Exploring Scientific Inquiry via Agent-Based Modeling

Dunja Šešelja

Munich Center for Mathematical Philosophy, LMU Munich
dunja.seselja@lrz.uni-muenchen.de

Abstract

In this paper I examine the epistemic function of agent-based models (ABMs) of scientific inquiry, proposed in the recent philosophical literature. In view of Boero and Squazzoni's (2005) classification of ABMs into case-based models, typifications and theoretical abstractions, I argue that proposed ABMs of scientific inquiry largely belong to the third category. While this means that their function is primarily exploratory, I suggest that they are epistemically valuable not only as a temporary stage in the development of ABMs of science, but by providing insights into theoretical aspects of scientific rationality. I illustrate my point with two examples of highly idealized ABMs of science, which perform two exploratory functions: Zollman's (2010) ABM which provides a proof-of-possibility in the realm of theoretical discussions on scientific rationality, and ArgABM (Borg et al., 2017b, 2018a,b), which provides insights into potential mechanisms underlying the efficiency of scientific inquiry.

1 Introduction

Computational modeling has in recent years become an increasingly popular method for the study of social aspects of scientific inquiry. In particular, agent-based models (ABMs) have been used as simulations of scientific inquiry, allowing for the examination of various socio-epistemological issues: from tensions between individual and group rationality (Grim, 2009; Grim et al., 2013; Zollman, 2007, 2010), to different social mechanisms that impact the efficiency of inquiry (Holman and Bruner, 2015; Weatherall, O'Connor, and Bruner, 2018), to different research strategies (Alexander, Himmelreich, and Thompson, 2015; Pöyhönen, 2017; Weisberg and Muldoon, 2009), etc. A common feature of ABMs developed in philosophy of science is that they are simple, 'thin' representations of scientific inquiry (Pöyhönen and Kuorikoski, 2016). The primary appeal of such models is that they allow for an easy insight into possible causal mechanisms underlying the phenomenon in question. The less components a model includes, the easier it gets to study causal dependencies between the given components. Nevertheless, such simplicity comes at a price: the model will likely end up being highly idealized, making it difficult to determine its relation to the real world. More precisely, the more idealized a model is, the harder it gets to exactly determine target phenomena it represents.

Despite their highly idealized character, many of the ABMs proposed in the literature have been motivated by concrete episodes from the history of science, suggesting potential explanations of the given cases (Holman and Bruner, 2015; O'Connor and Weatherall, 2017; Weatherall, O'Connor, and Bruner, 2018; Zollman, 2010). This has had two significant consequences for the reception of ABMs of science. On the one hand, these models have been considered to be primarily aiming at *explaining* real-world phenomena or at least providing 'how-possibly explanations' or 'proofs of possibility' that should be applicable to the given cases. On the other hand, the lack of robustness analysis of the given findings has cast doubt on their link to real-world phenomena, and hence on the relevance of these results for actual scientific inquiry (even in a how-possibly way).[1] As a result, it has been suggested that the vast majority of ABMs developed in philosophy of science are currently only exploratory, rather than explanatory (Frey and Šešelja, 2018b), and that they need to be 'thickened' and enhanced by empirical data to provide insights into actual scientific inquiry (Martini and Pinto, 2016).

Altogether, the current state of ABMs of science is such that it is unclear what their exact function is, unless we link them more directly to real-world phenomena.[2] The aim of this paper is twofold. On the one hand, I examine different strategies of relating ABMs of science to actual scientific practice and I specify challenges that these strategies face. On the other hand, I introduce a classification of ABMs of science, which will help in providing

[1]Similar worries have been directed at ABMs in social sciences, see e.g. Arnold, 2014.

[2]Beside explanation and exploration there may be various other functions of ABMs, see Edmonds, 2017; Edmonds et al., 2018.

a clear epistemic function to highly-idealized exploratory models without necessarily assigning them a mere temporary stage. The basic idea behind the latter is that ABMs may be explanatory of theoretical phenomena, which are nevertheless interesting as conceptual explorations of scientific rationality taken *in abstracto*.

I will start with Boero and Squazzoni's (2005) classification of ABMs in general, and with a critical look to it, classify ABMs of scientific inquiry (Section 2). In Section 3 I discuss some of the central procedures necessary for relating these ABMs to actual scientific practice. In Section 4 I assess ABMs proposed in the philosophical literature in view of the preceding discussion and argue that they are theoretical abstractions, performing exploratory roles. In Section 5 I illustrate these functions in terms of two examples: Zollman's (2010) model, and the argumentative ABM of scientific inquiry (Borg et al., 2017a,b, 2018a,b). While the first model plays the role of providing a proof of possibility in the realm of scientific rationality, the second model plays the role of pointing to new hypotheses about mechanisms relevant for the efficiency of scientific inquiry. Section 6 concludes the paper.

2 Classifying ABMs of science

To explicate the difference between simple (or 'thin') and complex (or 'thick') models,[3] it is useful to take a look at the classification of ABMs introduced by Boero and Squazzoni, 2005, in terms of the degree of specificity of the purported target, and the degree of complexity of the given model. The authors distinguish between case-based ABMs, typifications and theoretical abstractions—on the spectrum ranging from complex and specific models to simple and general ones. While the two dimensions of classification—specificity and complexity—may not necessarily coincide (as I'll argue later on in this section), of special interest for our current purposes is the dimension of specificity of the purported target. Let's take a closer look at each of these categories and see how they apply to ABMs of science.

