Robustness and Idealizations in Agent-Based Models of Scientific Interaction

(Forthcoming in *The British Journal for the Philosophy of Science*)

Daniel Frey[1] and Dunja Šešelja[*2]

[1] *Faculty of Economics and Social Sciences, Heidelberg University*
[2] *Institute for Philosophy II, Ruhr-University Bochum*

Abstract

The paper presents an agent-based model (ABM) of scientific interaction aimed at examining how different degrees of connectedness of scientists impact their efficiency in knowledge acquisition. The model is built on the basis of Zollman's (2010) ABM by changing some of its idealizing assumptions that concern the representation of the central notions underlying the model: epistemic success of the rivaling scientific theories, scientific interaction and the assessment in view of which scientists choose theories to work on. Our results suggest that whether and to which extent the degree of connectedness of a scientific community impacts its efficiency is a highly context-dependent matter since different conditions deem strikingly different results. More generally, we argue that simplicity of ABMs may come at a price: the requirement to run extensive robustness analysis before we can specify the adequate target phenomenon of the model.

1 Introduction

Recent studies into social aspects of scientific inquiry have been increasingly employing agent-based models (ABMs). A number of articles presenting results of such computer-based simulations have suggested that increased information flow among scientists can be epistemically harmful (Zollman, 2007, 2010, Grim, 2009, Grim et al., 2013, Kummerfeld and Zollman, 2016). A specific feature of these models is that they are highly idealized, based on simplified representation of scientific inquiry. As such, they can be characterized as 'minimal' or 'toy models' (Reutlinger, Hangleiter, and Hartmann, 2016), the virtue of which

[*]The order of authors is alphabetical; both authors contributed equally to this paper. To contact the authors, please write to dunja.seselja@rub.de.

is their capacity to offer simple explanatory hypotheses about the causal dependencies underlying represented phenomena.[1] It is then not surprising that the findings of these models have been adopted by philosophers of science and social epistemologists (e.g. Wray, 2011, Goldman and Blanchard, 2016).

Zollman's ABMs have in this respect been particularly influential. Nevertheless, a recent examination by Rosenstock, O'Connor, and Bruner, 2017 has shown that Zollman's (2007; 2010) results hold only for a small portion of the relevant parameter space. While the authors suggest that Zollman's findings should be interpreted as relevant only for the context of difficult scientific inquiry, the precise relation between these models and the real-world phenomena has remained an open question. In particular, it has remained unclear whether the results of the models are robust under changes in idealizing assumptions even if we restrict the target phenomenon to difficult inquiry.

In this paper we propose an ABM of scientific interaction, building on Zollman's (2010) model. The aim of our simulations is to address the above question by examining whether adding certain assumptions to Zollman's model – while keeping it as simple as possible – affects the conclusions drawn from it. In this way we will investigate whether these additional assumptions are 'difference-making' in the sense of Strevens, 2013 and Weisberg, 2007. According to the latter, "a minimalist model contains only those factors that *make a difference* to the occurrence and essential character of the phenomenon in question" (Weisberg, 2007, p. 642, italics in original). Hence, if the introduction of new factors, which are typical features of a difficult scientific inquiry, turns out to make a difference to the results, then we can conclude that the starting model was not adequately minimal for the given target phenomenon. Moreover, this will help in specifying the application context to which Zollman's results do apply, and with respect to which his model can be considered adequately minimal.

We will start by outlining the main features of Zollman's (2010) model (Section 2), which represents the following scenario: a scientific community is confronted with two rivaling theories, only one of which is true (or empirically successful), though scientists do not know which one. The question the model examines is how different degrees of connectedness among scientists impact their efficiency in converging on the right theory. We will then turn to our adjustments of some idealizing assumptions in this model. Our first adjustment concerns the representation of scientists' epistemic success in gathering evidence for a given theory over the course of their inquiry. In contrast to Zollman's representation of theories in terms of bandits with static probabilities of success, we will represent them in terms of *restless bandits*, which are such that their probability of success changes over time (Section 3). Next, we will introduce a critical component to the representation of scientific interaction, assuming that critique among scientists is epistemically beneficial (Section 4). Finally, we will introduce two additional assumptions concerning theory choice: first, rational inertia that scientists have towards their current theories (Section 5) and a threshold

[1]One of the most prominent examples of this kind of models is Schelling's model of social segregation (Schelling, 1971), which has remained influential in various domains of social sciences (see e.g. Bruch and Atwell, 2015).

within which rivaling theories are considered to be equally promising (Section 6). As we will argue, each of these four assumptions is either a typical feature of scientific inquiry in general, or of difficult inquiry in particular. By changing the assumptions one after another, we will show in which way they affect the performance of communities that are characterized by different degrees of connectedness.[2]

Our results suggest that the answer to this question varies from one scenario to another, depending on the conditions characterizing the given context of difficult inquiry. In Section 7 we discuss the significance of these results. While our model should still be understood as exploratory in nature (rather than providing normative conclusions about actual scientific inquiry),[3] it can nevertheless contribute to the specification of the domain of phenomena that Zollman's model adequately captures. Section 8 concludes the paper.

2 Zollman's (2010) model

Zollman's ABM is designed to answer the following question: given a situation in which a scientific community investigates multiple rivaling theories in the given domain, which social structures do most efficiently lead the community to a consensus on the best theory? Social structures here stand for different ways in which information flow among the given scientists occurs, specified in terms of the number of scientists and paths via which information is shared.

