences about real-world systems presupposes a systematic understanding of the ways in which the changes in the assumptions of the model changes its results. And to achieve a more systematic (and generalizable) understanding of the real world, one has to focus on an even broader range of related models.

Let us then move to the competing causal mechanisms thesis. We have claimed that the central contribution that the checkerboard model makes to our understanding of social processes is based

on the fact that it expands the menu of possible explanations of a wide variety of segregation processes. It provides an important building block for constructing causal scenarios.

As we have argued in the previous section, highly abstract theoretical models such as Schelling's checkerboard model present us with causal mechanism schemes. By way of doing this they expand the menu of available explanations for particular facts. That is, they expand the ways in which one may develop possible causal scenarios. As we have noted earlier, causal mechanism schemes serve as building blocks for causal scenarios and causal scenarios may contain multiple causal mechanism schemes. Thus, abstract models that introduce new causal mechanism schemes also enlarge the set of possible causal scenarios. The contribution of possible causal scenarios becomes apparent with the observation that the search for causal explanations often has the structure of eliminative induction.[5] The key characteristic of this process of scenario culling is that it is comparative. Explanatory hypotheses are not only evaluated by finding supporting or undermining evidence, but also by comparatively studying (and ruling out) competing hypotheses. In this process the alternative causal scenarios play an important role. As the aim of the research is to find evidence that discriminates between competing scenarios, knowledge about the scenarios and their presuppositions guides the search for new evidence. A well-specified causal scenario tells us what kinds of traces of its behavior should be found in the evidence. It is precisely this kind of evidence that is relevant for the process of cutting down the number of possible causal scenarios.

A pivotal presupposition of this process of comparative evaluation is the assumption that the alternative causal scenarios are known. A demonstration that there are unexplored, but viable, alternatives is quite damaging for the eliminative inference. This makes knowledge of the range of possible explanations crucial: we can only be confident of having the right explanation if we

---

[5] The inferential structure could also be called inference to the best explanation. However, this notion has multiple meanings (Lipton 2004). The version employed in the present argumentation is weak in the sense that it does not give the explanatory virtues (Lipton's loveliness) any evidential role (Ylikoski & Kuorikoski 2010). One could call it inference to the only explanation (Bird 2007).

have good reasons to believe that our list of alternatives is comprehensive. As a consequence, any contribution that revises the menu of possible explanations is of crucial epistemic value – the strategy of eliminative induction presupposes it. This observation helps us to understand why Schelling's checkerboard model made an important contribution to the explanatory understanding of residential segregation; its introduction expanded the menu of possible causes of residential segregation and changed the evidential landscape against which explanatory inferences are made.

The comparative nature of the assessment of competing causal scenarios provides a context for figuring out how an HPE can contribute to our understanding without providing the actual explanation. A number of different contributions can be distinguished. First, a new causal scenario – or a whole class of new scenarios – gives us a better grasp of the range of alternative causal scenarios. This knowledge makes the choice between alternative causal scenarios more secure. The more complete our menu of possible explanations is, the more confident we can be about the results of eliminative reasoning.

Second, the introduction of a new causal scenario (or the elimination of an old one) changes the evidential requirements set for accepting one causal scenario as the correct one. The change in evidential requirements happens in two ways. On the one hand, the demand for evidence rises: to justify a claim about the actual causal contribution one has to rule out one more competing explanation. A change in the menu of possible explanations brings with it a new set of challenges that the winning scenario has to overcome. This forces researchers to search for more stringent evidence. In the case of residential segregation an alternative explanation must rule out or incorporate the influence of mild discriminatory preferences. This is the second way in which a new HPE has epistemic relevance: it changes the requirements for the evidence necessary to support the causal scenario.