On one end of the spectrum, we have *case-based ABMs* of scientific inquiry, aiming to represent concrete cases from the scientific practice. Their target is thus a phenomenon that is restricted in terms of space and time and other empirical information, specific for the given case of scientific inquiry. As such, these ABMs tend to be 'thick' representations of science since they are calibrated towards the given empirical scenario and hence,

[3]Thanks to Daniel Singer for bringing to my attention adjectives 'thin' and 'thick', which are sometimes used to describe simple and complex ABMs.

they will include a variety of details relevant for it.[4]

The second type of ABMs—*typifications*—aim to represent a *class* of empirical phenomena. They aim at capturing key properties of the given type, while abstracting away from particularities of each individual phenomenon. Hence, a model designed to represent a certain type of scientific inquiry (for instance, an inquiry characterized by deceptive information sharing among scientists, or an inquiry that occurs in the context of theoretical diversity), would fall in this category, characterized by mid-level specificity.

Finally, at the other end of the spectrum we have *theoretical abstractions*: these are simple, highly idealized models, which often have purely exploratory function: testing new ideas, extending existing frameworks, etc.[5] As I suggest below, most ABMs in recent philosophical literature fall into this category.[6]

While Boero and Squazzoni assume that complexity correlates with specificity, and simplicity with generality, this may not necessarily be the case. As Bruce Edmonds has argued (Edmonds and Moss, 2004), simplifying a model won't necessarily make it more general. Edmonds suggests that simplification will lead to a greater generality of the model only if it satisfies one of the following conditions:

> - When what is simplified away is essentially irrelevant to the outcomes of interest (e.g. when there is some averaging process over a lot of random deviations)
> - When what is simplified away happens to be constant for all the situations considered (e.g. gravity is always 9.8m/s^2 downwards)
> - When you loosen your criteria for being approximately right hugely as you simplify (e.g. mover from a requirement that results match some concrete data to using the model as a vague analogy for what is happening)

[4] For instance, an ABM of scientific collaboration presented by Zamzami and Schiffauerova, 2017, which aims at examining knowledge transmission and productivity of inquiry, is calibrated on nanotechnology journal publications in Canada, and hence, it can be considered a case-based ABM.

[5] For a discussion on different exploratory functions of models see Gelfert, 2016, Chapter 4, for exploratory functions of simulations see Arnold, 2008, Chapter 6 (see also Footnote 18).

[6] The most common example of this class of ABMs in social sciences is Schelling's model of social segregation (Schelling, 1971).

> In other cases, where you compare like with like (i.e. you don't move the goalposts such as in (3) above) then it only works if you happen to know what can be safely simplified away. (Edmonds, 2018)

Hence, the link from simplicity to generality requires evidence, or at least an explication showing that one of the above conditions has been satisfied. Furthermore, complex models are not necessarily more specific. A higher number of parameters may increase the complexity of a model, but there is no reason to assume that this implies a narrower scope of the target phenomenon. Whether the target is narrow or wide is an independent question, which can be determined only by means of adequate validation procedures. The next section tackles this issue.

3 ABM validation: determining the adequate target

The literature on the validation of ABMs is vast (for recent discussions see Casini and Manzo, 2016; Gräbner, 2018; Thicke, 2018). The aim of this section is not to give an exhaustive list of validation methods suggested to this end.[7] For the current purposes it will suffice to give an overview of some of the central strategies that have been suggested in this context, and which are necessary if we wish to argue that a given ABM of scientific inquiry is a case-based model or a typification. For either of these ABM types, what needs to be validated is the link between the model and a given class of empirical phenomena (in case of typifications) or a concrete phenomenon (in case of case-based ABMs). Such a phenomenon is specified in the interpretation of the ABM as its purported target. The following two types of mutually interwoven processes represent central methodologies of validation: on the one hand, robustness analysis, and on the other hand, the process of empirically embedding the model.

Before I turn to them, a brief look at possible targets of simulations is in place. First, we can imagine a simulation that is descriptively adequate (e.g. representing a particular case-study) and which may be of sociological interest. However, such a simulation will be philosophically interesting only if we can use it to examine counterfactual dependencies, in view of which we can draw normative conclusions about the given phenomenon. In other words,

[7] In addition to validation, ABMs also require a process of *verification*, the aim of which is to evaluate how accurate the program of the model is (see e.g. Cooley and Solano, 2011).

simulations become useful for philosophical purposes once they represent a space of possibilities. Since such possibilities can range from empirical to merely logical ones,[8] the main challenge for drawing information that is relevant for the real-world phenomena is determining *which kind of possibility* the model represents. That is, unless we can specify which context precisely a given model represents, it will be impossible to draw information concerning counterfactual dependencies, such that we can reliably relate them to real-world phenomena. Unsurprisingly, this is especially challenging in case of highly idealized models, where it is unclear how 'counterfactually distant' from the real world a given model is. The following procedures help in addressing this challenge.

3.1 Robustness analysis

Two particularly relevant types of robustness analysis are the following ones:

a) *Sensitivity analysis*: this is a method of examining the robustness of results under changes in parameters of the model (Thiele, Kurth, and Grimm, 2014). It is used to determine the scope of parameters within which results of simulations remain stable.

b) *Derivational robustness analysis*: this is a method of examining the robustness of results under the changes in idealizing assumptions of the model (Lehtinen, 2017; Railsback and Grimm, 2011; Ylikoski and Aydinonat, 2014, p. 302-306). Since the robustness analysis of this kind can be rather complex and tedious, rather than starting from the model and altering each of its assumptions separately, it may be more efficient to examine the robustness of results by employing an altogether new model, which is structurally different, while it aims at the same target phenomenon.

