Epistemic Effects of Scientific Interaction: Approaching the Question with an Argumentative Agent-Based Model

Borg, AnneMarie; Frey, Daniel; Šešelja, Dunja; Straßer, Christian

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:
Borg, A., Frey, D., Šešelja, D., & Straßer, C. (2018). Epistemic Effects of Scientific Interaction: Approaching the Question with an Argumentative Agent-Based Model. *Historical Social Research*, *43*(1), 285-307. https://doi.org/10.12759/hsr.43.2018.1.285-307

Historical Social Research
Historische Sozialforschung

Anne Marie Borg, Daniel Frey, Dunja Šešelja & Christian Straßer:

Epistemic Effects of Scientific Interaction: Approaching the Question with an Argumentative Agent-Based Model.

doi: 10.12759/hsr.43.2018.1.285-307

Published in:

Historical Social Research 43 (2018) 1

Cite as:

Borg, Anne Marie, Daniel Frey, Dunja Šešelja, and Christian Straßer. 2018. Epistemic Effects of Scientific Interaction: Approaching the Question with an Argumentative Agent-Based Model. *Historical Social Research* 43 (1): 285-309. doi: 10.12759/hsr.43.2018.1.285-309.

Historical Social Research
Historische Sozialforschung

All articles published in HSR Special Issue 43 (2018) 1: Agent-Based Modeling in Social Science, History, and Philosophy.

Epistemic Effects of Scientific Interaction: Approaching the Question with an Argumentative Agent-Based Model

AnneMarie Borg, Daniel Frey, Dunja Šešelja & Christian Straßer *

Abstract: *»Epistemische Auswirkungen wissenschaftlicher Interaktion: Ein argumentatives agentenbasiertes Modell«.* The question whether increased interaction among scientists is beneficial or harmful for their efficiency in acquiring knowledge has in recent years been tackled by means of agent-based models (ABMs) (e.g. Zollman 2007, 2010; Grim 2009; Grim et al. 2013). Nevertheless, the relevance of some of these results for actual scientific practice has been questioned in view of specific parameter choices used in the simulations (Rosenstock et al. 2016). In this paper we present a novel ABM that aims at tackling the same question, while representing scientific interaction in terms of argumentative exchange. In this way we examine the robustness of previously obtained results under different modeling choices.

Keywords: Agent-based models, abstract argumentation, scientific inquiry, scientific interaction, social networks.

1. Introduction[1]

Recent investigations of social aspects of scientific inquiry have increasingly utilized agent-based models (ABMs) (e.g. Zollman 2007, 2010; Weisberg and Muldoon 2009; Douven 2010; Thoma 2015, etc). Computer models are beneficial for the tackling of several questions in the domain of methodology of science and social epistemology that are difficult to approach with qualitative

* AnneMarie Borg, Institute for Philosophy II, Ruhr-University Bochum, 44801 Bochum, Germany; annemarie.borg@rub.de.
Daniel Frey, Faculty of Economics and Social Sciences, Heidelberg University, Bergheimer Str. 58, 69115 Heidelberg, Germany; freyda87@gmail.com.
Dunja Šešelja, Institute for Philosophy II, Ruhr-University Bochum, 44801 Bochum, Germany; dunja.seselja@rub.de.
Christian Straßer, Institute for Philosophy II, Ruhr-University Bochum, 44801 Bochum, Germany; christian.strasser@rub.de.

1 Acknowledgements: The research of AnneMarie Borg and Christian Straßer was supported by a Sofja Kovalevskaja award of the Alexander von Humboldt-Foundation, funded by the German Ministry for Education and Research.

methods, such as historical case studies. One such question concerns the impact of different degrees and structures of information flow among scientists on their efficiency in acquiring knowledge. Zollman's pioneering work in this domain (Zollman 2007, 2010) suggested that information flow within highly connected social networks may be epistemically harmful, and that there is a trade-off between the success of agents and the amount of time they need to reach a consensus (namely, less connected networks need more time to converge but therefore they converge more often to the best hypothesis [Zollman 2007]). Even though Rosenstock, O'Connor, and Bruner (2016) have shown that the results of Zollman's models do not hold for a large portion of the relevant parameter space, other, structurally different ABMs have come to similar conclusions as Zollman (e.g. Grim 2009; Grim et al. 2013). As a result, whether the results of Zollman's and Grim's models can be considered informative of actual scientific inquiry has remained an open question. On the one hand, these findings could turn out to be a robust property, which can be shown by means of a range of models that employ different idealizations (Weisberg 2006). On the other hand, the findings could also turn out to be a feature resulting primarily from idealizations employed in these models, and unlikely to hold for actual scientific inquiry.

Of special interest are those idealizing assumptions related to the link between the information flow among scientists and their efficiency. One such idealization concerns the fact that interaction among scientists is usually represented as an update that does not take into account important contextual factors, which should further qualify the received information. For example, upon receiving information critical of her theory, an agent's evaluation of the latter may remain unaffected if she already has a counterargument (e.g., concerning the reliability of the given information). This is not only absent from Zollman's and Grim's models, but from other ABMs of scientific inquiry as well. While some authors have tried to relax this idealization by representing agents as "trusting" others only if they have sufficiently similar views (e.g. Hegselmann and Krause 2005), by introducing "noise" in the received information (Douven 2010), by assigning different weights to the opinions of agents (Riegler and Douven 2009), or by assigning different epistemic systems to agents (De Langhe 2013), these adjustments do not capture the argumentative nature of scientific interaction, such as the above mentioned assessment of an argument in view of possible counterarguments.

This points to another possibly problematic idealization frequently present in ABMs, which concerns heuristic behavior of scientists in face of argumentative attacks. While in ABMs such as Zollman's scientists always gather the evidence for the given theory in the same manner, it is important to distinguish between the heuristic behavior that aims at gathering new evidence and the one that aims at finding solutions to the encountered problems. In particular, when scientists face possible anomalies in their research programs, such anomalies trigger a search for counterarguments that could turn apparent refutations into confirmatory instances (Lakatos 1978). This is important in prohibiting a premature abandonment of an actually fruitful inquiry (Nickles 2006; Šešelja et al. 2012).

In this paper we offer a novel approach to the agent-based modeling of scientific inquiry and scientific interaction, which is based on argumentation, and which aims to soften the above mentioned idealizations. In this way we will examine the robustness of the results obtained by previous ABMs of scientific interaction, under different assumptions underlying the simulations. In contrast to most other ABMs of science, our model is based on the idea that an essential component of scientific inquiry is argumentative dynamics between scientists. To this end, we employ abstract argumentation frameworks as one of the design features of our ABM. Abstract argumentation has previously been shown fruitful for the modeling of scientific debates by Šešelja and Straßer (2013) and employed in an ABM of social behavior by Gabbriellini and Torroni (2014). We will show that this feature makes our model a fruitful tool for tackling various questions concerning social aspects of scientific inquiry, including the epistemic benefits of scientific interaction.