The third possible contribution is related. The new alternative may make *the standards of evaluation* for all the alternatives more demanding. This is clear in the case of Schelling's

checkerboard: one of its main contributions to the studies of residential segregation has not been in the area of explanatory claims but in the requirements for a satisfactory explanation. In other words, it made the requirements for an acceptable explanation of a segregation process much more stringent. After Schelling, 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 (Fossett 2006, Macy & van de Rijt 2006). This influence on standards of explanation was not limited to explanations of segregation. For example, Schelling's modeling strategy has an important historical role. The checkerboard framework also provided a template for thinking about a satisfactory social scientific explanation: the explanations should be explicit about the mechanisms, and the way in which the mechanisms bring about their effects should be studied systematically by means of agent-based computer simulations or other modeling techniques. More generally, the checkerboard model provides a practical illustration of the importance of going beyond sketchy verbal explanations (Hedström & Ylikoski 2010; Ylikoski 2013). This third contribution is, of course, not a property of all highly abstract theoretical models. However, influential theoretical models may change our standards of evaluation of alternatives by way of introducing new ways of looking at the phenomenon at stake. The above considerations also give a clear indication of the source of the epistemic value of theoretical HPEs. If the menu of possible explanations has a crucial role in the evaluation of competing causal scenarios, then the theoretical work that influences the menu is bound to be of crucial importance. Theoretical models provide building blocks for causal scenarios, thus expanding the toolbox for constructing new explanations for both novel and familiar *explananda*. Their purpose is not to explain any particular facts, but to refine, systematize, and expand the menu of possible explanations. While robustness analysis might look like one is playing around with a model without any empirical import, it actually helps scientists to understand how changes in outcomes depend on changes in assumptions and to identify the crucial characteristics of a model (such as

its tipping points) and their background conditions. Explaining particular facts may be the ultimate aim and main motivation for developing HPEs, but there is no reason to assume that sketching concrete causal scenarios is the only way to contribute to this project.

Many philosophical accounts of modeling do not have the means to make sense of this crucial role of abstract models. It is often suggested that these theoretical models are only explanatory or only apply in simple and highly idealized counterfactual situations (e.g., Cartwright 1983, 1999), or that they are merely explorations of an idealized model world (e.g., Hausman 1992). According to this view, Schelling's checkerboard model helps us see how it is possible that mildly discriminatory agents could bring about segregation in the checkerboard city (i.e., only in the model world). So, in this account, the model would provide explanatory insight only if the world were similar to the model constructed by Schelling. From this perspective the checkerboard model would produce a very limited sort of understanding about the causes of residential segregation. However, we claim that the epistemic import of abstract theoretical models is not limited to these fictional worlds.

Some philosophers (e.g., Cartwright 2009, Mäki 2009) think that models are abstract representations of complex phenomena: the simplifications and idealizations of theoretical models do not matter as the intention is only to represent intellectually isolated aspects of the phenomenon. We agree with this, but think that this view is still too limited. In our view, even if a model describes a causal mechanism that has no known instances, it might make an important contribution to our scientific understanding. Similarly, a model can be crucial to the proper understanding of a specific phenomenon even if it does not describe the actual cause influencing it.

The problem with these alternative views is that they do not fully recognize the epistemic import of highly abstract models such as Schelling's. They both focus on the individual model as an independent representation. As we have already pointed out, when an abstract model successfully

represents an explanatory relationship in some real-world phenomenon, it only does so against the background of a whole family of models. Take away this background and you also take away most of the explanatory understanding provided by the model. Similarly, the importance that scientists give to the exploration of their models can only be understood from this perspective. The purpose of playing around with a cluster of related models is to acquire knowledge about dependencies between different assumptions of the model. This understanding is not based on these models being representations of some particular empirical phenomena.

Rather than providing an understanding of particular empirical facts, theoretical models provide modal understanding: they tell how things could be. This is what happens with the checkerboard model. Alternatively, theoretical models can also tell what cannot happen by showing that certain mechanisms cannot generate the specified outcomes. The work on theoretical models also helps in articulating the essential elements and preconditions of causal mechanism schemes. The more detailed understanding of mechanisms makes it easier to adapt them to particular causal scenarios. These contributions explain (and at least partially justify) the modelers' feeling that they are contributing to scientific understanding by creating and examining abstract imaginary worlds.