While the robustness analysis serves to explore the stability of results under various changes of the model, an output of such an analysis will not necessarily help us in validating the link between the model and its purported target. On the one hand, if the robustness analysis shows that the results are highly stable and that they hold under numerous changes, this may indicate that the model is representative of a large scope of phenomena, which makes it more likely that it also represents the specific target in question. On the other hand, the results may indeed be sensitive to a variety of parameter changes, in which case we need some interpretative tools to determine which results are representative of which phenomenon exactly. For instance, if it turns out that the results of a certain ABM of scientific inquiry hold only

[8] See Verreault-Julien, 2018 for a discussion on different types of how-possibly explanations and their relation to the represented possibilities.

under certain parameter values, then we need some way of translating the parameters in the model into real-world values in order to determine what exactly the model tells us about actual scientific practice. In other words, what is needed is a link between the model and empirical information. This is done by empirically embedding the model.

3.2 Empirical embeddedness

In order to guide the robustness analysis towards the purported target phenomena of case-based ABMs and typifications, our methods need to be empirically embedded. Casini and Manzo, 2016 suggest three trends in the literature on ABMs, which have been used to this end:[9] a trend towards theoretical realism, a trend towards empirical calibration and a trend towards empirical validation. Let's see how these strategies apply to the validation of ABMs of scientific inquiry.

a) The first strategy—*enhancing theoretical realism*—suggests that models be built in view of relevant psychological and sociological theories (Casini and Manzo, 2016, p. 23). We may add that with respect to ABMs of science, an important source of such theoretical background are accounts proposed in traditional methodology of science and social epistemology. Hence, when deciding how to represent different aspects of inquiry, and which simplifications may turn out problematic, we may profit from ideas proposed in this literature. For instance, if we are modeling decision-making of scientists concerning which lines of inquiry they are to pursue, different assessments discussed by philosophers of science can be useful.[10] Similarly, if a model aims at representing the context of scientific disagreements, insights from the literature on peer disagreement may be helpful.[11] This information can be

[9]More precisely, the authors discuss these trends as conducive to the validation of models and their capacity to warrant causal inference. They also note that due to practical limitations these strategies may not always be available, in which case ABM validation should proceed by means of theoretical explorations, which include the two above mentioned robustness analysis, as well as 'dispersion analysis' (the examination of the stochastic character of the results) and 'model analysis' (the examination of events, behaviors and feedbacks executed within the model, which aim towards transparency of the coded processes).

[10]For instance, Nickles' (2006) distinction between the heuristic appraisal and epistemic appraisal motivates the behavior of agents in an argumentation-based ABM of science (Borg et al., 2017a,b, 2018a).

[11]For example, Douven, 2010 builds an ABM to examine some of the norms suggested in the informal literature on peer disagreements. Moreover, future ABMs of scientific disagreements could make use of the notion of higher-order evidence and its role in scientific disagreements (Straßer, Šešelja, and Wieland, 2015), which has so far largely remained

used for guiding the construction of both models tackling novel phenomena, and models aimed at examining the robustness of existing ABMs.

b) The second strategy—*empirically calibrating ABMs*—consists in using concrete numerical information as an input for parameters of the model. In other words, empirical data is used to build micro-specifications of a given ABM (Boero and Squazzoni, 2005), either when constructing novel models or when examining the robustness of the existing ones. Such data can range from the number of agents that represent a given scientific community, to their specific distribution on the given epistemic landscape, to the time span of a given inquiry, to the epistemic success of the represented scientific theories, etc. There are different possible sources of such information. First, historical knowledge about scientific episodes may be essential for an adequate representation of a given case-study by a case-based ABM.[12] Similarly, historical information about different episodes may be informative of ABMs that aim to be typifications (for instance, representing inquiry in a certain scientific discipline). Second, sociological studies about past or contemporary scientific episodes may provide valuable data for empirical calibration of ABMs. Finally, recent study of bibliometric data is a particularly promising avenue for empirical calibration of ABMs of science. Be it case-based ABMs or typifications, bibliometric data may serve as an input for an average number of agents in a given domain, their distribution across different sub-topics, their citation behavior (which may suggest lines of communication and type of social networks across the given community), their relative impact (in terms of citations), etc. (see e.g. Martini and Pinto, 2016; Perović et al., 2016; Thicke, 2018).[13]

c) The third strategy—*empirically validating ABMs*—consists in the analysis and comparison of a simulated macro behavior with the real-world macro behavior (Boero and Squazzoni, 2005). For instance, if a model aims at representing a certain episode from the history of science, then given specific initial conditions, the macro behavior of simulated agents should correspond to the historical knowledge about the given case-study. Such a procedure may be especially useful if combined with theoretical realism and empirical calibration, which can together guide the sensitivity analysis (by examining the macro behavior for the parameter values relevant for the given case or type of inquiry), and the derivational robustness analysis (by examining the

absent from ABMs of science.

[12]For a recent example see Frey and Šešelja, 2018b who use historical information about the mid-twentieth century research on peptic ulcer disease to examine the robustness of Zollman's (2010) ABM as a case-based model.