Our paper is structured as follows. In Section 2 we introduce the idea underlying our ABM and its main features. In Section 3 we present the central findings of the presented version of our model. In Section 4 we compare our findings with those obtained by other ABMs of science and discuss idealizations present in our current model, which should be taken into consideration when assessing the relevance of our results for the actual scientific practice. In Section 5 we present ideas for further enhancements of this ABM.

2. The Model

The aim of our ABM is to represent scientists engaged in a scientific inquiry with the goal of finding the best of the given rivaling theories, where they occasionally exchange arguments with other scientists, *pro* or *con* the given pursued theories. The model is used to tackle the question which structure of the information flow leads scientists to most efficiently discover the best theory, where efficiency is measured in terms of their success and the time they need to complete their exploration. In this section we will explain the central elements of our model, namely, the landscape which agents explore, the behavior of agents, and the notion of social networks employed in our simulations.[2]

[2] Our ABM is created in NetLogo (Wilensky 1999). The source code is available at <https://github.com/g4v4g4i/ArgABM/tree/HSR-Special-Issue-Agent-Based-Modelling> (Accessed January 31, 2018).

2.1 The Landscape

Agents, representing scientists, move along an *argumentative landscape*. The argumentative landscape, which represents rivaling theories in a given scientific domain, is based on a dynamic abstract argumentation framework. Let us explain what 'abstract' and 'dynamic' mean here.

An abstract argumentation framework, introduced by Dung (1995), is a pair consisting of a set of abstract entities A representing *arguments* and an *attack relation* $\rightsquigarrow$ over A. Similarly, the framework underlying our model consists of a set of arguments and an attack relation over this set. In addition to attacking each other, arguments may also be connected by a *discovery relation* $\hookrightarrow$. The latter represents the path which scientists have to take in order to discover different parts of the given theory, i.e., some argument b in the given theory can only be discovered after an argument a has been discovered for which $a \hookrightarrow b$.[3]

A scientific theory is represented as a conflict-free set of arguments (i.e. no argument in the theory attacks an argument in the same theory), connected by discovery relations, resulting in a tree-like graph. Formally, an argumentative landscape is given by a triple $\langle A, \rightsquigarrow, \hookrightarrow \rangle$where $A = \langle A_1, \ldots, A_m \rangle$ is partitioned in m many theories $T_i = \langle A_i, a_i, \hookrightarrow \rangle$ which are trees with $a_i \in A_i$ as a root and

$$\rightsquigarrow \subseteq \bigcup_{\substack{1 \leq i,j \leq m \\ i \neq j}} (\mathcal{A}_i \times \mathcal{A}_j) \quad \text{and} \quad \hookrightarrow \subseteq \bigcup_{1 \leq i \leq m} (\mathcal{A}_i \times \mathcal{A}_i).$$

Specifying $\rightsquigarrow$ like this ensures that the theories are conflict-free.

The abstractness of the framework concerns all of its elements. Instead of providing the concrete content and structure of the given arguments, we represent them as abstract entities. Similarly, we do not reveal the concrete nature of the attack or the discovery relation.

The framework is dynamic in the sense that agents gradually discover arguments, as well as attack and discovery relations between them. Given the abstract nature of arguments, we interpret them as hypotheses which scientists investigate. Occasionally scientists encounter defeating evidence, represented by attacks from other arguments, and then attempt to find defending arguments for the attacked hypothesis a (i.e., to find arguments in the same theory attacking arguments from other theories that attack a). This dynamic aspect is implemented by associating arguments with their *degree of exploration* for an agent at a given time point of a run of the simulation: for each agent ag and each argument $a \in A$, $expl(a, ag) \in \{0, \ldots, 6\}$ where 0 indicates that the argument is unknown to ag and 6 indicates that the argument is fully explored and

3 While the process of discovery is generally governed by this rule, there are some exceptions to it, which occur as a result of the communication between agents (see below Section 2.3).

cannot be further explored.[4] In view of this agents have subjective and limited insights into the structure of the landscape. Whether an attack or discovery relation between two arguments *a* and *a'* is visible to an agent *ag* depends on the degree of exploration *expl(a,ag):* the higher *expl(a,ag)* is, the more relation[s] between *a* and other arguments will be visible (additionally agents may learn about the landscape by communicating with other agents, see Section 2.3).

2.2 Basic Behavior of Agents

The model is round-based and each round agents perform actions that are among the following:

(A1) exploring a single argument, thereby gradually discovering possible attacks (on it, and from it to arguments that belong to other theories) as well as discovery relations to neighboring arguments;
(A2) moving to a neighboring argument along the discovery relation within the same theory;
(A3) moving to an argument of a rivaling theory.

While agents start the run of the simulation at the root of a given theory, they will gradually discover more and more of the argumentative landscape. This way each turn an agent operates on her own (subjective) fragment of the landscape, which consists of her discovered arguments which are explored by her to a specific degree, and her discovered (attack and discovery) relations between the arguments.

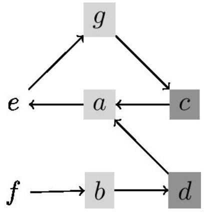
In order to decide whether to work on the current theory (A1 and A2 above), or whether to better start working on an alternative theory (A3) agents are equipped with the ability to evaluate theories. Every few rounds agents apply an evaluative procedure, with respect to the set of arguments and attacks they currently know (i.e. their subjective memory), in order to determine the current subjective *degree of defensibility* of each of the theories. The degree of defensibility of a theory is the number of defended arguments in this theory, where – informally speaking – an argument *a* is defended in the theory if it is not attacked or if each attacker *b* from another theory is itself attacked by some defended argument *c* in the current theory.

We give a more precise formal definition. We say that a subset of arguments *A* of a given theory *T* is *admissible* if for each attacker *b* of some *a* in *A* there is an *a'* in *A* that attacks *b* (*a'* is said to defend *a* from the attack by *b*). Since every theory is conflict-free (there are no *a* and *b* in *T* such that *a* attacks *b*), it can easily be seen that for each theory *T* there is a unique maximally admissible

[4] Our model is round-based (more on that in Section 2.2). Each round may be interpreted as one research day. Each of the 6 levels of an argument takes a researcher 5 rounds/days of exploration. Thus, each argument represents a hypothesis that needs altogether 30 research days to be fully investigated.

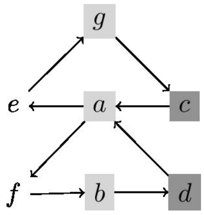
subset of T (with respect to set inclusion). An argument a in T is said to be *defended in T* if it is a member of this maximally admissible subset of T.[5] The *degree of defensibility* of T is equal to the number of defended arguments in T.