The main idea behind this ABM (based on the framework developed by Bala and Goyal, 1998) is that the process of scientific inquiry can be tackled as a type of *bandit problem.* The so-called bandit problems, usually discussed in the context of economics and game theory, concern the following question: if a gambler is confronted with different slot machines, at which point should she stop testing which machine maximizes her reward, and stick with the one that seems the best in this respect? Zollman suggests that an analogous situation occurs in the context of scientific inquiry, where we can ask: at which point should a scientist stop testing different hypotheses and stick with the most promising one? The payoff of a slot machine here corresponds to the success of applying the given hypothesis (or a method, or a theory), while the objective probability of success (OPS) of a slot machine corresponds to the OPS of the given hypothesis (or a method, or a theory). Hence, just like a gambler faces a

[2]Borg et al., 2017a,b, 2018 examine the robustness of Zollman's results under different modeling assumptions as well, by employing an argumentation based ABMs. While their models do not reproduce Zollman's results, the authors point out that they may represent a different target phenomenon from the one captured by Zollman's model. Kummerfeld and Zollman, 2016 present a variant of Zollman's (2010) model, showing that the results change if scientists intentionally keep on experimenting with an inferior theory. Similarly to the latter, we employ a model structurally related to Zollman's one, but rather than focusing only on assumptions underlying the representation of one's research strategy, we introduce other relevant assumptions, focusing on the context of difficult inquiry.

[3]As pointed out by Martini and Pinto, 2016, gaining normative insights on the basis of simulations also requires their empirical calibration (see also Footnote 24).

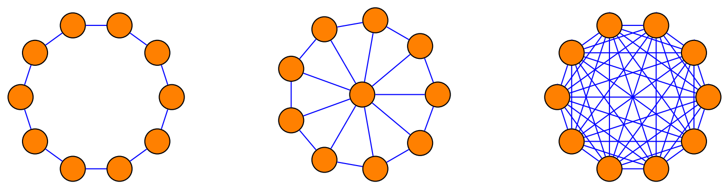

Figure 1: A cycle, a wheel and a complete graph. The nodes in each graph represent agents, while the edges that connect the nodes represent transmission of information between two agents.

choice between different slot machines, so does a scientist face a choice between different hypotheses.

Adding a social dimension to this problem raises the following question: if more than one gambler is trying to determine which bandit is the most profitable one, how does the information flow among the gamblers influence their respective choices? Analogously, we can ask: if more than one scientist is working in the same domain consisting of rivaling theories, how does the information flow among them influence their respective decisions as for which theory to pursue?

The model is designed as a computer simulation, which is round based. At the start of a run each agent is assigned a random prior value for each of the two rivaling theories. Each round an agent makes 1,000 pulls, each of which can be a success or failure, where the probability of success is given by the OPS of the respective theory. Agents then update their beliefs via Bayesian reasoning (modeled by means of beta distributions),[4] based on their own success and the success of some other agents, namely those with whom they are connected in a social network. Zollman investigates the efficiency of agents in converging onto the hypothesis with the best OPS in three kinds of social networks: a cycle, a wheel and a complete graph (see Figure 1). In the cycle, every agent exchanges information with two of her neighbors. The wheel has the same information flow as the cycle, except that in addition one agent exchanges information with all the other agents. Finally, in the complete graph every agent exchanges information with every other agent.

Zollman's results show that agents connected in the cycle score the best, those connected in the wheel score worse, and those connected in the complete graph score the worst. This suggests that information flow via highly connected groups can be epistemically harmful:

> It would appear here that the amount of information distributed is negatively impacting the ability of a social group to converge on the correct methodology. Initially suggestive information is causing everyone to adopt one particular methodology. (p. 28)

However, once agents are modeled as biased towards hypotheses with which

[4]For more details on Bayesian reasoning employed in Zollman's model, which we have partially retained in our own model, see Section 3.2.

they initially start, the results become inverse.[5] In other words, if agents are modeled as more resistant to changing their beliefs in view of new information, misleading initial results can't infect the entire community:

> ... when our agents have very extreme priors, even rather large amounts of information will not cause them to discard their prior beliefs. (p. 31)

Thus, Zollman's model suggests two ways of reducing errors in the learning process of a scientific community: either the information flow needs to be restricted, or scientists need to be initially biased towards their pursued hypotheses. In the remainder of the paper we will focus on the former suggestion, and examine under which conditions it holds once we alter some of the model's assumptions.[6]

3 Static vs. dynamic epistemic success

We will now take a closer look at four idealizing assumptions present in the above model and motivate their replacement. We start with the assumption that concerns the epistemic success of the given theories.

3.1 Introducing the notion of dynamic epistemic success

Zollman represents the OPS of the given theories or methodologies[7] as fixed values of 0.5 and 0.499, respectively. Note that the numerical proximity of these values is crucial for obtaining his results (as shown by Rosenstock, O'Connor, and Bruner, 2017). Now, in case of the classic interpretation of bandit problems, where a gambler is facing different slot machines, it indeed makes sense to assume that the OPS of these bandits remains stable over the course of time. It also makes sense to assume that a gambler should prefer the 0.5 bandit, even though its OPS is only 0.001 higher than the OPS of the other bandit: for a gambler this difference might result in significantly higher payoffs over the course of time.

However, when it comes to scientific inquiry, these assumptions don't seem very plausible. On the one hand, if the OPS of the two theories is meant to represent how likely it is that each is successfully applied in a given domain, then it is not clear why scientists are successful only if they converge on the 0.5 one.

[5] This is done by assigning more extreme distributions (i.e. drawing α/β from a larger interval) as priors from which agents start their updates, thus allowing for a more conservative outcome of the updating process. An interesting detail to note is that the reversal of results is only due to the cutoff point at which Zollman chooses to end his simulation. If the model is run sufficiently long, agents in all networks will end up on the correct theory: since extreme priors allow for the theoretical diversity to be preserved, agents in all networks have enough time to gather evidence and make the right choice.

