Examining Network Effects in an Argumentative Agent-Based Model of Scientific Inquiry

AnneMarie Borg[1], Daniel Frey[3], Dunja Šešelja[1,2], and Christian Straßer *[1,2]

[1]Institute for Philosophy II, Ruhr-University Bochum, Bochum, Germany
[2]Center for Logic and Philosophy of Science, Ghent University, Ghent, Belgium
[3]Faculty of Economics and Social Sciences, Heidelberg University, Germany

Abstract

In this paper we present an agent-based model (ABM) of scientific inquiry aimed at investigating how different social networks impact the efficiency of scientists in acquiring knowledge. The model is an improved variant of the ABM introduced in [3], which is based on abstract argumentation frameworks. The current model employs a more refined notion of social networks and a more realistic representation of knowledge acquisition than the previous variant. Moreover, it includes two criteria of success: a monist and a pluralist one, reflecting different desiderata of scientific inquiry. Our findings suggest that, given a reasonable ratio between research time and time spent on communication, increasing the degree of connectedness of the social network tends to improve the efficiency of scientists.

1 Introduction

Agent-based models (ABMs) have in recent years been increasingly utilized as a method for tackling socio-epistemic aspects of scientific inquiry [6, 18, 21, 23, 24]. The primary value of this approach is that it allows us to tackle questions that are difficult to answer with qualitative 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 [23, 24] suggested that a high degree of connectedness of a social network may be epistemically harmful, and that there is a trade-off between the success of agents and the time they need to reach a consensus. Even though it has been shown in [15] that this result, dubbed as "Zollman effect", does not hold for a large portion of the relevant parameter space, structurally different ABMs have come to similar conclusions [9, 10].

The highly idealized nature of these ABMs makes it difficult to assess how relevant their findings are for actual scientific inquiry [15]. On the one hand, idealization and abstraction are necessary for simulations that aim at representing complex real world phenomena [14]. On the other hand, unless we include the most important 'difference making' factors figuring in the target phenomenon, the model might not represent any realistic scenario. Instead, it may represent only a logical possibility, uninformative about the real world.[1] Since due to their high degree of idealization such models operate at a significant representational distance from their intended

The final publication is available at Springer via https://link.springer.com/chapter/10.1007/978-3-662-55665-8_27.

*The research by AnneMarie Borg and Christian Straßer is supported by a Sofja Kovalevskaja award of the Alexander von Humboldt Foundation and by the German Ministry for Education and Research.

[1]Accordingly, we can distinguish between models that provide *how actually* and *how possibly* explanations. While some have suggested that modeling a possibility is epistemically valuable [11], others have argued that if a model merely captures a possibility, it is epistemically and pragmatically idle [19]; instead, the presented possibility has to be understood within a specified context, which makes it relevant for real world phenomena.

target, we are faced with the methodological difficulty to judge their adequacy and epistemic merit. What we can do is to compare models with the same intended target that differ in various important respects, such as in the way they structure the phenomenon, in the way some of its elements are represented, in their degree of abstraction and idealization, etc. The possible gain may be that in virtue of this we are able to identify results that are robust under different modeling choices, and, based on that, to identify causes and explanatory relevant properties underlying these commonalities. Similarly, if specific types of models give rise to different outcomes, we may sharpen our understanding of their possible target phenomena and how they apply to them. For example, we may be able to identify distinct sub-classes of phenomena to which the different types of models apply.

The ABM presented in this paper has the same intended target phenomenon as the ones presented in [23, 24]. Nevertheless, the representation of social networks, the content of the exchanged information, and the behavior of agents is different, which raises the question whether the "Zollman effect" also occurs in our model. As will be demonstrated, it does not. Given the differences between our models, this result may not be surprising. Nevertheless, it suggests that further investigation is needed to determine to which target phenomena the results of each model apply.

Our model is based on a recently developed argumentative ABM of scientific inquiry (AABMSI) [3]. It represents scientific interaction as argumentative in nature. This means that instead of a simple information flow that lacks critical assessment, in AABMSI agents may refute previously accepted or new information in view of counterarguments. Second, information about a given scientific theory is represented as consisting of a set of arguments (rather than being fully aggregated in a single value, representing one's credence in the given theory [23, 24] or the epistemic success of the theory [9, 10]). As a result, information sharing is represented as concerning particular parts of the given scientific theory, rather than an aggregated attitude about the whole theory. Third, receiving criticism triggers a search for defense of the given attacked argument. Fourth, the model takes into account that sharing information costs time and hence, that there is a trade-off between time spent on research and time spent on interaction.