Figure 1: Argumentation Framework 1



theory	defended	degree of def.
$T_1 = \{e, f\}$ (white)	$\{f\}$	1
$T_2 = \{a, b, g\}$ (gray)	$\{\}$	0
$T_3 = \{c, d\}$ (dark gray)	$\{\}$	0

Figure 2: Argumentation Framework 2



theory	defended	degree of def.
$T_1 = \{e, f\}$ (white)	$\{\}$	0
$T_2 = \{a, b, g\}$ (gray)	$\{a, b, g\}$	3
$T_3 = \{c, d\}$ (dark gray)	$\{\}$	0

Figure 1 depicts a situation with three theories as it might occur from the perspective of a given agent: T_1 consisting of arguments e and f (white nodes), T_2 consisting of arguments a, b, and g (gray nodes), and T_3 consisting of arguments c and d (dark gray nodes). The arrows represent attacks, we omit discovery relations. We are now interested in the degrees of defensibility our agent would ascribe to the given theories. The table shows which arguments are defended in each theory and their corresponding degree of defensibility. The only defended argument in this situation is f in T_1. Note for instance that in T_3 the argument d is not defended since no argument in T_3 is able to defend it from the attack by b. Although the argument f in T_1 attacks b, it doesn't count as a

[5] Since theories are conflict-free, our notion of admissibility comes down to the same as the one introduced in Dung (1995). In the terminology of Dung (1995) our sets of defended arguments correspond to Dung's preferred extension (which are exactly the maximally admissible sets) with the difference that we determine these sets relative to given theories.

defender of *d* for theory T_3 when determining the defended arguments in T_3 since in our account a theory is supposed to defend itself.

Figure 2 depicts the situation after an attack from *a* to *f* has been discovered. Consider theory T_2. In this situation *a* defends *b* from the attack by *f, b* defends *a* from the attack by *d, a* defends *g* from the attack by *e* and *g* defends *a* from the attack by *c*. Hence, all arguments are defended resulting in a degree of defensibility of 3.

Agents decide to move to a rivaling theory if the degree of defensibility of their current theory is below a relative threshold compared to the theory with the highest degree of defensibility, i.e., the theory with the most defended arguments.[6]

To make a decision between options A1 and A2, each agent employs the following heuristic: every time step she considers all arguments in her direct neighborhood (relative to the discovery relation) that could possibly be the next ones to work on. With a certain probability (determined in the interface of the model) she will then move to one of these arguments, or alternatively keep on exploring her current argument. In case the argument she is currently at is fully explored, she will try to move on to a neighboring argument, and if such an argument isn't visible (e.g., if she has reached the end of a branch of her theory) she will try to move to the parent argument, or in case it is fully explored, she will move to another not fully explored argument in the same theory.

The decision making of agents also includes some prospective considerations. If during her exploration an agent discovers an attack on the argument *a* she is currently investigating, she will attempt to discover a defense for it. A defending argument may be found among the visible neighboring arguments of *a* (relative to the discovery relation). In case *a* is attacked by an argument *b*, an agent *ag* working on *a* will 'see' outgoing attack arrows $a' \rightsquigarrow b$ from an already discovered child argument *a'* of *a*, even in cases where $a' \rightsquigarrow b$ is not yet discovered by *ag* since *a'* is insufficiently explored by *ag*.[7] This way our agent knows that exploring *a'* may help in defending *a*. If no potential defender of *a* is visible she keeps on exploring *a* in the hope of discovering new neighbors and thus new potential defenders.

This idea corresponds to the situation in which a scientist discovers a problem in her hypothesis and attempts to find a way to resolve it. While she may not have a solution ready at hand, she may have a method for finding such a

---

[6] The relative threshold can be selected in the interface of the model under *strategy threshold* and it allows for theories with similar degrees of defensibility to be considered as equally good. In addition, agents have a degree of 'inertia' which can be selected in the interface under *jump threshold*. Where *n* is the jump threshold, an agent can remain for the duration of *n* evaluations on a theory which is not one of the subjectively best ones.

[7] In such cases, where an attack relation is merely 'seen' but not yet discovered by an agent, it is not yet considered as a defense when the respective theory is evaluated by the agent.

solution (for example, going back to the laboratory and conducting some new experiments).[8]

2.3 Social Networks

An agent discovers the argumentative landscape by investigating arguments (as described in Section 2.2) or by means of exchanging information about the landscape with other agents. We will now describe the latter aspect. As mentioned above, the presented version of the model is designed to investigate how different social networks impact the efficiency of scientists in discovering the best theory. In contrast to other ABMs employing the idea of social networks (e.g. Zollman 2007, 2010; Grim 2009), which represent connectivity only in view of different types of graphs that connect agents, we distinguish between two types of social networks. First, our agents are divided into *collaborative networks* which consist of up to five individuals who start from the same theory.[9] While each agent gathers information (i.e. the attack and discovery relations between arguments) on her own, every five steps this information is shared with all other agents forming the same collaborative network.

Second, besides sharing information with agents from the same network, every five steps each agent shares information with agents from other collaborative networks with a given *probability of information sharing* that is determined before the run of the simulation.[10] This way the agents form ad-hoc and random *communal networks* with agents from other research collaborations. A higher probability of information sharing leads to a higher degree of interaction among agents. Finally, we represent *reliable* and *deceptive* scientists by allowing for different approaches to the sharing of information between networks. A reliable agent shares all the information she has gathered during her exploration of the current theory, while a deceptive agent does not share information regarding discovered attacks on her current theory. This corresponds to the idea that deception consists in providing some while withholding other information, thus leading the receiver to a wrong inference (Caminada 2009).

Agents share information either in a unidirectional (an agent sends information to another agent) or a bidirectional way (two agents exchange infor-

---

8 Such a heuristic response to apparent refutations belongs to what Lakatos has dubbed the *negative heuristics* of a research program (Lakatos 1978).

9 Our model allows for collaborative groups also to be formed of agents that are positioned on different theories. While in this paper we have focused on the former type, examining the behavior of the latter remains for future research.

10 While agents share their full subjective knowledge within their respective collaborative networks, the information which they share with agents from other networks concerns recently obtained knowledge on the theory (arguments and attacks in an agent's current neighborhood) which they are currently exploring. This corresponds to a situation in which a scientist writes a paper or gives a talk that presents arguments for and/or against hypotheses regarding the theory she is currently pursuing.

mation both ways). Moreover, our model takes into account the fact that receiving information is time costly: when an agent receives information, she will not explore the argument on which she is standing nor move. This corresponds to the idea that scientists need to invest time when reading papers by other scientists, which they would otherwise devote to their own research.