Finally, abstract theoretical models improve the modeler's understanding of his or her epistemic tools. Scientists use external representations, such as models and simulations, as they are unable to represent the objects of their interest without them. This is mediated understanding: an object is understood via a tool that also needs to be understood. To use a model or a simulation as a representation of a real-world phenomenon, one has to know both the possibilities and limitations of one's epistemic tool. This understanding is often achieved by exploring the formal properties of the model or by experimenting with the simulation. Often these models and simulations are theoretical in the sense that they are not intended to represent any particular phenomena. An improved understanding of our epistemic tools provides us with a better understanding of abstract model worlds, which in turn makes it possible to make more correct what-if inferences about

real-world phenomena. In our view, theoretical exploration is not (necessarily) an idle activity motivated by the wish to avoid hard questions about the adequacy of modeling assumptions. Rather, it may be a crucial part of developing preconditions for the proper scientific understanding of real-world phenomena.

One of our central claims is that when modelers express their intuitive judgment that their models provide explanatory insight, this judgment is ultimately based on the above contributions to their understanding. Thus, understanding the nature of how-possibly explanations is crucial for making sense of theoretical models. And while the modelers might often overestimate these contributions (Ylikoski 2009), they are right in being sensitive to the changes in the menu of possible explanations.

5. Conclusion

In this paper we have presented a number of claims. First, we have argued that abstract theoretical models can better be understood in the context of other models. This is a claim about the methodology of the philosophy of modeling. The scientific contribution of an abstract model can only be fully understood if it is considered in the context of a family of related models (our cluster thesis 1) and of competing explanations for the same phenomenon (our cluster thesis 2).

Second, we have defended the view that theoretical modeling should be understood as how-possibly reasoning. Our key suggestion is that there are two kinds of how-possibly explanations: possible causal scenarios and causal mechanism schemes. Understanding their relations is a necessary step in making sense of the epistemic contribution of highly abstract models.

Thirdly, we have introduced the notion of the menu of possible explanations and argued that by modifying this menu, theoretical models can be epistemically relevant for the explanation of particular empirical phenomena, even in cases in which they do not identify the causal mechanism that is actually responsible for the effect to be explained. In fact, the main point of theoretical HPEs is to refine, systematize, and expand the menu of possible explanations, not to

explain any particular facts.

Finally, the idea that scientific understanding consists of an ability to make correct what-if inferences helps us to see how the above contributions of HPEs are indeed contributions to scientific understanding. First, the way in which HPEs influence the structure of the eliminative inference makes it clear that they influence the probability that we choose the correct causal scenario. Similarly, the way in which a new HPE might lead to new, more stringent standards for explanatory evaluation clearly increases our understanding. The more demanding standards make it more probable that we do not end up with an incorrect causal scenario. Third, as theoretical HPEs contribute to our general ability to construct an explanation, they also expand the scope of our explanatory understanding. They allow us to make what-if inferences about a wider set of phenomena.

While we have presented these points using only one example, Schelling's checkerboard model, we believe that our observations can be generalized to other abstract theoretical models that are taken to have explanatory import. One example that we can readily provide is models of the emergence of money as presented in Aydinonat (2008). However, we believe that our framework is more general and helpful in understanding other theoretical models in economics and biology.

References

Aydinonat, N. E. (2007). Models, conjectures and exploration: An analysis of Schelling's checkerboard model of residential segregation. *Journal of Economic Methodology*, *14*, 429–454. doi:10.1080/13501780701718680

Aydinonat, N. E. (2008). *The invisible hand in economics: How economists explain unintended social consequences*. Abington: Routledge.