In this paper we present an improved and more encompassing variant of AABMSI, aimed at examining the effects of different social networks on the efficiency of scientific inquiry. The model represents a situation in which scientists pursue different scientific theories, with the aim of determining which one is the best, and where they exchange arguments regarding their pursued theories. Compared to the model presented in [3], the current model introduces a number of improvements: 1) it employs the notion of social networks that is typically used in other ABMs of science, such as the complete graph, the wheel and the cycle [23, 24, 9, 10] – this allows for the representation of an increase in information sharing proportional to the population size, which is absent from AABMSI; 2) the heuristic behavior of agents is represented in a more adequate way; 3) an additional criterion of success has been introduced in order to examine the robustness of the results under different standards relevant in scientific practice; 4) the time cost of learning is now proportional to the amount of new information received by the agent; 5) the model has been computationally improved, allowing for statistical analysis of the data and for more reliable results.

Our findings suggest that a higher degree of connectedness leads to a more efficient inquiry, given a reasonable ratio between research time and time spent on communication. While our ABM is still too idealized to draw normative conclusions concerning scientific inquiry that would be useful, for instance, to policy makers,[2] it represents a step further in this direction.

The paper is structured as follows. In Sec. 2 we explicate the main features of our ABM. In Sec. 3 we present the main results of the simulations. In Sec. 4 we compare our model with an argumentation-based ABM introduced in [8], as well as with other ABMs of scientific interaction. We conclude the paper in Sec. 5 by suggesting further enhancements of our model and future research avenues.

[2]To this end, one of the tasks for future research is the empirical calibration of the parameters used in the model, as pointed out in [13].

2 The Model

Agents in our model represent scientists who inquire into a number of rivaling theories in a given scientific domain. Theories are represented by sets of arguments which scientists can discover and investigate. This is modeled in terms of an argumentative landscape where scientists can investigate arguments by spending time on them, which allows them to discover other arguments or argumentative attacks between arguments of different theories. In certain intervals scientists evaluate theories on the basis of their knowledge about the argumentative landscape and decide which theory to pursue further. Additionally, scientists are situated in communication networks. By communicating with other scientists they enhance their knowledge about the argumentative landscape which may help them to make more informed evaluations.

In the remainder of this section we describe the main components of our model in more detail: the underlying argumentative landscape, the behavior of agents and the notion of social networks.[3]

2.1 The argumentative landscape

An *abstract argumentation framework*, as introduced by Dung [7], is a directed graph $\mathcal{AF} = \langle \mathcal{A}, \rightsquigarrow \rangle$ where $\mathcal{A}$ is a set of abstract entities called *arguments* and $\rightsquigarrow$ is an *attack relation* between arguments. Where $a, b \in \mathcal{A}$, a *attacks* b if $a \rightsquigarrow b$. A set of arguments $\mathcal{S}$ is *conflict-free* if there are no $a, b \in \mathcal{S}$ such that $a \rightsquigarrow b$.

For our purposes it is useful to add another relation between arguments, the *discovery relation* $\hookrightarrow$. It represents possible paths scientists can take to discover arguments:[4] if $a \hookrightarrow b$, then b can be discovered by the scientists if a has been previously discovered.

A theory is represented by a conflict-free set of arguments connected in a tree-like graph by discovery relations. Argumentative attacks exist only between arguments of different theories. Formally, an *argumentative landscape* is a triple $\mathcal{L} = \langle \mathcal{A}, \rightsquigarrow, \hookrightarrow \rangle$ where the set of arguments $\mathcal{A}$ is partitioned into m theories $\langle \mathcal{A}_1, \ldots, \mathcal{A}_m \rangle$ such that for each $i \in \{1, \ldots, m\}$ the theory $T_i = \langle \mathcal{A}_i, a_i, \hookrightarrow \rangle$ is a tree with root $a_i \in \mathcal{A}_i$ and

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

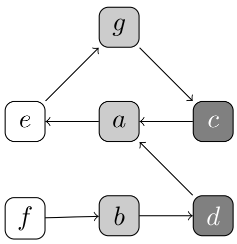
This definition of the attack relation $\rightsquigarrow$ ensures that each theory is conflict-free and that attacks only occur between members of different theories.