In order to get a better picture of the information flow represented in our model and to better understand how costly sharing information is, let's recall the interpretation (see Footnote 4) according to which each round stands for one research day, so that five rounds represent one research week. Our agents then represent scientists who each Friday first evaluate all the available theories in order to decide which one to pursue. Then from next Monday to Thursday they engage in the pursuit of their chosen theory. On Fridays, however, before performing the assessment of theories, there is a chance that they share information with scientists from other collaborative groups. In case a scientist indeed receives information from another scientist, she skips pursuing her current theory on this Friday (suppose she is at a workshop or she has to review a paper). For example, if the probability of information sharing is set to 0.3, then each Friday there is a 30% chance for each of our scientists to interact with a scientist from another group. If the probability of information sharing is set to the maximum (i.e. to 1), then every Friday each of our scientists will necessarily share information with a scientist from another group, that is, Friday will be a "communication day" for everyone.

3. The Main Findings

In this section we will first specify the parameters used in the simulations and then present our main results.

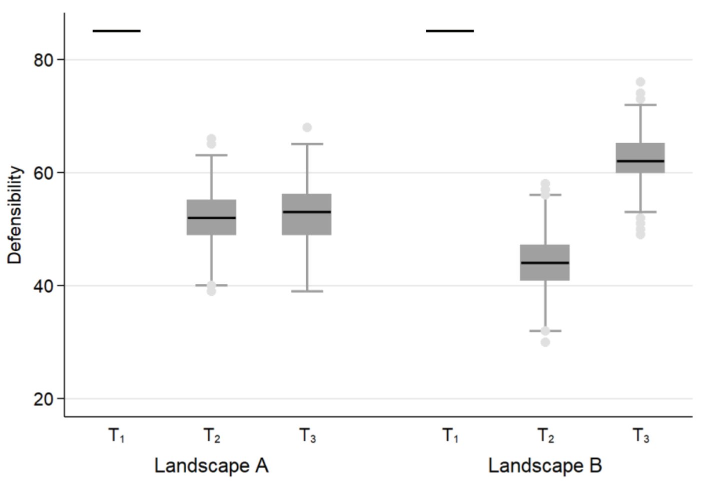
3.1 Parameters Used in Simulations

We have run the simulation 10,000 times with 10, 20, 30, 40, 70, and 100 agents by varying the following settings:

1) different probabilities of an agent communicating with agents from other collaborative networks, namely: 0.3, 0.5, and 1;
2) different approaches to communicating, namely: reliable and deceptive agents;
3) two different landscapes (A and B), both representing three theories (T_1 – T_3), where T_1 in both cases has the maximum degree of defensibility (i.e. all

of its 85 arguments are fully defensible), while T_2 and T_3 have different degrees of defensibility (Figure 3).[11]

Figure 3: Defensibility on Different Landscapes



Tukey boxplot for 1000 data points each: the bottom and top of the box are the first and third quartiles, and the band inside the box is the median. Whiskers extend to the most distant value within 1.5 IQR from each end of the box and outliers are shown as light-gray dots.

On the one hand, landscape A represents a situation in which T_2 and T_3 are clearly worse than T_1 while not being completely problematic (as it would, for instance, be the case with pseudo-scientific theories). Hence, in both theories approximately half of their arguments are defensible in the objective landscape in the majority of setups (Figure 3). On the other hand, landscape B represents a situation in which T_2 is clearly worse than the other two theories, while T_3 has a degree of defensibility closer to the fully defensible T_1. For example, T_2 cor-

[11] Theories are modelled as trees of depth 3, where each argument (except for the leaves) has 4 child-arguments, amounting to altogether 85 arguments. Here is a list of other parameters and a short explanation. The *move probability* is set to 0.5. Together with the degree of exploration of the argument an agent is situated at, it determines for each agent the chance of moving to another argument. The *visibility probability* concerns the probability with which a new attack relation is discovered when an agent further explores her current argument. It is set to 0.5. The *research speed* determines the number of time steps an agent has to work on an argument before that argument reaches its next level of exploration. It is set to 5. The *strategy threshold* is set to 0.9. It expresses that each theory with a degree of defensibility that is at least 90% of the degree of defensibility of the best theory is considered good enough to work on by our agents. The *jump threshold* has been explained in footnote 6 and it is set to 10.

responds to a theory that doesn't have many successful hypotheses (e.g. they fail at offering successful explanations of the relevant phenomena), and whose arguments are thus very much attacked.[12] In contrast, T_3 stands for a theory that can explain many phenomena, but still less than T_1.

The program runs until one of the theories is completely explored. At that point all the agents have one more chance to make their final evaluation and choose their preferred theory. As mentioned above, our main research question is how efficient agents are in each of the above listed scenarios, where efficiency is assessed in terms of the success of agents in acquiring knowledge, as well as in terms of the time steps needed for the run to be completed.

We have examined this question with respect to two criteria of success:

- *monist criterion*, according to which a run is considered successful if, at the end of the run, all agents have converged onto the objectively best theory;
- *pluralist criterion*, according to which a run is considered successful if, at the end of the run, there is no theory for which the number of agents working on it is greater than the number of agents working on the objectively best theory.

The monist criterion corresponds to the idea that the scientific community is successful in case it reaches a consensus on the objectively best theory. This approach has commonly been employed in ABMs of science, such as Zollman's and Grim's ones mentioned above. Nevertheless, this criterion does not reflect the view of philosophers of science who endorse a pluralist view on scientific inquiry. According to scientific pluralism, a parallel existence of rivaling scientific theories is considered epistemically and heuristically beneficial for the goals of the scientific community (e.g. Longino 2002; Kitcher 2011; Chang 2012). Hence, from a pluralist perspective, a primary epistemic goal is not the convergence of all scientists onto the same theory (in fact, some even consider it undesirable), but rather to assure that no fruitful theory has been prematurely abandoned. In our model this is represented as a test for whether the best theory is among the most actively investigated ones.

3.2 Results

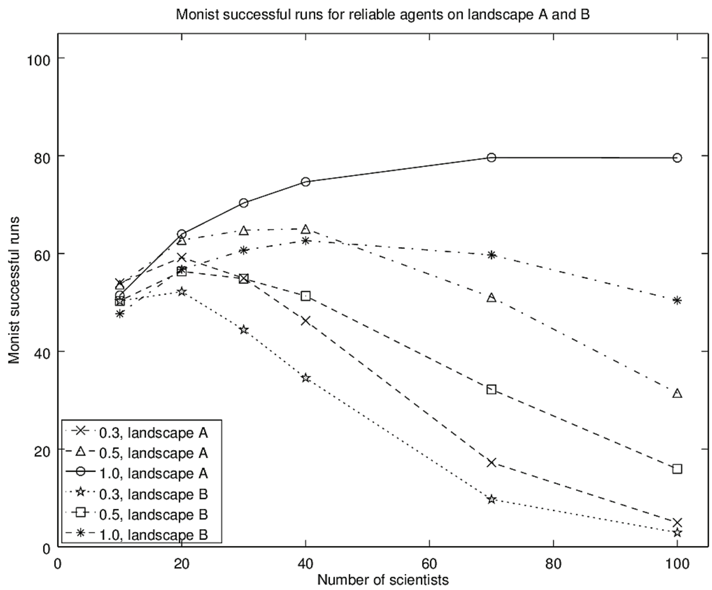
In what follows we present the most significant results of our simulations.