[13]The above mentioned model by Zamzami and Schiffauerova, 2017 is a case in point.

output given the assumptions relevant for the given case or type of inquiry).[14]

4 Taking stock of ABMs of science: valid unless proven otherwise?

Looking at ABMs developed in the field of philosophy of science[15] we may notice that the vast majority of them are designed as intentionally simple: by simplifying the representation of scientific inquiry in terms of factors that are included in the model we can have an easier insight into causal dependencies between these factors. Moreover, how general or specific they are has remained largely open. On the one hand, as mentioned in Section 1, these models have frequently been taken as providing potential explanatory mechanisms of concrete episodes from the history of science. On the other hand, validation procedures establishing the link between models and respective case studies have been typically omitted: the majority of them have not been subjected to any systematic robustness analysis or the process of empirical embedding, which would relate them to concrete cases of scientific inquiry.

Moreover, for those ABMs that *prima facie* display a plausible social mechanism and for which no further validation seems necessary, there is typically the danger of triviality. If the simulation displays a process which plausibly holds for the real-world phenomenon *in virtue* of our knowledge of that phenomenon, where any further robustness analysis of the simulation seems unnecessary, then we may ask: why run the simulation in the first place? As Arnold, 2008 argues: "the results [of simulations] should not already be deducible without any model or simulation from the empirical description of the process." (p. 191).

Hence, each ABM lacking robustness analysis (i.e. sensitivity analysis and derivational robustness analysis) cannot be reliably considered a non-trivial case-based model or a typification. In other words, the assumption of validity unless proven otherwise is unsuitable as a guideline for the interpretation of non-trivial ABMs. Just like with any scientific claim, justification needs to proceed in terms of evidence (obtained by reliable methods), where the burden of proof is on those proposing an interpretation of a model, which assigns it a representational power with respect to real-world phenomena. The reason why validity cannot be granted without robustness-based evidence lies in the highly idealized nature of these models. Even if the model

[14]See, for example, Frey and Šešelja, 2018b for some initial steps in this direction.

[15]For instance, all the models listed in Section 1.

seems plausible, some of its underlying assumptions may be representative only of a relatively small subset of the given target phenomenon.[16]

In contrast, theoretical abstractions may not be representative of any real-world phenomena, but they may rather show how a given socio-epistemic mechanism occurs under certain conditions, even though such conditions may be unlikely in any realistic context of scientific inquiry, and they may moreover offer no clear ways of making interesting counterfactual inferences about real-world phenomena.[17]

Altogether, this means that the majority of these models fall into the category of theoretical abstractions, performing only exploratory function.[18] As a result, they cannot be reliably used for making inferences about any concrete real-world phenomena. More precisely, some consequences of the current exploratory status of ABMs of science are:

1. *They cannot be reliably used as case-based models:* this means that without systematic validation procedures (and possibly also further enhancements) these ABMs are insufficiently reliable as representations of any concrete case of scientific inquiry.

2. *They cannot be reliably used as typifications:* this means that without systematic validation procedures (and possibly also further enhancements) these ABMs are insufficiently reliable as representations of any concrete case of scientific inquiry.

[16]For instance, different representations of knowledge acquisition may result in strikingly different outcomes: e.g. representing scientists as gathering information by making pulls (each of which is success or a failure) from a given probability distribution as in Zollman-inspired ABMs may lead to different findings than if agents gather information about different *parts* of the given rivaling theories, where some parts of each theory are defensible or indefensible (see e.g. Borg et al., 2018b).

[17]It is important to notice a difference between these two types of results. On the one hand, ABMs may generate valuable normative insights about real-world phenomena by representing certain counterfactual scenarios. On the other hand, they may be so idealized and simplified to the point of being *too counterfactually distant* from actual scientific practice, so that no relevant inferences about actual science can be drawn from the model.

[18] Arnold, 2008 classifies simulations according to their purpose, dividing them into those employed at the 'conceptual level' and those employed at the 'application level'. The former are then distinguished into proof-of-possibility simulations and exploratory simulations, while the letter include predictive simulations and explanatory simulations (p. 187). In the current paper I use the term 'exploratory' in the sense of Arnold's first larger category (conceptual level simulations) mainly because proof-of-the-principle ABMs tend to be toy-models, not necessarily referring to any real-world phenomena, and as such, they are exploratory of the given conceptual space.

3. What is more, making these models *more complex* won't necessarily make them more adequate candidates for the above two categories.

While claims 1. and 2. are supported by the discussion in the previous and this section, let me turn now to claim 3. Even though the majority of highly-idealized ABMs in the philosophical literature are simple in character, making these models more complex—by adding assumptions and parameters—isn't a straightforward path to their validation. Just like in case of simple ABMs, the link between complex models and their real-world targets needs to be warranted if the model is to become a case-based ABM or a typification. Moreover, there is a trade-off when it comes to the kinds of robustness analysis which pose a challenge to the validation of simple or complex models. On the one hand, in case of simple models, a small number of parameters makes the sensitivity analysis relatively easy. In contrast, derivational robustness analysis presents a challenge for simple models since it requires a construction of new ABMs (either as enhancements of the existing one or as newly constructed models). On the other hand, the situation with complex models is reverse: while derivational robustness analysis can be easy at least in the sense of removing certain assumptions from the model, sensitivity analysis poses a challenge to complex ABMs due to a large number of parameters. To elevate this difficulty, different types of screening procedures, which facilitate sensitivity analysis, have been suggested in the literature on simulations (see Thiele, Kurth, and Grimm, 2014).