At the beginning of the simulation only the roots of each theory (representing a basic underlying hypothesis) are visible to the agents. During the simulation the agents gradually discover arguments and attacks between them (see also Section 2.2). Each argument a has a *degree of exploration*: $\mathfrak{expl}(a) \in \{0, \ldots, 6\}$ where 0 means that the argument has not been discovered and 6 means that the argument has been fully researched and it cannot be explored further. By spending time on an argument a, a scientist increases its degree of exploration. The higher the degree of exploration of a, the higher is the likelihood that the connected arguments will become visible. This concerns both arguments connected to a via the discovery relation in the same theory and arguments that attack a or that are being attacked by a. If $\mathfrak{expl}(a) = 0$ [resp. $\mathfrak{expl}(a) = 6$] no [resp. all] relation[s] from a to other arguments have been discovered.

Agents only have subjective knowledge of the landscape, which consists of arguments, their respective degrees of exploration, and attacks. During the simulation agents gain knowledge, on the one hand, by exploring the landscape, and on the other hand, by communicating with other agents. This way two agents may have different knowledge about the degree of exploration of an argument or about discovered relations. See also Section 2.3.

[3]Our ABM is created in NetLogo [22]. The source code is available at: `https://github.com/g4v4g4i/ArgABM/tree/LORI-VI-2017`.

[4]Other ways of discovering arguments and attacks (via social networks) are discussed below.



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 1: Argumentation Framework 1

2.2 Basic behavior of agents

Our model is round based. Each round every agent performs actions that are among the following:

1. the agent investigates the argument a she is currently situated at (i.e. she increases $\mathfrak{expl}(a)$) and while doing so she gradually reveals outgoing discovery relations as well as attacks from and to a;
2. the agent explores her current branch of the theory further by moving along a discovery relation to a neighboring argument (that she can see);
3. the agent leaves her current theory and moves to an argument of a rivaling theory (that she can see).

Every round each agent decides (based on a certain probability) whether to stay on her current argument (option 1) or to move to a new argument in her direct neighborhood (relative to the discovery relation, option 2). If she has reached a leaf of her branch, which is fully explored (i.e., $\mathfrak{expl}(a) = 6$), she backtracks on this branch to find an argument that is not fully explored. In case this fails, she moves to another not fully explored argument in the same theory.

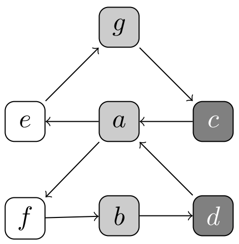
Additionally, every 5 rounds agents consider whether they are still working on the theory they consider the best (with respect to their current subjective knowledge of the landscape). Depending on this decision, they continue with 1. and 2., or move to an alternative theory and start to explore that one (3.). Their decision is based on an evaluation of the *degree of defensibility* of a theory.

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.

Let us give a more precise formal definition. First, 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 (we say that a' defends a from the attack by b). Since every theory is conflict-free, 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.[5] The *degree of defensibility* of T is equal to the number of defended arguments in T.

Example 1 *Fig. 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 theory 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*

[5]Given that theories in our model are conflict-free, our notion of admissibility is actually the same as the one introduced in [7]. In Dung's terminology, our sets of defended arguments correspond to preferred extension (which are exactly the maximally admissible sets), except that we determine these sets relative to given theories.



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 2: Argumentation Framework 2

b, it doesn't count as a 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.

Fig. 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.

A scientist decides to work on another theory T' instead of her current theory T if the degree of defensibility of T' surpasses the one of T by a certain margin. We represent the objectively best theory as a fully defended one.

Moreover, agents are equipped with heuristic abilities. An agent that encounters an attack from an argument b on the argument a, she is currently working on, will try to find a defense for this argument. For this she will consider all the arguments in her subjective knowledge that belong to the same theory as a. If there is an argument a' among these that can *potentially defend* a from the attack, she will begin to investigate it. That a' potentially defends a from b means that there is an attack from a' to b but this attack has not yet been added to the subjective representation of the landscape of our agent (e.g., since $\mathfrak{expl}(a')$ is too low or since it has not been communicated to her). This means that agents are equipped with 'professional hunches' which help them to tackle problems in their theories.[6]

2.3 Social networks

Besides discovering the argumentative landscape by exploring it on her own (see Section 2.2), an agent can share information about the landscape with other agents.

At the start of a simulation agents are divided into local *collaborative networks*, each consisting of exactly five individuals working on the same theory. During the simulation each agent gathers information (i.e., the degree of exploration of arguments, discovery and attack relations) on her own. Agents of the same collaborative network have the same subjective knowledge of the landscape since whenever an agent learns something new, this is communicated with the other agents in the same collaborative group.