[12] For example, an undefended attack from theory T_i to theory T_k can be understood as a situation in which T_i has a hypothesis that explains a phenomenon more successfully than a corresponding hypothesis in T_k.

Information Sharing among Reliable Agents

For smaller groups of reliable agents (up to 20) the impact of information sharing is rather small (Figures 4 and 5). From 30 agents on, we observe a positive impact of an increase in information sharing on the successful convergence as well as on the pluralist kind of success, in case of both landscapes. In contrast, the negative effect on time steps needed is minimal (Figure 6).[13] If we impose a high communication cost of 3 rounds (instead of the standard cost of 1 round), then if agents communicate less (with probability 0.3), they reach the end of the run quicker than if they communicate more (with probability 1). As expected, the effect is reversed in case communication has no cost (see Figure 7).

Figure 4: Monist Success of Reliable Agents



Monist success of reliable agents, with different probabilities of information sharing. Each data point is generated on the basis of 10.000 runs. The same holds for the other plots below.

[13] While Figures 6, 7, and 10 concern the monist criterion, the results were in all cases similar for the pluralist criterion.

Figure 5: Pluralist Success of Reliable Agents

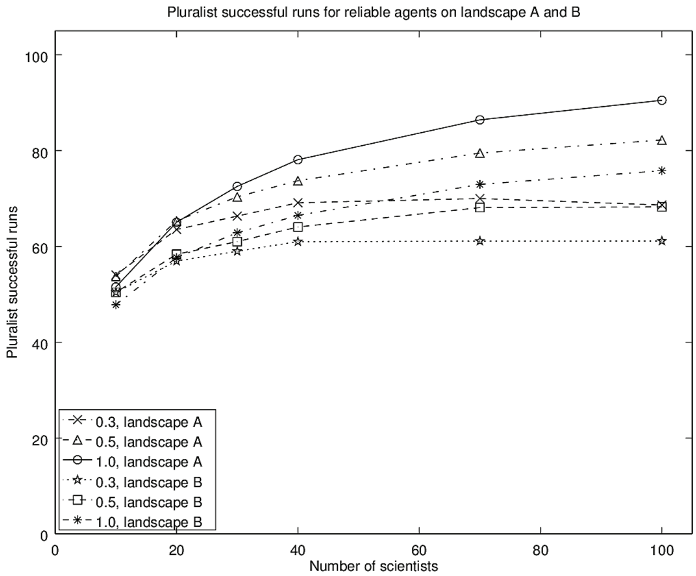


Figure 6: Time Steps Needed for Successful Runs (Monist Criterion) with Reliable Agents

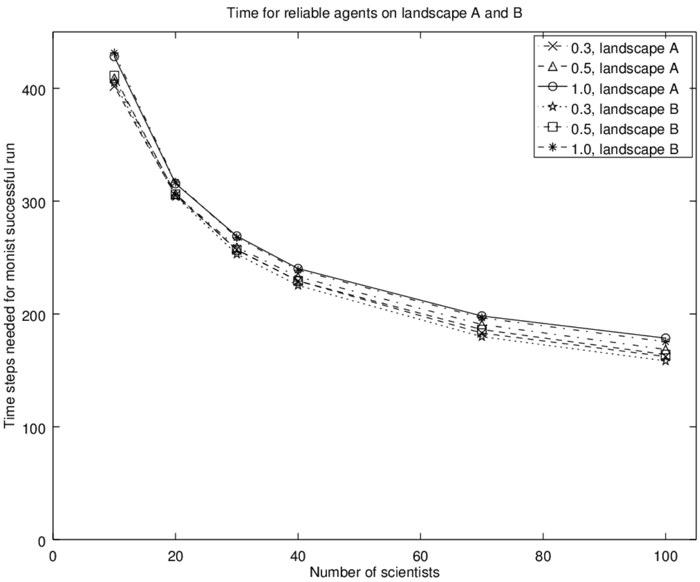
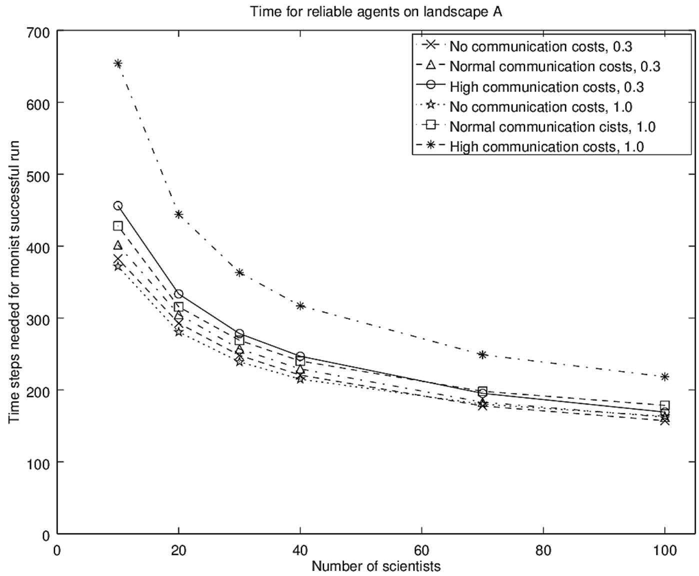


Figure 7: Time Steps Needed for Successful Runs (Monist Criterion) with Reliable Agents and Different Communication Costs



Information Sharing Among Deceptive Agents

While for smaller groups of deceptive agents a higher degree of information sharing has a relatively small impact, we notice positive effects in cases of larger communities (Figure 8), without slowdowns (Figure 10). In case of the pluralist criterion of success, different degrees of information sharing lead to similar degrees of success (Figure 9).

Reliable vs. Deceptive Agents

If we compare the groups with same degrees of information sharing, reliable agents tend to be clearly more successful than the deceptive ones, while being only slightly slower.