5 What can we learn from theoretical abstractions?

If theoretical abstractions primarily serve an exploratory function (either by exploring a conceptual space and providing a proof-of-principle, or by exploring a causal space of mechanisms relevant for real-world phenomena), what exactly can we learn from them? In this section I illustrate insights of this kind by means of two highly-idealized ABMs of science: one of which has provided a proof-of-possibility in the realm of theoretical discussions on the tension between the individual and group rationality, and one of which has provided novel insights into possible mechanisms underlying the efficiency of scientific inquiry.

5.1 Zollman's modeling in view of bandit problems

One of the most prominent classes of ABMs of science are Zollman's (2007; 2010) models. Inspired by (Bala and Goyal, 1998), the models represent scientific inquiry by employing the so-called bandit problems. Bandit problems, well known in the field of statistics and economics, concern situations in which a gambler (or a group of gamblers) is trying to maximize their payoff when confronted with multiple slot machines (bandits) that have different probabilities of success. Analogously, we can imagine a scientist confronted with multiple rivaling hypotheses, trying to determine which one is the best. At the beginning of the simulation[19] scientists are assigned random prior probabilities for two rivaling hypotheses, each of which has a designated objective probability of success, unknown to the agents-scientists. Agents always choose to pursue a theory which they consider to be better. Throughout the simulation they update their beliefs in view of their own findings, and by receiving information from other agents with whom they are connected in a social network.[20] Zollman employs three types of social networks (the so-called cycle, wheel and the complete graph) and his results suggest that the degree of connectedness of the scientific community is inversely proportional to the success of scientists in converging on the objectively better hypothesis. The reason why a fully connected community often fails in converging on the best hypothesis is that initial findings by scientists may be misleading, but due to the full connectedness, they may spread quickly throughout the whole community, resulting in a premature abandonment of the objectively better hypothesis.

Zollman's model is indeed highly-idealized: it abstracts away from different types of interactions among scientists (e.g. mere exchange of obtained evidence vs. critical interaction), it assumes that scientists always prefer a better hypothesis without any inertia towards their previous choice of inquiry, etc. Moreover, the results of his runs pass neither sensitivity analysis nor derivational robustness analysis. On the one hand, Rosenstock, O'Connor, and Bruner, 2017 have shown that as soon as certain parameter values are slightly changed (e.g. the values for the objective probability of success assigned to two hypotheses), all networks are equally successful. On the other hand, Frey and Šešelja, 2018a have shown that changing some of the assumptions in the model (e.g. adding the idea that scientists don't abandon their current hypothesis as soon as they learn that the rivaling one

[19]I describe here Zollman's (2010) model, which is a generalized version of his (2007) one.

[20]Every round an agent makes 1,000 pulls, each of which can be a success or failure, where the probability of success is given by the objective probability of success of the respective theory. Agents then update their beliefs via Bayesian reasoning (modeled by means of beta distributions), in view of their own success and the success of other agents with whom they are linked in a social network.

is superior since they have a 'rational inertia' towards the former) changes Zollman's results as well. In view of these findings, it is difficult to argue that Zollman's ABM represents a typical context of scientific inquiry, or that we can draw from it normative conclusions about actual scientific practice.

Nevertheless, the model still provides philosophically valuable insights: it illustrates a case of the tension between individual and group rationality as a theoretical notion, that is, irrespective of how realistic such a scenario is.[21] As such, it contributes to the conceptual exploration of rationality, where it is legitimate to push some assumptions to extreme in order to observe their consequences. We do the same in case of epistemic modal logic, for instance, which includes the 'positive introspection axiom', according to which if one knows p then one knows that one knows p, and which has been criticized as highly problematic (see e.g. Williamson, 2002, p. 114-134). Yet, epistemic logic is nonetheless considered a valuable contribution to attempts at formally modeling knowledge. As explorations of rationality these models are *philosophically* valuable since they tell us something about the theoretical phenomena they represent. In case of Zollman's model, we have learned about conditions under which a certain degree of connectedness of a social network, representing information flow among scientists, may be epistemically harmful. Whether these conditions ever occur in the real world is a separate question.

In this sense, Zollman's ABM provides a proof-of-possibility that individual and group rationality may not always go hand-in-hand, and that communication structure may be an underlying mechanism leading to this tension. While his models can of course serve the purpose of grounding further enhancements that aim at providing normatively relevant information about actual scientific research, it doesn't get its epistemic value primarily from such a temporary heuristic role. To the contrary: I suggest that its primary role consists in explaining the above mentioned theoretical phenomenon.

5.2 Argumentation-based ABM (ArgABM)

Inspired by Abstract Argumentation Frameworks,[22] Borg et al.'s (2017; 2017; 2018) model represents scientific inquiry as an argumentative exchange be-

[21] I am indebted to Christian Straßer for an inspiring discussion that resulted in a number of ideas appearing in this section.