Additionally, the collaborative groups form a *community network*. These have one of the following three structures: a *cycle*, in which each collaborative group is connected to exactly two other groups, a *wheel* which is similar to the cycle, except that a unique group is connected to every other group, and a *complete graph* where each group is connected to all other groups (Fig. 3). Every five rounds, randomly chosen *representative* agents of the collaborative groups communicate along the communication channels of the community network. The different network structures allow us to represent varying degrees of information flow in the scientific community, with the cycle representing the lowest and the complete graph the highest degree of information sharing.[7]

Representative agents do not share their whole knowledge of the landscape with agents from other collaborative networks. Instead, they share the knowledge they have obtained recently

[6] Such hunches are not considered when agents evaluate theories.

[7] In contrast to the current model, in AABMSI [3] network structures are generated probabilistically in specific time intervals.

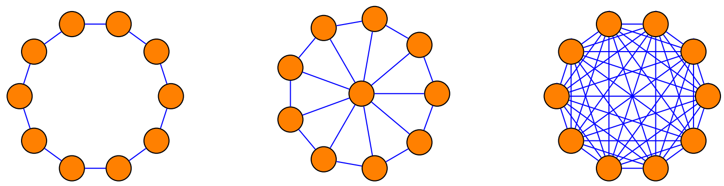


Figure 3: A cycle, a wheel and a complete graph. Each node is a collaborative group, while the edges represent communication channels.

by their own exploration of the landscape, which consists of the argument they are currently at, its neighboring arguments connected via the discovery relation, their respective degrees of exploration, and attacks to and from their current argument. One way of interpreting this limited knowledge sharing is by considering this to be a situation where the agent writes a paper or gives a presentation on her current research results.

An alternative interpretation of collaborative groups is that they represent one individual working on 5 different arguments (papers, hypotheses, etc.) in a given theory. Whenever this individual communicates with other individuals from the community network it exchanges the information from the neighborhood of one of the 5 arguments she is currently engaged with.

Our model takes into account that information sharing and especially receiving information is time costly: agents who receive information (the representative agents of each collaborative network) are blocked from further exploration for a certain number of rounds. The costs of information sharing are proportional to the amount of new information obtained.[8]

Finally, we distinguish two types of information sharing, characterizing two types of scientists. A *reliable* agent shares all her recent discoveries, whereas a *deceptive* agent withholds the information about attacks on her current theory. This way, agents receiving information from a deceptive agent from theory T might come to a more favorable interpretation of T than if they would have communicated with a reliable agent [4].

3 The main findings

We will now specify the parameters used in the simulations and then present our most significant results.

3.1 Parameters used in simulations

We have run simulations with a landscape consisting of 3 theories. The landscape is created in three steps: First, each theory is represented by a tree (as explained in Section 2.1) of depth 3 such that each node has 4 children (except for the leaves) resulting in 85 arguments per theory. Second, with a chance of 0.3, each argument gets randomly attacked by an argument from another theory. Third, for every argument in theories T_2 and T_3 that attacks an argument in theory T_1 and that is not attacked by an argument from T_1 we add an attack from some random argument in T_1. Thus, it is made sure that T_1 is the objectively best theory and as such is fully defended from all attacks. In this way we wish to represent a scenario in which theories that are rivals to the best one are worse, though not completely problematic (as it would be the case with pseudo-scientific theories).

Simulations were run 10.000 times for each of the scenarios with 10, 20, 30, 40, 70 and 100 agents. The scenarios are created by varying::

1. the community network: in the form of a cycle, a wheel and a complete graph;

[8]A representative agent is excluded from research for 1–4 rounds: she always pays the basic cost of information sharing which is 1 round, and in addition, for every 2 fully explored arguments she will pay an additional round. The cost of learning an attack is equivalent to learning one degree of exploration of an argument.

2. two types of information sharing: reliable and deceptive.[9]

A simulation stops when one of the theories is completely explored. At this point all the agents have one more chance to make their final evaluation and choose their preferred theory. We then evaluate whether the agents have been successful according to the following criteria (where T_1 is the objectively best theory):

1. *monist criterion*, according to which a run is considered successful if, at the end of the run, all agents have converged onto T_1;
2. *pluralist criterion*, according to which a run is considered successful if, at the end of the run, the number of agents working on T_1 is not smaller than the number of agents working on any of the other theories.

The monist criterion is the standard notion of success, often employed in other ABMs of science, e.g. [23, 24]. The pluralist criterion, on the other hand, is motivated by the philosophical conception of scientific pluralism, according to which a parallel existence of multiple theories in a given scientific domain is considered epistemically and heuristically beneficial e.g. [5]. This means that the convergence of all scientists onto the objectively best theory isn't a primary epistemic concern for pluralists. Rather, what matters is that the best theory is one of the most actively researched theories.[10]

3.2 Results

In this section we describe the most important findings of the simulations.