Figure 8: Monist Success of Deceptive Agents

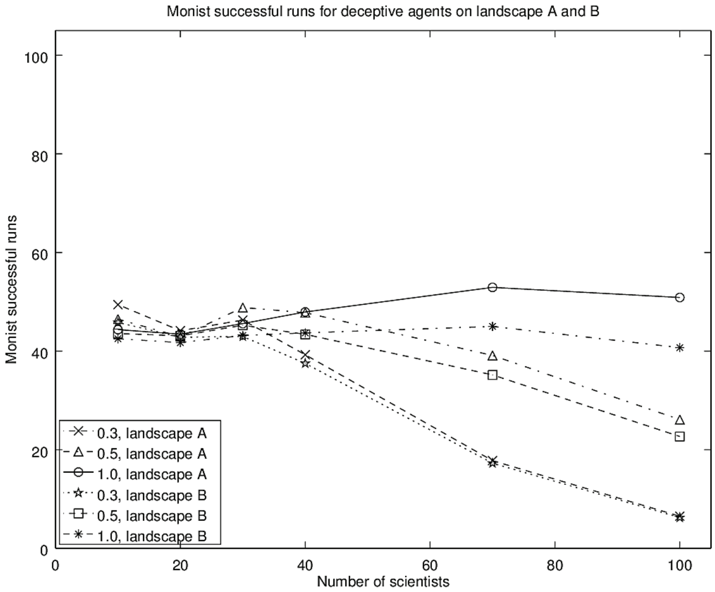


Figure 9: Pluralist Success of Deceptive Agents

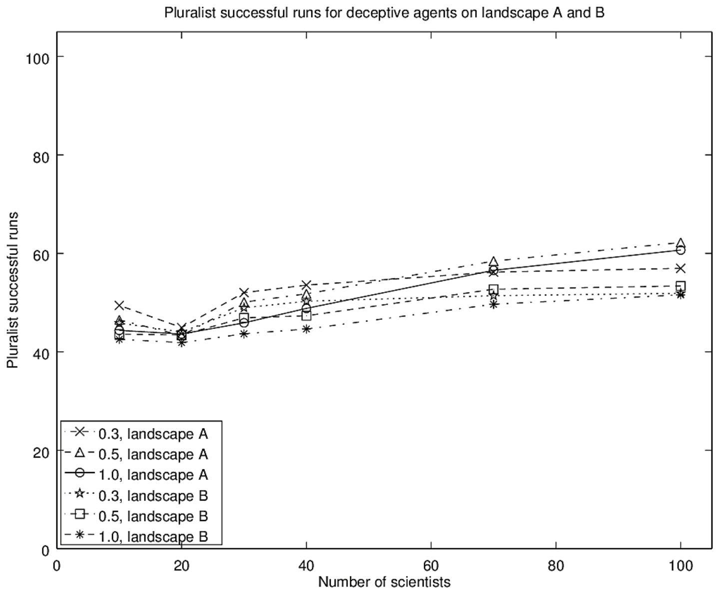
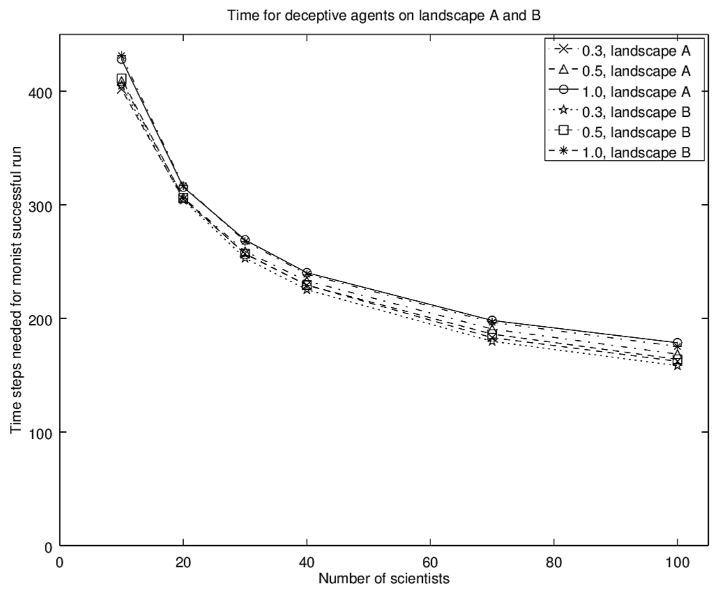


Figure 10: Time Steps Needed for Successful Runs (Monist Criterion) with Deceptive Agents



Efficiency of Agents on Two Landscapes

Landscape A tends to be explored more successfully than landscape B, while the time steps needed remain similar. This holds for both reliable and deceptive agents.

Size of the Scientific Community

Which size of the scientific community is optimal depends on the notion of success we choose (i.e. the monist or the pluralist one). With the monist criterion larger populations of 70 and 100 agents are outperformed by smaller populations (with an optimum around 20 and 30). A possible explanation is that with larger sized populations information circulates less among research groups,[14] which may prevent them from reaching consensus, i.e. from converg-

[14] Since the size of the research groups is not proportional to the population size, the communication network on "Fridays" will with larger populations be less dense than with smaller populations. The above explanation is confirmed by the results in Borg et al. (2017), where we present a variant of the model in which the communication across the population is proportional to its size: larger populations in this variant of the model are more successful than smaller populations, in case they are highly connected.

ing onto one theory. In contrast, the increase in the population size tends to be beneficial with respect to the pluralist criterion of success.

4. Discussion

In this section we will first analyze the results presented in the previous section and compare them with findings obtained by other ABMs of science. We will then turn to a critical analysis of some idealizations present in our model.

Comparison of Our Results with Other ABMs of Science

Let us start from our two landscapes and the implications of their differences on the behavior of agents. The fact that agents are more successful on landscape A than on landscape B indicates that landscape B represents an inquiry that is more difficult than the one represented by landscape A. This is not surprising: while A corresponds to a scenario in which the best theory is objectively much better than either of its two rivals, both of which are similarly weak, B corresponds to a scenario in which one of the rivals of the best theory is very weak, while another is comparatively stronger (see Figure 3). Hence, in the latter case, it is more difficult to determine which theory is objectively the best one.[15]

In case of both landscapes our highest degree of information sharing – which corresponds to the situation in which each scientist shares information with a scientist from another collaborative group once a week – is epistemically beneficial. Note that this especially holds for landscape B, that is, in the context of a difficult inquiry. This result contrasts with conclusions drawn from other ABMs of scientific interaction (Zollman 2007, 2010; Grim, 2009; Grim et al. 2013). In particular, we have observed no "Zollman effect" – namely, the positive difference between successful convergence in a network with a low degree of information sharing versus a network with a high degree of information sharing (see Rosenstock et al. 2016) (see Figure 4). Does this also mean that

[15] In terms of concrete historical examples, the latter case could be compared with the research on peptic ulcer disease (PUD) in the second half of the 19th and the first half of the 20th century. The objectively weakest theory corresponds to the 19th century view that factors such as "Old age, privation, fatigue, mental anxiety, and intemperance" are the main causes of PUD (Grob 2003). Two much stronger theories, which were the 20th century dominant rivals were based, respectively, on the 'bacterial hypothesis,' according to which bacteria are the main etiological agent, and the 'acidity hypothesis,' according to which excessive secretion of stomach acid causes the ulcer. While the weakest theory was rejected as insufficiently substantiated early on, the bacterial hypothesis was prematurely abandoned in the mid-20th century, only to be revived in the 1980s with Warren and Marshall's discovery of *Helicobacter pylori*, the major cause of PUD (Šešelja and Straßer 2014).

our results challenge the latter? In order to answer this question, we should first clarify the differences between our respective models.