[22] Abstract Argumentation Frameworks were pioneered by Dung, 1995 and previously used for the modeling of scientific debates by Šešelja and Straßer, 2013. In what follows I give an informal overview of ArgABM. For all the details see the original papers on the model.

tween scientists pursuing rivaling research programs. Throughout each run of the simulation agents-scientists explore an 'argumentative-landscape', gradually discovering arguments in favor or against their current theory. Each theory (or a research program) is represented as consisting of a number of arguments. These arguments are represented abstractly, as nodes in a directed graph, connected via a 'discovery relation'. The discovery relation represents paths that agents take when moving on the landscape, from one argument to another. Moreover, arguments belonging to one research program can attack arguments of one of the rivaling programs. The landscape then consists of different argumentative (rooted) trees, with nodes as arguments[23] and edges as discovery relation, where an argument in one tree may attack an argument in another tree (see Figure 1).

Roughly speaking, an argument in a theory is considered defended if it is not attacked, or if there is another defended argument in the same theory, which attacks the attacker-argument.[24] Moreover, the landscape is pre-defined in such a way that one theory is the best in the sense that it is fully defended from all its attackers in the objective landscape. Over the course of a run, agents gather knowledge about the objective landscape, which consists of arguments in favor of each theory, and attacks on these arguments. In addition to gathering knowledge on their own, they also learn about the landscape from other agents with whom they are linked in a social network. In view of this knowledge, agents evaluate the theories. A run is successful if agents manage to converge on the objectively best theory.[25]

As the authors explicitly state, ArgABM was primarily designed as an exploratory model, aimed at testing the robustness of previously proposed ABMs of scientific interaction (such as Zollman's ones). While it has more parameters than Zollman-inspired ABMs, and it allows for more assumptions to be integrated in the model, it still lacks the validation procedures discussed in Section 3.

The results obtained by ArgABM suggest that more connected groups perform significantly better than the less connected ones under a variety of

[23]All theories are trees of the same size, i.e. consisting of the same number of arguments.

[24]More precisely, we call a subset of arguments A of a given theory T *admissible* iff for each attacker b of some a in A there is an a' in A that attacks b. Since every theory in the model is conflict-free (in the sense that no two arguments in the same theory attack one another), it can easily be shown that for each theory T there is a unique maximally admissible subset of T (with respect to set inclusion). An argument a in T is said to be *defended in* T iff it is a member of this maximally admissible subset of T.

[25]The model employs an additional, more permissive criterion of success, according to which agents are successful if the best theory is at least as populated as any of the rivaling theories.

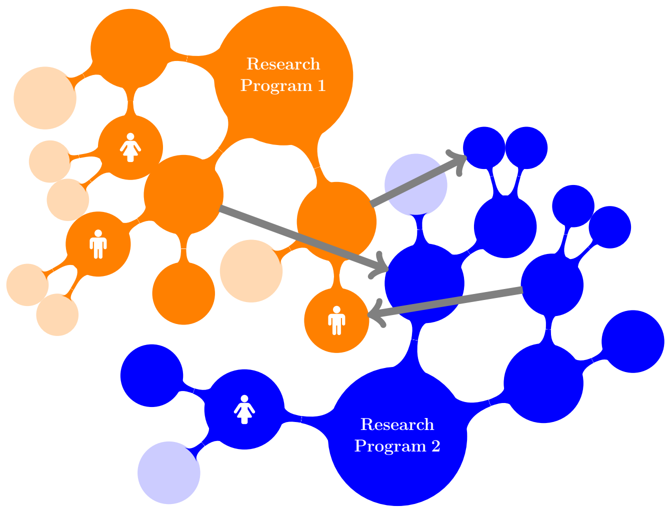

Figure 1: An example of an argumentative landscape consisting of 2 theories (or research programs). Darker shaded nodes represent arguments that have been investigated by agents and are thus visible to them; brighter shaded nodes stand for arguments that aren't visible to agents. The biggest node in each theory is the root argument, from whcih agents start their exploration via discovery relation, which connects arguments within one theory. Arrows stand for attacks from an argument in one theory to an argument in another theory (Borg et al., 2018b).

conditions. And while these findings appear contrary to conclusions drawn from Zollman's ABMs, the authors caution that structural differences between these ABMs may indicate that each model represents a specific *kind of* inquiry. This is the first important exploratory result of ArgABM: different ABMs of science may be representing different types of inquiry, which means that more research should be done towards determining the specific context of inquiry represented by each of the models.

Second, a recent iteration of ArgABM by Borg et al., 2018b examines different evaluation procedures underlying theory-choice performed by scientists. For instance, scientists may prefer a theory that has more defended arguments than its rivals. Alternatively, they may prefer a theory that has a lower number of undefended arguments (which can be understood as anomalies in the given theory). As the authors show, these assessments may result in different preference orders on the given theories, and in strikingly different outcomes of the simulation. Hence, this result is a novel insight into a potential mechanism that may impact the efficiency of scientists. While the link between the modeled evaluations and actual scientific practice (i.e. evaluations underlying theory-choice performed by actual scientists) remains an open question, this result points to the significance of a factor that has previously often been omitted from ABMs of science.[26]

Before closing this section, let me add that one could argue that just like Zollman's model, ArgABM actually provides a proof-of-possibility for different theoretical phenomena (the relation between the degree of connectedness of scientists and their efficiency, the relative performance of different evaluation procedures underlying theory-choice, etc.). Indeed, the two exploratory functions discussed here are closely related: a proof-of-possibility may also be an insight into a potential novel mechanism, and the other way around. In fact, the exact function a model performs is largely a matter of the context in which it is proposed, including initial motivations for developing the model and ways in which it is employed: as the first model tackling a given question, a model designed to test the robustness of previous proposals, etc.