[6] For a model that examines the impact of biases on scientific inquiry, building on Zollman's one, see Holman and Bruner, 2015.

[7] He uses the terms 'theory' and 'methodology' interchangeably. See below Section 3.2 for a possible reason to stick with only one of these notions.

Clearly, applying the 0.499 theory might still be fruitful in certain contexts, and retaining it as a fruitful alternative seems to be a wise choice. On the other hand, if one theory is meant to be superior to another to the extent that abandoning its rival is warranted, then we would expect the difference between the two theories to increase over the course of inquiry. More precisely, we would expect scientists working on the objectively better theory to eventually become successful in obtaining corroborating evidence more than 50% of the time as they improve their methodology and therefore reduce the rate of false negatives. Conversely, we would expect that scientists pursuing the objectively worse theory become less and less successful in corroborating their theory, as they reduce the rate of false positives (but see below for a discussion on other possible dynamics).

The research on peptic ulcer disease (PUD), which Zollman, 2010 uses as a case study motivating his model, is in fact a nice example of the above dynamics. The bacterial hypothesis of PUD exhibited a low degree of success in the 1940s and it was abandoned in favor of the objectively worse acidity hypothesis. Nevertheless, the former had its comeback with Warren and Marshall's discovery of Helicobacter pylori, bacteria that turned out to be the major cause of the disease. Even though the two hypotheses appeared equally promising in the first half of the twentieth century, the contemporary antibiotic treatment (based on the bacterial hypothesis) exhibits the empirical success rates of around 90% vs. less than 30% for the antacid treatment (Hosking et al., 1994; Moayyedi et al., 2000), a far cry from the almost indistinguishable success rates for both theories employed in Zollman's paper.[8] Thus, instead of a static notion of epistemic success of the given theories, a more plausible option seems a dynamic notion, according to which scientists gradually improve their methods.

3.2 Implementation and results for the basic setup

We will model the dynamic aspect of inquiry by representing the epistemic success of the theories in terms of *restless bandits*, that is, bandits whose OPS changes over time. More precisely, we will assume that if scientists pursue a true theory its OPS will gradually increase, while in case they pursue a false theory its OPS will decrease. Hence, we will start from the assumption that one's methodology improves over time, where such improvement is truth-conducive. In this way we will represent a gradually increasing gap between the epistemic success of the two theories.

In order to keep things conceptually clear, we will replace the notion of OPS with the notion of a scientist's *current probability of success* (CPS), assigned

[8]A more charitable reading of Zollman's scenario would be to understand it as representing scientific inquiry within a limited time frame which is too narrow for scientists to sufficiently improve their methodology in order to realize they have made a mistake. The success rates of 0.5 (resp. 0.499) would in that case represent only temporary snapshot of how well the two hypotheses perform on average (which might be static for a sufficiently short time frame). In view of this interpretation we could ask, what would happen if we aimed to represent a larger time span of inquiry, throughout which scientists gradually improve their methodologies – which is precisely what our model is designed to answer.

to each theory, representing the probability of gaining corroborating evidence for it, given the current state of one's inquiry. For the sake of simplicity we will refer to the two theories as the 'true' and the 'false' one (we comment on other possible interpretations below in this section). Each theory is assigned an *actual probability of success* (APS), which represents the ideal probability of its success: APS(True theory)=1 and APS(False theory)=0. This means that if a scientist is pursuing the true theory, inquiry should lead her to more and more successful applications of the given methodology. In contrast, if she is pursuing the false one, her applications of it should become less and less successful.[9]

Our assumptions are motivated by cases such as the above mentioned research on PUD, where successful applications of the bacterial hypothesis increased, while those of the acidity hypothesis decreased. Another example fitting this scenario is the continental drift debate, where the research program built on the basis of Wegener's hypothesis of continental drift gradually gained success, while its rivals (the so-called contractionism and permanentism) gradually became weaker (see Šešelja and Weber, 2012). Another option would be to assume that both (or more) rivaling hypotheses gain success at different rates. For instance, the objectively worse hypothesis may in the beginning make greater improvements than its rival, though it is eventually surpassed by the latter. Yet another option would be to assume that only the objectively better theory gains success, while the objectively worse one remains the same (i.e. to represent only the better theory in terms of a restless bandit). This latter case might especially be applicable to the modeling of rivaling *methods*, for instance in technological research, where the worse one may stagnate while its rival shows improvement. While in Zollman's model theories, hypotheses and methodologies are treated interchangeably, this point raises the question whether dynamic epistemic success has to be represented differently in each of these cases. We will thus restrict our interpretation to theories in the sense of research programs which can gain or loose epistemic success (as it was in the above mentioned cases of PUD and the continental drift debate).We will say a few more words on alternative forms of dynamics in Section 7.

We implement the above ideas as follows.[10] At the start of the simulation each theory is assigned an initial value for the current probability of success (CPS). Like Zollman, we will take those to be 0.5 for the true theory and 0.499 for the false one, for all agents. Every x rounds (e.g. every 100 rounds), an agent's CPS(T) (where T stands for a theory in general) will slightly increase in case of the true theory, and slightly decrease in case of the false one. More precisely, the CPS assigned by an agent to the given theory after an update, expressed in terms of the CPS of the same theory before the update will be as

[9]Our results would also hold under less idealized values for APS, such as, for example, 0.3 and 0.9 respectively, as long as the ratio by which CPS of each theory changes is adjusted accordingly. Moreover, runs tend to end long before each agent's CPS reaches the according APS.