First of all, we have to be careful when comparing the structure of social networks between our models. On the one hand, in Zollman's and Grim's models agents are connected in one network, throughout which information (gradually) spreads, from one agent to another. In contrast, our model employs two types of networks: small collaborative ones, which are complete in the sense that each agent shares information with every other agent, and communal ones, which are every time generated in a random ad-hoc way, in the sense that whether an agent shares information is determined probabilistically and with whom she shares it is random. In addition, within communal networks agents share only recently acquired information concerning their current theory. As a consequence, in our communal networks not all information necessarily spreads throughout the entire community. Moreover, the highest degree of information sharing in our model is lower than Zollman's and Grim's complete networks (where each agent shares information with every other agent).

Nevertheless, under the assumption that the intended target phenomenon in our approach as well as in these ABMs is actual scientific interaction, our findings indeed pose a challenge to the contrary results by the latter. Since, as we argue below, there is no reason to assume that any of the above mentioned ABMs represent scientific interaction more adequately than our ABM does, our results show that the findings of the latter are not robust under different assumptions and idealizations underlying the models. Moreover, this leaves the question – whether each of these ABMs represents a specific type of actual scientific inquiry, or just a logical possibility not representative of any real world phenomena – open for future research. In fact, one may expect that scientific inquiry is a diversified target phenomenon and that ultimately, the assumptions underlying some models fit some types of scientific inquiry better than others, and that one model may not fit them all. In this case, and once the intended target domains of the respective models have been rendered more precisely, we may be able to avoid the interpretation according to which the diverging results obtained by them are challenging each other or undermining each other's robustness.

Let us now highlight the features of our model that allow for a more adequate representation of scientific inquiry than this has so far been done with ABMs of science.[16] First of all, our model addresses idealizing assumptions identified in other ABMs (see Section 1), namely (i) the inadequate representation of heuristic behavior, as well as (ii) the inadequate representation of in-

[16] Note that these features do not imply that the results of more idealized models are necessarily less informative, less explanatory, or less reliable. We only wish to show that our model can be used to examine whether idealizations underlying other models lead to results that are not robust under less idealized circumstances.

formation flow among scientists. Regarding (i), scientists in our model are equipped with certain prospective considerations, explicated in Section 2.2. In the current version of the model, such prospective considerations play a role in methods that guide agents in their inquiry. In particular, if an agent discovers an attack on her current argument, she tries to find a defense for it. Regarding (ii), the notion of information flow is in contrast to many other ABMs of science not just a simple update of information. Instead, exchange of information is represented as argumentative, which means that received information is critically assessed. Hence, when an agent receives new information regarding a discovered argument, she will assess it as acceptable (in case it can be defended in view of her knowledge base) or as unacceptable (in case it cannot be defended in view of her knowledge base).

Finally, our model incorporates the fact that receiving information (such as reading articles by other scientists) costs time.

Idealizations Present in Our ABM

In spite of all the above mentioned features, our ABM is still based on a number of idealizations, the impact of which should be examined in future research.

First, heuristics of agents are highly simplified, including their search for defense in view of discovered attacks. Even though agents may recognize a defense in case it is located in their surroundings (i.e. in one of the child-arguments of the given attacked argument, which an agent currently explores), a defense may be present at an entirely different branch of the tree. It is not surprising then that we noticed hardly any effect of the heuristic behavior in our runs when comparing them to runs without heuristic behavior. Equipping agents with more adequate tools and thus representing a scientist as having methods for finding solutions for the current anomalies of the theory, is another task for future research.[17] It also remains a task for a future version of the model to incorporate such features into the evaluations performed by agents in view of which they can judge how promising their theory is. Such an assessment would more aptly capture heuristic appraisal, which informs scientists how worthy of pursuit different theories are, rather than how confirmed (or defensible) they are in view of the available evidence.

Second, anomalies of a given theory are currently represented solely as attacks from one of the rivaling theories. An improved version of the model should allow for anomalies to be represented also as counter-evidence that is independent from those other theories. To this end, we will include a more direct representation of evidence and counter-evidence in a future version of the model.

[17] For a step further in this direction see Borg et al. (2017).

Third, which parameter settings for building landscapes in our model are representative of specific types of scientific controversies is an open question. While we have tried to represent different situations of a scientific inquiry with our two landscapes, more work needs to be done to allow for an empirical calibration of our model.[18]

Fourth, as mentioned in Section 2.2, the calculation of the degree of defensibility of a theory is based on the number of defensible arguments of that theory: if it turns out that another theory has more defensible arguments (above a relative threshold), after some deliberation the agent jumps to that theory. This means that our agents represent scientists who are easily impressed by new discoveries, that is, by hypotheses that in view of new results seem to be corroborated. Hence, if a rivaling theory appears to have more such hypotheses, our scientists will switch to pursuing it, although later studies may reveal that some of those hypotheses have turned out to be wrong. Note that changing the way agents evaluate theories may have a strong impact on our current results. For instance, our agents could be modeled as scientists who never abandon their current theory unless a number of its undefended arguments (i.e. hypotheses) passes a certain relative threshold, in comparison to the rivaling theories. Such agents wouldn't easily get misled by early apparent successes of a rivaling theory, but they could be misled by early apparent problems in their own or in one of the rivaling theories. Investigating the impact of different types of evaluations underlying decisions of scientists for which theory they are to pursue remains a task for future research.

In view of these remarks, it is important to interpret the results of our runs cautiously when it comes to their relevance for actual scientific practice. The primary function of our model in its current state is, on the one hand, to provide the basis for testing the robustness of the results obtained by other ABMs tackling similar questions, in order to determine whether they provide a *robust property*. A robust property is, according to Weisberg "a dynamic or static property common to many models making different idealizing assumptions" (2006, 736), and a necessary step in the robustness analysis of a set of models representing a given target phenomenon.[19] On the other hand, the aim of our ABM is to serve a heuristic purpose of constructing more informative models of scientific interaction (Reutlinger et al. 2016). The modular nature of our

[18] For the importance of empirically calibrating simulations of this kind see Martini and Pinto (2016).