Reliable vs. deceptive agents. With respect to both criteria of success, reliable agents are clearly more successful than the deceptive ones (Fig. 4 and 5), while being only slightly slower (Fig. 6).[11]

The degree of connectedness. In case of reliable agents, the complete graph tends to outperform the wheel and cycle, with respect to both criteria of success (Fig. 4 and 5), as well as with respect to the speed of exploration (Fig. 6). In other words, a higher degree of connectedness tends to lead to a more efficient inquiry.

In case of deceptive agents, the situation is a bit trickier. On the one hand, higher degrees of connectedness are also beneficial for the success according to the monist criterion (Fig. 5). However, the effect of connectedness is inverse for the pluralist criterion given populations with up to 70 agents (Fig. 4). A possible explanation for this asymmetry between the two success criteria is that deceptive agents in more connected networks share more false positives. As a result, if one theory is explored by a larger number of agents than either of the other theories, for the agents on this theory it will be easier to attract the whole population to it, leading to a fast, possibly wrong, convergence. In contrast, deceptive agents in the less connected networks will spread less information among each other, resulting in fewer cases of wrong convergence. For populations larger than 70 agents these premature convergences of the complete graph are prevented by the fact that each theory has on average enough researchers praising it and pointing out problems with the other theories so that the false positives are debunked as such more often.

[9] Further parameters, with short explanations, are as follows. The *move probability* (set to 0.5) together with the degree of exploration of the argument an agent is situated at, determines the chance that she will move to another argument every 5 rounds (the move incentive is further decreased by $\frac{1}{5}$ for time steps in between). The *visibility probability* (set to 0.5) is the probability with which a new attack is discovered when an agent further explores her argument. The *research speed* (set to 5) determines the number of time steps an agent has to work on an argument a before a reaches its next level of exploration. The *strategy threshold* (set to 0.9) concerns the fact 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 be researched by agents. The *jump threshold* (set to 10) concerns the number of evaluations an agent can remain on a theory that is not one of the subjectively best ones.

[10] While our criterion is moderately pluralist, a more radical version would make plurality a necessary condition of success (i.e. populations would be punished for converging on one theory). We leave this consideration for future research.

[11] The plots concern the landscape consisting of three theories. The results were similar in case of two theories in all the discussed respects, except that the agents were comparatively more efficient.