[10]The model presented in this paper is programmed in NetLogo (Wilensky, 1999). The open-source code is available on GitHub: `https://github.com/daimpi/SocNetABM/tree/RobIdeal`. Our code also includes Zollman's (2010) ABM, as a nested variant.

follows:

$$\mathrm{CPS}_{\mathrm{after}}(T) = \mathrm{CPS}_{\mathrm{before}}(T) + f(d) \tag{1}$$

where $f(d) = \frac{d}{1000}$ and $d = \mathrm{APS}(T) - \mathrm{CPS}_{\mathrm{before}}(T)$.

Let's clarify this a bit. Every x rounds, the bandit of each agent will improve its CPS towards its APS. Such an improvement is a function of the distance d between the CPS and the APS of the given theory (before the update). This means that the CPS can be improved towards APS in greater steps in the beginning of the inquiry; the closer a scientist gets to the theory's true mean, the smaller improvements she makes.[11] How often such global improvements in CPS occur is a parameter of the model.

It is important to mention though that even at the beginning of the inquiry the improvements a scientist makes are very small: for example, assuming that the improvements occur every 100 rounds, the first time an agent on the true theory receives an improvement, her CPS will change in the following way: $\mathrm{CPS}(\text{True theory}) = 0.5 + 0.0005 = 0.5005$. Similarly, for an agent pursuing the false theory $\mathrm{CPS}(\text{False theory}) = 0.499 - 0.000499 = 0.498501$.

Just like in Zollman's model, we represent an individual scientist's beliefs concerning the given theory in terms of beta distributions where the priors are determined by each agent drawing an α and β for each theory from a continuous uniform distribution over the interval $(0, 4]$. We will call the mean of the beta distribution a scientist's *subjective probability of success* (SPS) for that theory.[12]

An important consequence of our implementation of dynamic epistemic success is that due to methodological improvements, all our scientists will *eventually* discover which theory is objectively better (similarly to Zollman's agents who start with extreme priors, see Footnote 5). Hence, even if scientists prematurely abandon the better of the two hypotheses, they will eventually get back to it. This means that the question of the efficiency of inquiry shifts from 'How often are scientists successful?' to 'How long does it take them to get it right?'. Note that this doesn't mean that consensus on the false theory is not captured by the model: the scientific community may still abandon the true theory for a large time frame of the given run. As a result, such a community will be much slower in switching back and converging on the true theory. Thus, instead of measuring efficiency in terms of the percentage of successful runs (as Zollman does), we will instead measure the efficiency in terms of time that scientists need in order to converge on the true theory.[13]

[11] We have implemented the change in CPS of each theory in terms of diminishing marginal returns in order to represent the idea that one's methodological improvements result in greater successes of finding corroborating evidence in the beginning than at a later point in inquiry. While this is clearly an idealization, replacing it with a steady change in CPS is not likely to have any major effects on our results since most runs end long before CPS(T) approaches the value of APS(T).

[12] The mean of the beta distribution is given by the ratio of successes to pulls: $\mathrm{SPS}(T) = \frac{\alpha(T)}{\alpha(T)+\beta(T)} = \frac{\mathrm{successes}(T)}{\mathrm{successes}(T)+\mathrm{fails}(T)} = \frac{\mathrm{successes}(T)}{\mathrm{pulls}(T)}$.

[13] While we could have introduced an arbitrary cutoff point which would allow for distinguishing successful from unsuccessful runs (as Zollman does in the case of extreme priors), such a cutoff would be unmotivated as long as we don't map the time in the model to the real

A consequence of our implementation is a simulation that coheres with a (conditional) self-corrective thesis of science, according to which scientific methods are self-corrective and truth-conducive.[14] Nevertheless, we do not require a realist interpretation of this thesis (and hence we also do not require a realist notion of truth conduciveness). The model could equally be understood as representing two rivaling theories within a given historical framework, where one is more empirically successful than another, in view of the methodological standards adopted at the time. Under such interpretation, what we do assume is that if scientists have an access to two theories, one of which is from an objective point of view clearly more successful than the other, then even if they initially dismiss the former, they will eventually discover they have made a mistake and return to it. At the same time, we do not exclude a realist interpretation of the self-corrective thesis. It is worth mentioning though that our model (just like Zollman's) does not represent a situation in which new theories that are superior to their predecessors are discovered throughout the inquiry.

A consequence of replacing a static epistemic success with a dynamic one (and keeping all the other assumptions unchanged) is that we can reproduce Zollman's result concerning the order of the cycle and the complete graph in terms of their efficiency. Figure 2 represents the results for the cycle and complete graph, under the assumption that scientists receive the global improvement in CPS every 100 rounds.[15] We will call this scenario our *basic setup*.[16]

Since the assumption of a dynamic epistemic success concerns the way in which we measure efficiency, we will employ it as the basis for introducing further adjustments to Zollman's model. There are two reasons for this. On the one hand, keeping two approaches to the representation of efficiency would make our paper significantly longer. Second, in contrast to the other changes in assumptions that will be introduced in the next three sections, dynamic epistemic success isn't a specific feature whose absence would represent a realistic context of inquiry. As we have argued at the beginning of this section, if scientists are considered successful in their inquiry only if they converge on one of the rivaling theories, then it seems highly questionable to assume that these theories start

time (see also Section 7). Instead, our approach could be roughly understood as subsuming Zollman's original runs and examining what happens if agents begin to gradually improve their methodologies: on the one hand, agents who would be successful in Zollman's model will complete the run relatively quickly in our model (they may even finish before dynamic epistemic success kicks in for the first time). On the other hand, agents who would be unsuccessful in Zollman's model, also tend to be generally slower in our model (e.g. if they have a consensus on the wrong theory, it may take them many rounds to converge on the better theory).

[14] The self-corrected thesis has been most prominently advanced by Peirce, Popper and Reichenbach, and more recently by Mayo, 2005; for an early criticism see Laudan, 1981.