Benard, S., & Willer, R. (2007). A wealth and status-based model of residential segregation. *Journal of Mathematical Sociology*, *31*, 149–174. doi:10.1080/00222500601188486

Bird, A. (2007). Inference to the only explanation. *Philosophy and Phenomenological Research*, *74*, 424–432. doi:10.1111/j.1933-1592.2007.00028.x

Brandon, R. (1990). *Adaptation and environment.* New Jersey: Princeton University Press.

Bruch, E., & Robert M. (2006). Neighborhood choice and neighborhood change. *American Journal of Sociology*, *112*, 667–709. doi:10.1086/507856

Bruch, E., & Robert M. (2009a). Preferences and pathways to segregation: Reply to van de Rijt, Siegel, and Macy. *American Journal of Sociology*, *114*, 1181–1198. doi:10.1086/597599

Bruch, E., & Mare, R. (2009b). Segregation dynamics. In P. Hedström & P. Bearman (Eds.), *The Oxford handbook of analytical sociology* (pp. 269–293). England: Oxford University Press.

Cartwright, N. (1983). *How the laws of physics lie*. Oxford: Clarendon Press.

Cartwright, N. (1999). *The dappled world: A study of the boundaries of science*. England: Cambridge University Press.

Cartwright, N. (2009). If no capacities then no credible worlds. *Erkenntnis*, *70*, 45–58. doi:10.1007/s10670-008-9136-8

Casti, J. L. (1989). *Alternate realities: Mathematical models of nature and man*. New York: Wiley and Sons.

Clark, W. A. V., & Fossett, M. (2008). Understanding the social context of the Schelling segregation model. *PNAS*, *105*, 4109–4114. doi:10.1073/pnas.0708155105

de Donato, X. & Zamora-Bonilla, J. (2009). Credibility, idealisation, and model building: An inferential approach. *Erkenntnis*, *70*, 101–118. doi:10.1007/s10670-008-9139-5

Dray, W. (1957). *Law and explanation in history*. England: Oxford University Press.

Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies: Social science from the bottom up*. Washington, DC: Brookings Institution Press.

Forber, P. (2010). Confirmation and explaining how possible. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *41*, 32–40. doi:10.1016/j.shpsc.2009.12.006

Fossett, M. (2006). Ethnic preferences, social distance dynamics, and residential segregation: Theoretical explorations using simulation analysis. *Journal of Mathematical Sociology*, *30*, 185–274. doi: 10.1080/00222500500544052

Fossett, M. (2011). Generative models of segregation: Investigating model-generated patterns of residential segregation by ethnicity and socioeconomic status. *Journal of Mathematical Sociology*, *35*, 114–145. doi:10.1080/0022250X.2010.532367

Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology & Philosophy*, *21*, 725–740. doi:10.1007/s10539-006-9054-6

Goering, J. (2006). Shelling redux: How sociology fails to make progress in building and empirically testing complex causal models regarding race and residence. *Journal of Mathematical Sociology*, *30*, 299–317. doi:10.1080/00222500500544144

Grüne-Yanoff, T. (2009). Learning from minimal economic models. *Erkenntnis*, *70*, 81–99. doi:10.1007/s10670-008-9138-6

Hausman, D. (1992). *The inexact and separate science of economics*. England: Cambridge University Press.

Hedström, P., & Ylikoski, P. (2010). Causal mechanisms in the social sciences. *Annual Review of Sociology*, *36*, 49–67.

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

Kuorikoski, J., & Lehtinen A. (2009). Incredible worlds, credible results. *Erkenntnis*, *70*, 119–131. doi:10.1007/s10670-008-9140-z

Kuorikoski, J., Lehtinen, A., & Marchionni C. (2010). Economic modeling as robustness analysis. *The British Journal for the Philosophy of Science*, *61*, 541–567. doi: 10.1093/bjps/axp049

Lee, F. S. (2003). William O. Coleman. Economics and its enemies: Two centuries of anti-economics. [Review of the book *Economics and its enemies: Two centuries of anti-economics*, by W. O. Coleman]. *History of Economics Review*, *38*(Summer). Retrieved from http://www.freepatentsonline.com/article/History-Economics-Review/179424926.html

Lipton, P. (1991). *Inference to the best explanation*. Abington: Routledge.