6 Conclusion

In this paper I have analyzed a class of highly-idealized ABMs of scientific inquiry, proposed in the literature in philosophy of science and social epistemology, suggesting they should be considered as exploratory models. To this end, I have argued that the majority of these models belong to the category of theoretical abstractions, which means that what they represent is unspecified. As a result, these models cannot be reliably used for drawing inferences about actual scientific practice, at least as long as they don't pass an adequate validation procedure. Such a procedure consists in different types of robustness analysis, guided by the process of empirically embedding the model towards its purported target. Nevertheless, I have argued that in the lack of such validation procedures, theoretical abstractions are still epistemically valuable by being informative of theoretical phenomena, i.e. by providing conceptual insights about scientific rationality and the process of scientific inquiry.

Finally, it is important to notice that insights obtained by such exploratory ABMs might have been overlooked were these models immediately calibrated towards concrete cases of scientific inquiry. This suggests that exploratory ABMs are not just a preliminary stage in the development of

[26] An exception is the epistemic landscape ABM proposed by Currie and Avin, 2018, which represents the diversity of methods preferred by scientists during their inquiry.

empirically validated models, merely providing the ground for an eventual epistemic benefit. While they do play that role as well, their additional function consists in providing conceptual explorations of scientific rationality and revealing potential causal mechanisms underlying scientific inquiry, which may remain a blind spot in those ABMs that are immediately informative of real-world phenomena.

References

Alexander, Jason McKenzie, Johannes Himmelreich, and Christopher Thompson (2015). "Epistemic landscapes, optimal search, and the division of cognitive labor". In: *Philosophy of Science* 82.3, pp. 424–453.

Arnold, Eckhart (2008). *Explaining altruism: A simulation-based approach and its limits*. Vol. 11. Walter de Gruyter.

— (2014). "What's wrong with social simulations?" In: *The Monist* 97.3, pp. 359–377.

Bala, Venkatesh and Sanjeev Goyal (1998). "Learning from neighbours". In: *The review of economic studies* 65.3, pp. 595–621.

Boero, Riccardo and Flaminio Squazzoni (2005). "Does empirical embeddedness matter? Methodological issues on agent-based models for analytical social science". In: *Journal of artificial societies and social simulation* 8.4.

Borg, AnneMarie, Daniel Frey, Dunja Šešelja, and Christian Straßer (2017a). "An Argumentative Agent-Based Model of Scientific Inquiry". In: *Advances in Artificial Intelligence: From Theory to Practice: 30th International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2017, Arras, France, June 27-30, 2017, Proceedings, Part I*. Ed. by Salem Benferhat, Karim Tabia, and Moonis Ali. Cham: Springer International Publishing, pp. 507–510. ISBN: 978-3-319-60042-0.

— (2017b). "Examining Network Effects in an Argumentative Agent-Based Model of Scientific Inquiry". In: *Logic, Rationality, and Interaction: 6th International Workshop, LORI 2017, Sapporo, Japan, September 11-14, 2017, Proceedings*. Ed. by Alexandru Baltag, Jeremy Seligman, and Tomoyuki Yamada. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 391–406.

— (2018a). "Epistemic effects of scientific interaction: approaching the question with an argumentative agent-based model". In: *Historical Social Research* 43.1, pp. 285–309.

Borg, AnneMarie, Daniel Frey, Dunja Šešelja, and Christian Straßer (2018b). "Theory-choice, transient diversity and the efficiency of scientific inquiry". In: *Forthcoming*.

Casini, Lorenzo and Gianluca Manzo (2016). "Agent-based models and causality: a methodological appraisal". In: *(The IAS Working Paper Series). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-133332*.

Cooley, Philip and Eric Solano (2011). "Agent-based model (ABM) validation considerations". In: *Proceedings of the Third International Conference on Advances in System Simulation (SIMUL 2011)*, pp. 134–139.

Currie, Adrian and Shahar Avin (2018). "Method Pluralism, Method Mismatch & Method Bias". In: *Philosopher's Imprint*.

Douven, Igor (2010). "Simulating peer disagreements". In: *Studies in History and Philosophy of Science Part A* 41.2, pp. 148–157.

Dung, Phan Minh (1995). "On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games". In: *Artificial Intelligence* 77, pp. 321–358.

Edmonds, Bruce (2017). "Different Modelling Purposes". In: *Simulating Social Complexity*. Springer, pp. 39–58.

— (2018). "A bad assumption: a simpler model is more general". Review of Artificial Societies and Social Simulation, 28th August 2018. https://roasss.wordpress.com/2018/08/2
2/.

Edmonds, Bruce and Scott Moss (2004). "From KISS to KIDS–an 'anti-simplistic'modelling approach". In: *International workshop on multi-agent systems and agent-based simulation*. Springer, pp. 130–144.

Edmonds, Bruce, Christophe le Page, Volker Grimm, Cristina Montanola, Paul Ormerod, Hilbert Root, and Flaminio Squazzoni1 (2018). "Different Modelling Purposes". In: *Forthcoming*.

Frey, Daniel and Dunja Šešelja (2018a). "Robustness and Idealization in Agent-Based Models of Scientific Interaction". In: *British Journal for the Philosophy of Science* https://doi.org/10.1093/bjps/axy039.