[15] We let each simulation run for up to 100,000 rounds, for populations consisting of 4 to 11 scientists, and we recorded the point when agents converge on the true theory (by convergence on a theory we mean that all agents end up on that theory, without switching back to the rival). The same holds for the plots presented in the remainder of the paper. Each of them shows results based on 10,000 runs for each data point.

[16] The order of the cycle and the complete graph remains the same if we allow for the global improvement in CPS to occur every 10 rounds, except that both networks converge much faster. We will get back to the issue of time and its representation in this model in Section 7.

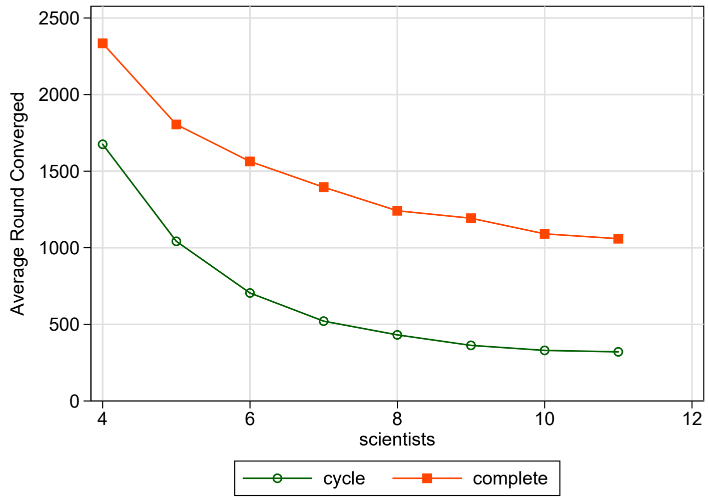

Figure 2: The average time agents need for successful convergence on the true theory under the assumption that global improvement in CPS happens every 100 rounds.

with the OPS of 0.5 and 0.499 respectively, and that they do not change these values over the course of research. Hence, examining what would happen if we discharged this assumption, while varying others, doesn't seem very interesting.

In the next three sections we will thus examine the impact of some additional assumptions, which represent the presence of specific features of scientific inquiry, on the results of our basic setup.

4 Critical interaction

4.1 Introducing critique

Interaction in Zollman's model includes critique or critical interaction neither explicitly, nor implicitly. In particular, the model doesn't include the assumption that criticism can be epistemically beneficial in the sense of helping scientists to reveal their mistakes. However, looking at actual scientific inquiry, when scientists exchange evidence that supports one of the rivaling theories and undermines another, we expect their opponents to engage in critical assessment of their own results and methodology, reexamining whether they have perhaps made a mistake. As a result, critical interaction plays an important role in disclosing errors that regularly appear in scientific research, namely false positives and false negatives. More generally speaking, critique tends to be truth conducive since it allows for false beliefs to be exposed as such (Betz, 2012).[17] This means that criticism is not only normatively relevant component of scientific interaction, but it is also descriptively adequate for many contexts of scientific inquiry. This especially concerns times of scientific controversies when inquiry can be difficult, and when critical exchange occurs via peer review procedure, published articles, or discussions during scientific conferences. Hence, specifying whether interaction represented in the model involves a critical component or not is important for determining the context to which the results of the model are supposed to apply.

4.2 Implementation and results

We will introduce a critical component to the representation of scientific interaction by assuming that:

a) criticism is truth conducive;

b) it occurs between proponents of rivaling theories.

These two assumptions cohere with the Millian view on rational scientific inquiry, according to which critical interaction with experts whose views conflict with one's own is essential for the justification of our own beliefs (see e.g. Moffett, 2007).

We assume that criticism is triggered every time an agent pursuing T_x receives information from an agent pursuing T_y, such that T_y turns out to be better than she has expected. More precisely, scientist S_1 working on T_x is affected by criticism whenever the success rate of the rivaling theory (reported by scientist S_2 working on T_y) from the pulls in the most recent round is higher than the value S_1 has had in her memory, i.e. in case:

$$S_1 : \mathrm{SPS}_{\mathrm{before}}(T_y) < \mathrm{SPS}_{\mathrm{after}}(T_y).$$

In other words, the receiver of information is affected by criticism every time she corrects her belief about the rivaling theory in a positive direction.

Similarly to the implementation of dynamic epistemic success, we assume that critical interaction allows agents to slightly improve the CPS of their current theory towards the APS of that theory. The CPS(T_x), assigned by S_1 after criticism has been triggered, is calculated according to Equation 1. For example, take an agent who is pursuing the true theory and currently has the following beliefs: SPS(True theory) = 0.51 and SPS(False theory) = 0.49 Imagine now that she receives information from an agent pursuing the false theory showing that on her pulls in the most recent round, the false theory has manifested the success ratio of 0.5. This could be understood as her opponent saying that he has found a new explanation of a phenomenon which our scientist previously

[17] The view that criticism is epistemically beneficial goes back to John Stuart Mill, and was later endorsed by Karl Popper and the school of critical rationalism. For an ABM of science implementing this view in a different way than we do here see Chavalarias, 2017.

didn't think could be explained by the other theory. We interpret this as a situation in which a scientist discovers that there has been an error in her overall assessment of the given domain, which makes her reexamine and improve her methodology.[18] As a result, her next pull will be made from a slightly improved bandit: if her CPS(True theory) was e.g. 0.5 before she has engaged in critical interaction, it will be upgraded to $0.5 + \frac{1-0.5}{1000} = 0.5005$. And conversely, if she were on the false theory, receiving information from an opponent on the true theory, her CPS would slightly decrease (in the same way as explicated above with respect to the implementation of dynamic epistemic success).[19]

Adding this assumption to our basic setup produces the results in Figure 3. The cycle still outperforms the complete graph. Both networks perform slightly better than in the absence of critical interaction, which is of course not surprising since we have assumed that critical interaction is epistemically beneficial.