[19] As Weisberg writes: "During this stage it is important to collect a sufficiently diverse set of models so that the discovery of a robust property does not depend in an arbitrary way on the set of models analyzed" (2006, 737). Given that the assumptions underlying our model are quite different from those underlying Zollman's or Grim's models, we find it suitable for this purpose.

model together with its specific design features makes it a fruitful basis for further improvements, which can provide insights into real world phenomena.[20]

5. Conclusion and Outlook

In this paper we have presented an ABM of scientific inquiry which makes use of abstract argumentation, aiming to model the argumentative nature of scientific inquiry. The presented version of our model is designed to tackle the question how different degrees of information flow among scientists affects the efficiency of their knowledge acquisition. Our results suggest that an increased information sharing tends to be epistemically beneficial, and that this holds for both, the conditions of easier inquiry and the conditions of a more difficult one, where determining which of the rivaling theories is the best one is harder. While we have argued that our model allows for a de-idealization of some important assumptions underlying scientific inquiry in comparison to the previous ABMs of science, we have also emphasized a number of currently employed idealizations, the impact of which remains to be investigated in future research. We will conclude the paper by showing the fruitfulness of our ABM for the investigation of related questions concerning social aspects of scientific inquiry.

First, our model can be enhanced with different types of research behavior, such as “mavericks” and “followers,” introduced in Weisberg and Muldoon (2009) (see also Fernández Pinto and Fernández Pinto 2018 in this issue). Next, the model can be enhanced with other aspects of scientific inquiry, such as an explanatory relation and a set of explananda (Šešelja and Straßer 2013). This would allow for an investigation of different evaluative procedures which agents perform when selecting their preferred theory (e.g. in addition to the degree of defensibility, agents can take into account how much their current theory explains, or how well it is supported by evidence). Furthermore, a number of enhancements available from the literature on abstract argumentation, such as probabilistic semantics (Thimm 2012), values (Bench-Capon 2002), etc. can be introduced in future versions of our ABM.

References

Trevor J. M. Bench-Capon. Value based argumentation frameworks. *arXiv*, cs.AI/0207059, 2002.

[20] To this end, analyzing the sensitivity of the output of our model to changes in parameters (e.g. via methods proposed in Thiele et al. [2014]) remains a task for future research.

Borg, AnneMarie, Daniel Frey, Dunja Šešelja, and Christian Straßer. 2017. Examining network effects in an argumentative agent-based model of scientific inquiry. In *Proceedings of the 6th International Conference on Logic, Rationality and Interaction (LORI)*, vol. forthcoming of *FoLLI Series on Logic, Language and Information*. Berlin, Heidelberg: Springer.
Caminada, Martin. 2009. Truth, lies and bullshit: distinguishing classes of dishonesty. In *Social Simulation Workshop at the International Joint Conference on Artificial Intelligence (SS@IJCAI)*, Pasadena.
Chang, Hasok. 2012. *Is Water H2O? Evidence, Pluralism and Realism*. Dordrecht: Springer.
De Langhe, Rogier. 2013. Peer disagreement under multiple epistemic systems. *Synthese* 190: 2547-56.
Douven, Igor. 2010. Simulating peer disagreements. *Studies in History and Philosophy of Science Part A* 41 (2): 148 -57.
Dung, Phan Minh. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77: 321-58.
Fernández Pinto, Manuela, and Daniel Fernández Pinto. 2018. Epistemic Landscapes Reloaded: An Examination of Agent-Based Models in Social Epistemology. *Historical Social Research* 43 (1): 48-71. doi: 10.12759/hsr.43.2018.1.48-71.
Gabbriellini, Simone, and Paolo Torroni. 2014. A new framework for ABMs based on argumentative reasoning. In *Advances in Social Simulation, Proceedings of the 9th Conference of the European Social Simulation Association*, ed. Bogumił Kamiński and Grzegorz Koloch, 25-36. Berlin, Heidelberg: Springer.
Grim, Patrick. 2009. Threshold phenomena in epistemic networks. In *AAAI Fall Symposium: Complex Adaptive Systems and the Threshold Effect*, 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. *Episteme* 10 (4): 441-64.
Grob, Gerald N. 2003. The rise of peptic ulcer, 1900-1950. *Perspectives in biology and medicine* 46 (4): 550-66.
Hegselmann, Rainer, and Ulrich Krause. 2005. Opinion dynamics driven by various ways of averaging. *Computational Economics* 25 (4): 381-405.
Kitcher, Philip. 2011. *Science in a democratic society*. Amherst, NY: Prometheus Books.
Lakatos, Imre. 1978. *The methodology of scientific research programmes*. Cambridge: Cambridge University Press.
Longino, Helen. 2002. *The Fate of Knowledge*. Princeton: Princeton University Press.
Martini, Carlo, and Manuela Fernández Pinto. 2016. Modeling the social organization of science. *European Journal for Philosophy of Science*: 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. Jutta Schickore and Friedrich Steinle, 159-82. Dordrecht: Springer.
Reutlinger, Alexander, Dominik Hangleiter, and Stephan Hartmann. 2016. Understanding (with) toy models. *The British Journal for the Philosophy of*

Science, <https://epub.ub.uni-muenchen.de/28990/1/Toy%2520Models_philsci_archive.pdf> (Accessed January 25, 2018).
Riegler, Alexander, and Igor Douven. 2009. Extending the hegselmann-krause model iii: From single beliefs to complex belief states. *Episteme* 6 (2): 145-63.
Rosenstock, Sarita, Cailin O'Connor, and Justin Bruner. 2016. In epistemic networks, is less really more? *Philosophy of Science* 84 (2): 234-52.
Šešelja, Dunja, and Christian Straßer. 2013. Abstract argumentation and explanation applied to scientific debates. *Synthese* 190: 2195-217.
Šešelja, Dunja, and Christian Straßer. 2014. Heuristic reevaluation of the bacterial hypothesis of peptic ulcer disease in the 1950s. *Acta Biotheoretica* 62: 429-454.
Šešelja, Dunja, Laszlo Kosolosky, and Christian Straßer. 2012. Rationality of scientific reasoning in the context of pursuit: drawing appropriate distinctions. *Philosophica* 86: 51-82.
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. *Journal of Artificial Societies and Social Simulation* 17 (3): 11.
Thimm, Matthias. 2012. A probabilistic semantics for abstract argumentation. In *ECAI*, 750-5.
Thoma, Johanna. 2015. The epistemic division of labor revisited. *Philosophy of Science* 82 (3): 454-72.
Weisberg, Michael. 2006. Robustness analysis. *Philosophy of Science* 73 (5): 730-42.
Weisberg, Michael, and Ryan Muldoon. 2009. Epistemic landscapes and the division of cognitive labor. *Philosophy of science* 76 (2): 225-52.
Wilensky, Uri. 1999. *NetLogo*. <http://ccl.northwestern.edu/netlogo/> (Accessed January 31, 2018). Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.
Zollman, Kevin J. S. 2007. The communication structure of epistemic communities. *Philosophy of Science* 74 (5): 574-87.
Zollman, Kevin J. S. 2010. The epistemic benefit of transient diversity. *Erkenntnis* 72 (1): 17-35.