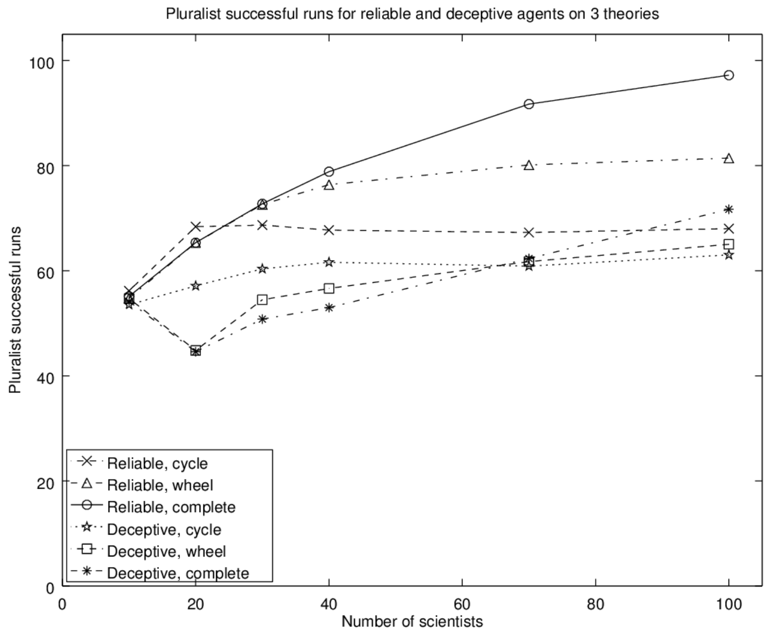


Figure 4: Pluralist criterion of success, reliable and deceptive agents

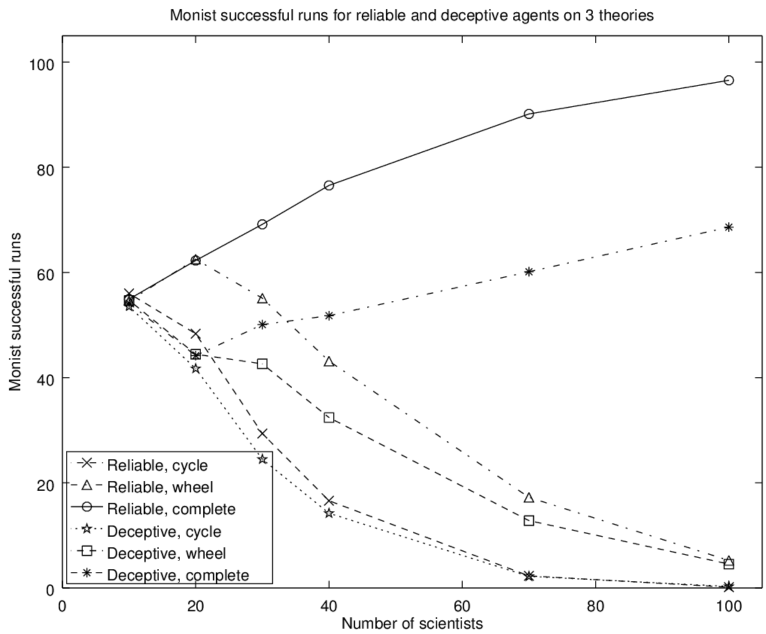


Figure 5: Monist criterion of success, reliable and deceptive agents

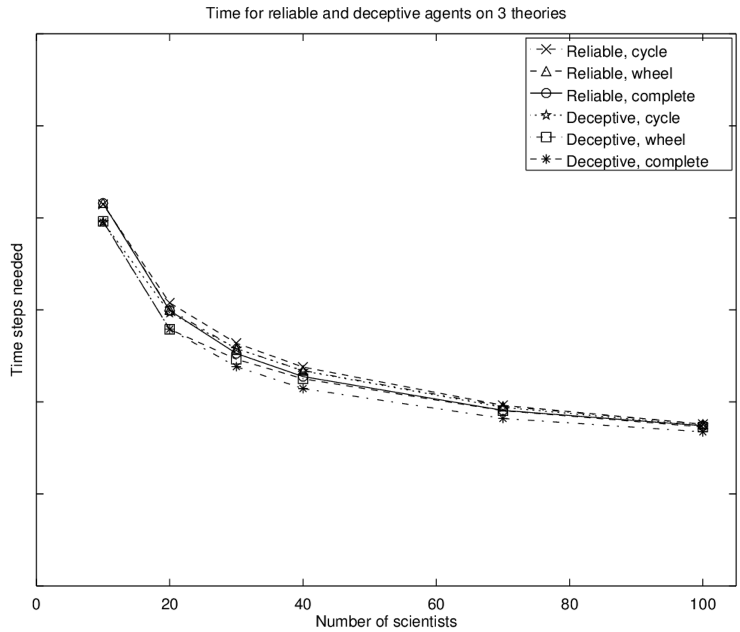


Figure 6: Time needed

Size of the community. Larger populations tend to be more efficient if agents are connected in the complete graph. In contrast, larger populations of agents in less connected networks perform similarly to smaller ones with respect to the pluralist criterion, and drastically worse with respect to the monist criterion. The latter results from the lack of sufficient information flow, which is why larger populations fail to converge on any (including the best) theory.

4 Discussion

In this section we will first comment on some of our results and then we will compare our model with other ABMs designed to tackle similar questions.

We focus on two particularly interesting results: the impact of deceptive information sharing in contrast to the reliable one, and the impact of the degree of connectedness on reliable agents. Concerning the former, deceptiveness has not been studied in ABMs of science to a large extent.[12] This is, however, an important issue for the efficiency of scientific inquiry. For instance, withholding results that would undermine one's theory, typical for scientific fraud, is an example of deceptive information sharing. Even though deception is clearly problematic from the perspective of ethics of scientific conduct, its effects on the efficiency of the given scientific community aren't immediately clear. For instance, one could assume that presenting one's theory in positive light, in spite of the early problems can attract new researchers and help in developing it further. Our findings suggest that this is in general not the case, i.e. that deception tends to be epistemically harmful. More precisely, assuming that the whole community consists of deceptive scientists, and that scientists prefer theories that have the highest number of defensible results, deception leads to significantly less successful (though slightly faster) inquiry. Whether these results hold also under other assumptions remains to be examined in future research.

[12]One example of an ABM that studies deception in science is [12], which examines the effects of a deceptive agent in a community of epistemically pure agents. The authors show that in general a higher degree of connectivity helps against deceptive information. While our model doesn't examine the case of mixed (reliable and deceiving) agents, our results are, generally speaking, in line with their results.