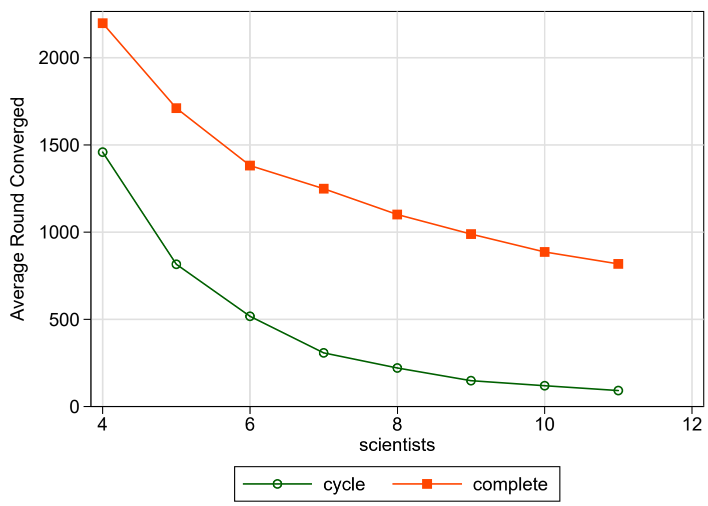

Figure 3: The average time for successful convergence under the assumption that global improvement in CPS happens every 100 rounds, and that agents interact critically.

[18]Given the simple character of the model, it is impossible to represent all the aspects of critical exchange or of theory assessment. As a result, even though we are adding some new assumptions to the model, we still have to employ various idealizations. For example, since agents can pursue only one theory, we assume that in this way they are also able to point out possible problems in their rivals' theory. For a more complex representation of critical interaction see Borg et al., 2017b, 2018.

[19]It is important to note though that the dynamics underlying critical interaction may very well vary from one scientific domain to another, and that the way we implement it here doesn't necessarily hold for all contexts of scientific inquiry.

5 Inertia of inquiry

5.1 Introducing rational inertia

A peculiar feature of Zollman's agents is that they are easily swayed by new evidence. That is, they easily give up on a theory they've been pursuing if, for example, the initial evidence suggests that the rival is superior. Nevertheless, in the context of actual scientific inquiry scientists tend to retain their current theory at least for a certain time period. More precisely, they will stick with it unless they are convinced that it can no longer be saved from the defeating evidence. This is not necessarily irrational behavior: if a scientist knows her current evidence is insufficient to determine whether the theory could eventually be accepted – as it may easily occur in times of difficult inquiry– it would be irrational to abandon it before attempting its further development, and rational to stick to it for a while longer (see Kelp and Douven, 2012). In addition, changing one's inquiry usually includes a number of costs (e.g. acquiring additional knowledge, new equipment, etc.), which is another motivation for such inertia.[20] Hence, we can assume that in some contexts of (difficult) inquiry scientists have a rational inertia towards their current theory, and they take some time to examine whether the theory can be improved before deciding to abandon it.

Note that such inertia should not be confused with Zollman's notion of extreme priors with which scientists may initiate their inquiry. While the latter notion does exhibit a type of inertia, this holds only for the initial phase of research. As soon as the prior value is overcome in view of updates, agents are no more inert towards their theories. For example, if Zollman's agent equipped with extreme priors starts with $\alpha = \beta = 3,000$ it may take many rounds before the probability of the pursued theories is altered to the extent that she changes her current theory. However, after this point, updates will continue in the same way as in the case of smaller priors and no further inertia will be displayed. In contrast, our notion concerns inertia for any new instance of inquiry, that is, a point where an agent starts pursuing the other theory.

5.2 Implementation and results

We implement this feature in terms of a *jump threshold*: agents 'jump' to the rivaling theory only after the latter has turned out to be better than their current theory for a certain number of rounds according to their beliefs, where the specific number of rounds is a parameter of the model.

Figure 4 shows the results of adding a jump threshold of 10 rounds to the basic setup of our model. Both the cycle and the complete graph perform similarly and although scientists take longer to switch theories, the average

[20]While our intention is to keep the current model simple, in a more complex model such inertia could result from the process of optimization in which not fully myopic agents take into account the costs of changing their inquiry.

time they need to converge is less than in the basic setup. Moreover, we observe that the complete graph surpasses the cycle in case of larger groups.

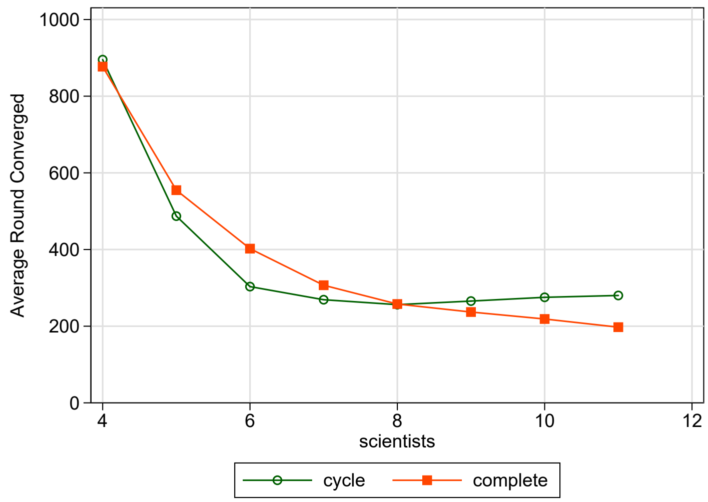

Figure 4: The average time for successful convergence under the assumption that global improvement in CPS happens every 100 rounds, and that there is a jump threshold of 10.