Macy, M. W., & van de Rijt, A. (2006). Ethnic preferences and residential segregation: Theoretical explorations beyond detroit. *Journal of Mathematical Sociology, 30*, 275–288. doi:10.1080/00222500500544086

Mäki, U. (2009). MISSing the world: Models as isolations and credible surrogate systems. *Erkenntnis*, *70*, 29–43. doi:10.1007/s10670-008-9135-9

Muldoon, R., Smith, T., & Weisberg, M. (2012). Segregation that no one seeks. *Philosophy of Science*, *79*, 38–62. doi:10.1086/663236

Pancs, R., & Vriend, N. J. (2007). Schelling's spatial proximity model of residential segregation revisited. *Journal of Public Econometrics*, *91*, 1–24. doi:10.1016/j.jpubeco.2006.03.008

Rosser, J. B. Jr. (1999). On the complexities of complex economic dynamics. *Journal of Economic Perspectives*, *13*(4), 169–92. doi:10.1257/jep.13.4.169

Schelling, T. C. (1969). Models of segregation. *American Economic Review*, *59*, 488–493. Retrieved from http://www.jstor.org/stable/1823701

Schelling, T. C. (1971a). Dynamic models of segregation. *Journal of Mathematical Sociology*, *1*, 143–186. doi:10.1080/0022250X.1971.9989794

Schelling, T. C. (1971b). On the ecology of micromotives. *The Public Interest*, *25*(Fall), 61–98. Retrieved from http://www.nationalaffairs.com/public_interest/detail/on-the-ecology-of-micromotives

Schelling, T. C. (1972). The process of residential segregation: Neighbourhood tipping. In A. H. Pascal (Ed.), *Racial discrimination in economic life* (pp. 157–185). Idaho Falls, ID: Lexington

Books.

Schelling, T. C. (1978). *Micromotives and macrobehavior*. London & New York, NY: W. W. Norton.

Sugden, R. (2000). Credible worlds: The status of theoretical models in economics. *Journal of Economic Methodology*, *7*, 1–31. doi:10.1080/135017800362220

Sugden, R. (2009). Credible worlds, capacities and mechanisms. *Erkenntnis*, *70*, 3–27. doi:10.1007/s10670-008-9134-x

van de Rijt, A., Siegel, D., & Macy M. (2009). Neighborhood chance and neighborhood change: A comment on Bruch and Mare. *American Journal of Sociology*, *114*, 1166–1180. doi:10.1086/588795

von Wright, G. H. (1971). *Explanation and understanding*. Abington: Routledge.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. England: Oxford University Press.

Woodward, J. (2006). Some Varieties of Robustness. *Journal of Economic Methodology*, *13*, 219–240. doi:10.1080/13501780600733376

Ylikoski, P. (2009). The illusion of depth of understanding in science. In H. De Regt, S. Leonelli, & K. Eigner (Eds.), *Scientific understanding: Philosophical perspectives* (pp. 100–119). Pittsburgh: Pittsburgh University Press.

Ylikoski, P. (2011). Social mechanisms and explanatory relevance. In P. Demeulenaere (Ed.), *Analytical Sociology and Social Mechanisms* (pp. 154–172). England: Cambridge University Press.

Ylikoski, P. (2013). Agent-based simulation and sociological understanding. *Perspectives on Science*, forthcoming.

Ylikoski, P., & Kuorikoski J. (2010). Dissecting explanatory power. *Philosophical Studies*, *148*, 201–219. doi:10.1007/s11098-008-9324-z