Our findings suggest that increasing the degree of connectedness of the communication network tends to be epistemically beneficial. This contrasts with findings obtained in other ABMs [9, 10, 23, 24], according to which agents connected in a cycle perform better than agents connected in a complete graph. In order to see whether and in which sense our results challenge the latter results we first need to highlight some differences between our approach and these ABMs.

First, while in the latter agents are directly connected in the given networks, we employ a more structured approach by distinguishing between collaborative (local research) groups and communal networks between these groups. Note that in the real world it is impossible for each member of a larger scientific community to invest time in communicating with all the other members for the simple fact that communication (such as, e.g. reading papers) costs time, which could otherwise be spent on doing research. Thus, we have started from the assumption that scientists only have limited time for interacting with members of other collaborative groups, while they fully share information within their own collaborative projects. Altogether this means, however, that the highest degree of connectedness examined in our model gives rise to less information flow than in models in which agents are directly linked into a complete graph. Second, the content of the information shared among our agents is more localized and patchy since only representative agents exchange information about local aspects of their theory, namely their current argument and its neighborhood. As a result, agents from different collaborative groups end up having more diverse subjective representations of the landscape than e.g. agents in [23, 24]. Third, our representation of interaction includes a critical component. This is important for a model designed to examine the efficiency of scientific knowledge acquisition due to the fact that criticism has been shown to be truth conducive since it allows for false beliefs to be exposed as such [2]. Finally, similarly to ABMs that employ epistemic landscapes [9, 10, 21], our argumentative landscape allows for the representation of the process of discovery and its timeline. However, unlike the models employing epistemic landscapes, the information that agents encounter via an argumentative landscape is defeasible. This feature allows not only for the representation of critical interaction, but also for specific heuristic behaviors,[13] such as the search for a defense of an argument in case it has been attacked.[14] In this sense, our respective results might concern a different kind of scenario and thus refer to different target phenomena.

Having explicated the differences between our and other ABMs, it is important to notice though that there is no reason to assume that our model introduces more problematic idealizing assumptions than the previous ABMs of scientific interaction when it comes to the representation of a typical scientific inquiry. To the contrary, it includes a number of assumptions directly relevant for its adequate representation. Thus, our findings suggest that the results obtained by others models might not hold for usual cases of scientific inquiry. Instead, they may hold only for some very specific contexts. Which subclass of the phenomenon of scientific inquiry each of these models reliably represents, remains to be tackled in future research.

Finally, let us compare our model with Gabbriellini and Torroni's (G&T) ABM [8]. Their aim is to study polarization effects, e.g., in online debates. Similarly to our approach, their model is based on an abstract argumentation framework. Agents start with an individual partial knowledge of the given framework and enhance their knowledge by means of communication. Since G&T do not model inquiry, their agents cannot discover new parts of the graph by means of 'investigating' arguments. Rather, they exchange information by engaging in a dialogue. This way, agents may learn about new arguments and attacks but also remove attacks. Whether new information is incorporated in the knowledge of an agent depends on the trust relation between the discussants. The beliefs of agents are represented by applying Dung-style admissibility-based semantics to the known part of the argumentation framework of an agent. This is quite different from our model

[13] Interestingly, comparing the results of our model that employs the heuristic behavior (HB) and the results produced when HB is removed, shows that HB has hardly any impact on the success of agents, and in some cases it even slightly lowers their success. This seems to suggest that HB, by making agents stay on an undefended argument, waiting to find how to defend it, shields not only the best theory but also the worse ones, leading to an overall less successful inquiry. Examining this issue in more detail remains a task for future research.

[14] Another important difference between our ABM and those in [23, 24] is that the latter examine a fringe case of epistemically similar theories, which makes distinguishing the best one difficult.

where the underlying graph topology is given by several discovery trees of arguments representing scientific theories and attacks between them. This additional structure of the argumentation graph is essential since we do not model the agents' beliefs in individual arguments but rather evaluative stances of agents that inform their practical decision of which theory to work on. While an admissibility-based semantics would lead to extensions that feature unproblematic sets of arguments from different theories (ones that form conflict-free and fully defensible sets), in our approach agents pick theories to work on. For this, they compare the merits of the given theories, pick the one that is most defended, and employ heuristic behavior to tackle open problems of theories. It will be the topic of future research to include dialogue protocols that are relevant for scientific communication, such as information-seeking, inquiry and deliberation dialogues [20].

5 Conclusion

In this paper we have presented an argumentative ABM aimed at modeling the argumentative nature of scientific inquiry. The model is designed to examine how different kinds of social networks affect the efficiency of scientists in acquiring knowledge. Under the assumption that, in order to conduct their inquiry, scientists only have limited time to spend on communicating with others, our results suggest that a high information flow tends to be epistemically beneficial.