Figure 5 shows what happens once we add critical interaction to this scenario. As expected, both cycle and complete graph are now more efficient. It is important to notice that even though critical interaction is per design epistemically beneficial, its impact on the ordering of the networks in terms of their performance isn't straightforward. On the one hand, agents connected in a complete graph will be advantaged since they more often interact, and hence they may more often *critically* interact. On the other hand, this advantage is counteracted by the fact that they will faster end up pursuing only one of the theories, in which case there can be no more critical interaction among them.[21]

6 Threshold within which theories are equally promising

6.1 An inquiry that is even more difficult

As mentioned above, Zollman's model is best described as representing the situation of a difficult inquiry (as Rosenstock, O'Connor, and Bruner, 2017 suggest) due to the fact that the rivaling theories are very similar in terms of their OPS.

[21]In fact, data analysis shows that agents in the cycle e.g. in Figure 7 (see below) on average interact critically more often than agents in the complete graph.

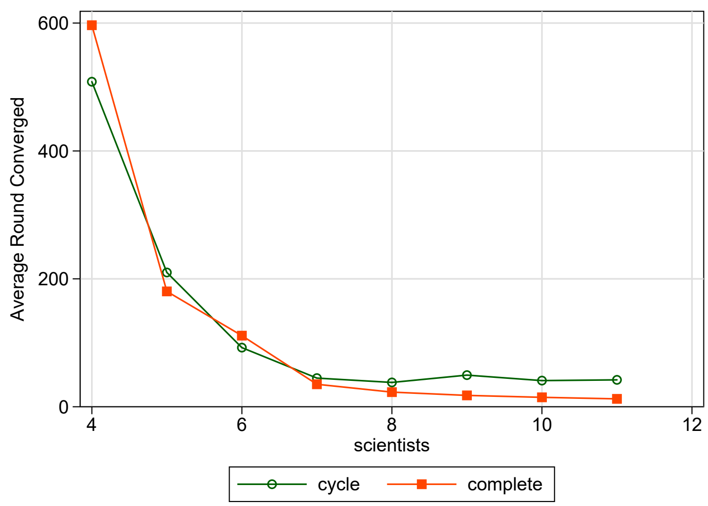

Figure 5: The average time for successful convergence with global improvement in CPS every 100 rounds, jump threshold of 10, and the assumption that agents critically interact.

A specific feature of Zollman's scientists is that they are able to distinguish between the success of these theories no matter how similar they are. So even when theories are very close to one another in terms of SPS that scientists assign to them, they can perfectly determine which one is better. Nevertheless, there are at least two ways in which this assumption can be challenged.

First, one way to think of difficult inquiry is as follows: if scientists are confronted with theories that are very similar in terms of epistemic support, it may be impossible for them to determine which one is better. For instance, they may face difficulty of aggregating the relevant criteria of evaluation (such as explanatory power, consistency, fruitfulness, etc.), which may impede their ability to make an overall preference for only one theory. As a result, it may be impossible for them to say which rival is *more* worthy of pursuit than another, at least as long as they appear to be very similar.

Second, as we have already mentioned in previous sections, the process of inquiry might reveal new evidence, in light of which one's former conclusions could turn out to be wrong. In view of this, rational scientists conducting a difficult inquiry will employ a dose of caution when evaluating rivaling theories, precisely to avoid prematurely discarding an objectively better theory. This means that they will reject a theory not simply after they have seen it perform worse multiple number of times (as discussed in Section 5), but only after its rival has become sufficiently superior to it.

These two points motivate the assumption that scientists employ a threshold which a theory has to surpass in order to count superior to its rival. Let's see what happens when we add this assumption to our model. We will then examine the effects of combining this assumption with those we have introduced in the previous sections.

6.2 Implementation and results

In order to account for the above presented scenario, we introduce a threshold value to the theory assessment: the rival theory counts as better only if it surpasses one's own theory by the margin of 0.1 (in terms of SPS that an agent assigns to the theories). We will call this parameter *theory threshold*.

Figure 6 shows what happens when we add this assumption to our basic setup. Even though the complete graph is better, both networks need a long time to converge on the true hypothesis. Note that this result suggests that the interpretation by Rosenstock, O'Connor, and Bruner, 2017, according to which Zollman's ordering of the networks holds for the context of difficult inquiry, may not apply to all contexts of difficult inquiry. Our simulations suggest that if the inquiry is difficult in the sense that it includes the aggregation problem (or in view of our second motivation: if scientists require that a theory is sufficiently inferior to its rival before they reject it), the complete network outperforms the cycle.

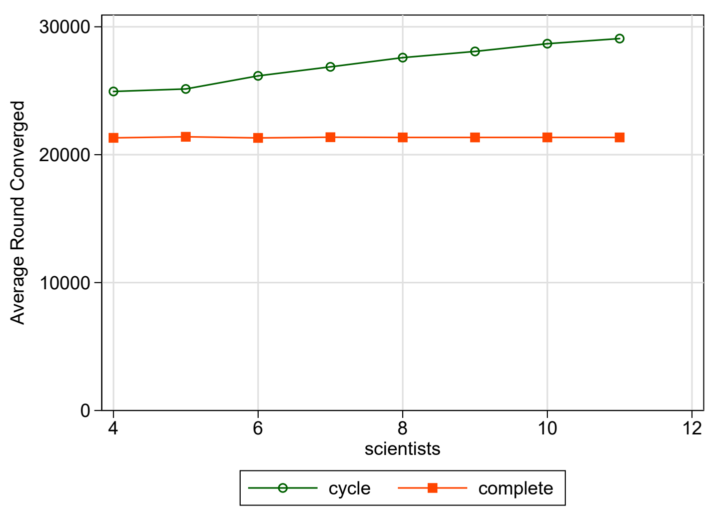

Figure 6: The average time for successful convergence with global improvement in CPS every 100 rounds, and a theory threshold of 0.1.