— (2018b). "What is the Epistemic Function of Highly Idealized Agent-Based Models of Scientific Inquiry?" In: *Philosophy of the Social Sciences* https://doi.org/10.1177/0048393118767085.

Gelfert, Axel (2016). *How to do science with models: a philosophical primer*. Springer.

Gräbner, Claudius (2018). "How to Relate Models to Reality? An Epistemological Framework for the Validation and Verification of Computational Models". In: *Journal of Artificial Societies and Social Simulation* 21.3, p. 8. ISSN: 1460-7425. DOI: 10.18564/jasss.3772. URL: http://jasss.soc.surrey.ac.uk/21/3/8.html.

Grim, Patrick (2009). “Threshold Phenomena in Epistemic Networks.” In: *AAAI Fall Symposium: Complex Adaptive Systems and the Threshold Effect*, pp. 53–60.

Grim, Patrick, Daniel J Singer, Steven Fisher, Aaron Bramson, William J Berger, Christopher Reade, Carissa Flocken, and Adam Sales (2013). “Scientific networks on data landscapes: question difficulty, epistemic success, and convergence”. In: *Episteme* 10.04, pp. 441–464.

Holman, Bennett and Justin P Bruner (2015). “The problem of intransigently biased agents”. In: *Philosophy of Science* 82.5, pp. 956–968.

Lehtinen, Aki (2017). “Derivational robustness and indirect confirmation”. In: *Erkenntnis*, pp. 1–38.

Martini, Carlo and Manuela Fernández Pinto (2016). “Modeling the social organization of science”. In: *European Journal for Philosophy of Science*, pp. 1–18.

Nickles, Thomas (2006). “Heuristic Appraisal: Context of Discovery or Justification?” In: *Revisiting Discovery and Justification: Historical and philosophical perspectives on the context distinction*. Ed. by Jutta Schickore and Friedrich Steinle. Netherlands: Springer, pp. 159–182.

O’Connor, Cailin and James Owen Weatherall (2017). “Scientific polarization”. In: *arXiv preprint arXiv:1712.04561*.

Perović, Slobodan, Sandro Radovanović, Vlasta Sikimić, and Andrea Berber (2016). “Optimal research team composition: data envelopment analysis of Fermilab experiments”. In: *Scientometrics*, pp. 1–29.

Pöyhönen, Samuli (2017). “Value of cognitive diversity in science”. In: *Synthese* 194.11, pp. 4519–4540.

Pöyhönen, Samuli and Jaakko Kuorikoski (2016). “Modeling epistemic communities”. In: *The Routledge Handbook of Social Epistemology (forthcoming)*. Ed. by M. Fricker, P. J. Graham, D. Henderson, N. Pedersen, and J. Wyatt. Routledge.

Railsback, Steven F and Volker Grimm (2011). *Agent-based and individual-based modeling: a practical introduction*. Princeton University Press.

Rosenstock, Sarita, Cailin O’Connor, and Justin Bruner (2017). “In Epistemic Networks, is Less Really More?” In: *Philosophy of Science* 84.2, pp. 234–252.

Schelling, Thomas C (1971). “Dynamic models of segregation”. In: *Journal of mathematical sociology* 1.2, pp. 143–186.

Šešelja, Dunja and Christian Straßer (2013). “Abstract argumentation and explanation applied to scientific debates”. In: *Synthese* 190, pp. 2195–2217.

Straßer, Christian, Dunja Šešelja, and Jan Willem Wieland (2015). "Withstanding Tensions: Scientific Disagreement and Epistemic Tolerance". In: *Heuristic Reasoning*. Ed. by Emiliano Ippoliti. Studies in Applied Philosophy, Epistemology and Rational Ethics. Springer, pp. 113–146.

Thicke, Mike (2018). "Evaluating Formal Models of Science". In: *Forthcoming*.

Thiele, Jan C, Winfried Kurth, and Volker Grimm (2014). "Facilitating parameter estimation and sensitivity analysis of agent-based models: A cookbook using NetLogo and R". In: *Journal of Artificial Societies and Social Simulation* 17.3, p. 11.

Verreault-Julien, Philippe (2018). "How could models possibly provide how-possibly explanations?" In: *Studies in History and Philosophy of Science Part A*.

Weatherall, James Owen, Cailin O'Connor, and Justin Bruner (2018). "How to Beat Science and Influence People: Policy Makers and Propaganda in Epistemic Networks". In: *The British Journal for the Philosophy of Science*. https://doi.org/10.1093/bjps/axy062.

Weisberg, Michael and Ryan Muldoon (2009). "Epistemic landscapes and the division of cognitive labor". In: *Philosophy of science* 76.2, pp. 225–252.

Williamson, Timothy (2002). *Knowledge and its Limits*. Oxford University Press on Demand.

Ylikoski, Petri and N Emrah Aydinonat (2014). "Understanding with theoretical models". In: *Journal of Economic Methodology* 21.1, pp. 19–36.

Zamzami, Nuha and Andrea Schiffauerova (2017). "The impact of individual collaborative activities on knowledge creation and transmission". In: *Scientometrics* 111.3, pp. 1385–1413.

Zollman, Kevin J. S. (2007). "The communication structure of epistemic communities". In: *Philosophy of Science* 74.5, pp. 574–587.

— (2010). "The epistemic benefit of transient diversity". In: *Erkenntnis* 72.1, pp. 17–35.