A variety of enhancements can be added to our model in order to make it apt for tackling similar or related questions. First, our current notion of the degree of defensibility represents scientists who prefer theories that exhibit a greater number of defensible results than their rivals. An alternative notion of defensibility would punish theories for having more anomalies (indefensible arguments) than their rivals, thus representing scientists who stick to their theories as long as they are not too anomalous (irrespective of how many positive results they have). Second, adding an explanatory relation and a set of explananda [16] would allow for a more refined representation of the desiderata of scientific theories and 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 argumentation frameworks, such as probabilistic semantics [17], values [1], etc. can be introduced in future variants of our ABM. In addition to examining the impact of social networks, the model can be used to examine different heuristic behaviors and evaluations that guide scientific inquiry.

References

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

[2] Gregor Betz. *Debate dynamics: How controversy improves our beliefs*, volume 357. Springer Science & Business Media, 2012.

[3] AnneMarie Borg, Daniel Frey, Dunja Šešelja, and Christian Straßer. *An Argumentative Agent-Based Model of Scientific Inquiry*, pages 507–510. Springer International Publishing, Cham, 2017.

[4] Martin Caminada. Truth, lies and bullshit: distinguishing classes of dishonesty. In *In: Social Simulation Workshop at the International Joint Conference on Artificial Intelligence (IJCAI)*, 2009.

[5] Hasok Chang. *Is Water H2O? Evidence, Pluralism and Realism.* Springer, 2012.

[6] Igor Douven. Simulating peer disagreements. *Studies in History and Philosophy of Science Part A*, 41(2):148–157, 2010.

[7] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.*, 77:321–358, 1995.

[8] Simone Gabbriellini and Paolo Torroni. A new framework for ABMs based on argumentative reasoning. In *Advances in Social Simulation*, pages 25–36. Springer, 2014.

[9] Patrick Grim. Threshold phenomena in epistemic networks. In *AAAI Fall Symposium: Complex Adaptive Systems and the Threshold Effect*, pages 53–60, 2009.

[10] Patrick Grim, Daniel J Singer, Steven Fisher, Aaron Bramson, William J Berger, Christopher Reade, Carissa Flocken, and Adam Sales. Scientific networks on data landscapes: question difficulty, epistemic success, and convergence. *Episteme*, 10(04):441–464, 2013.

[11] Stephan Hartmann, Alexander Reutlinger, and Dominik Hangleiter. Understanding (with) toy models. *The British Journal for the Philosophy of Science*, 2016.

[12] Bennett Holman and Justin P Bruner. The problem of intransigently biased agents. *Philosophy of Science*, 82(5):956–968, 2015.

[13] Carlo Martini and Manuela Fernández Pinto. Modeling the social organization of science. *European Journal for Philosophy of Science*, pages 1–18, 2016.

[14] Ryan Muldoon and Michael Weisberg. Robustness and idealization in models of cognitive labor. *Synthese*, 183(2):161–174, 2011.

[15] Sarita Rosenstock, Cailin O'Connor, and Justin Bruner. In epistemic networks, is less really more? *Philosophy of Science*, 2016.

[16] Dunja Šešelja and Christian Straßer. Abstract argumentation and explanation applied to scientific debates. *Synthese*, 190:2195–2217, 2013.

[17] Matthias Thimm. A probabilistic semantics for abstract argumentation. In *ECAI*, pages 750–755, 2012.

[18] Johanna Thoma. The epistemic division of labor revisited. *Philosophy of Science*, 82(3):454–472, 2015.

[19] Raphael Van Riel. The content of model-based information. *Synthese*, 192(12):3839–3858, 2015.

[20] Douglas Walton and Erik CW Krabbe. *Commitment in dialogue: Basic concepts of interpersonal reasoning*. SUNY press, 1995.

[21] Michael Weisberg and Ryan Muldoon. Epistemic landscapes and the division of cognitive labor. *Philosophy of science*, 76(2):225–252, 2009.

[22] Uri Wilensky. Netlogo.(http://ccl.northwestern.edu/netlogo/). *Center for Connected Learning and Computer Based Modeling, Northwestern University*, 1999.

[23] Kevin J. S. Zollman. The communication structure of epistemic communities. *Philosophy of Science*, 74(5):574–587, 2007.

[24] Kevin J. S. Zollman. The epistemic benefit of transient diversity. *Erkenntnis*, 72(1):17–35, 2010.