If we now add critical interaction to this scenario, both networks become much faster, though the complete graph still outperforms the cycle. The results of these simulations are shown in Figure 7.

Finally, adding inertia to agents doesn't change the situation much. Figure

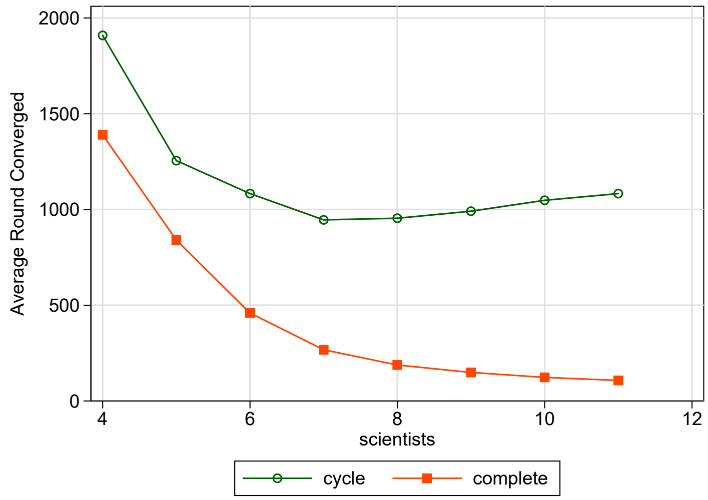

Figure 7: The average time for successful convergence with global improvement in CPS every 100 rounds, a theory threshold of 0.1, and the assumption that agents critically interact.

8 shows the results of adding a jump threshold of 10 to the previous setup. It is interesting to notice that while adding inertia to the basic setup had a strong impact on the results (see Figures 2 and 4) it now hardly has any. This is due to the fact that the benefits of inertia – preventing a premature rejection of a theory – are now covered by the theory threshold, which allows for theoretical diversity to be retained for a large portion of a run.

7 Discussion

The above results indicate that the question whether and to which extent the degree of connectedness impacts the efficiency of scientific inquiry is a highly contextual issue. Figure 9 summarizes the performance of the cycle and complete graph under different conditions, presented in the previous sections. On the one hand, our results show that the complete graph may very well outperform the cycle even in the context of difficult inquiry (see Figure 10), and even if we don't assume that scientists intentionally keep on pursuing the worse theory (as examined by Kummerfeld and Zollman, 2016).[22] This is because

[22]Even though our assumption of rational inertia allows for the preservation of diversity of pursued theories, its interpretation is quite different from the assumption employed by Kummerfeld and Zollman, 2016, where the diversity is preserved by agents always pursuing the inferior theory with a certain probability. In contrast to the latter case where each agent pursues two theories, in the former case an agent stays on her current theory until she is convinced that there are good reasons to abandon it.

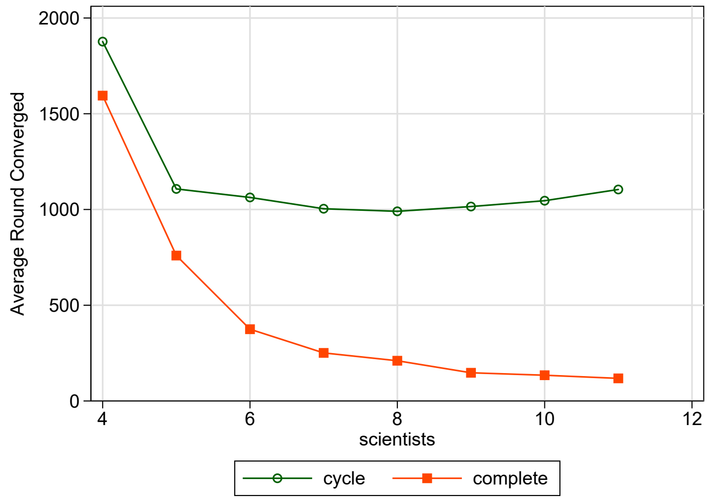

Figure 8: The average time for successful convergence with global improvement in CPS every 100 rounds, a theory threshold of 0.1, a jump threshold of 10, and the assumption that agents critically interact.

some additional factors may positively impact theoretical diversity. These are, on the one hand, rational inertia (which can also be understood as a cautious methodological approach or as an optimization in view of costs and benefits of changing inquiry), and on the other hand, the interval within which theories are considered equally promising. On the other hand, the presence of critical interaction and rational inertia may in some contexts increase the efficiency to a greater extent than does increasing or decreasing the degree of connectedness (see Figure 11).

Looking at the runs where decreasing network density seems to provide an advantage, our results confirm a conjecture by Rosenstock, O'Connor, and Bruner, 2017 that "there are better solutions, in these cases, to the problem" (p. 251). Figure 11 shows that in those cases having scientists with some inertia yields a much higher rate of improvement than trying to limit the flow of information among them.

Altogether, this shows not only that the results of Zollman's (2010) may not hold for all situations of difficult inquiry, but also that the issue of connectedness and its impact on the efficiency of knowledge acquisition is much more context dependent (even in the case of difficult inquiry) than this might have seemed in view of Zollman's results. Consequently, any application of these modeling results to concrete cases of scientific inquiry should aim at carefully establishing that the target domain of the given case study corresponds to the particular context represented by the model.

However, looking at the way Zollman's results have been used in the litera-

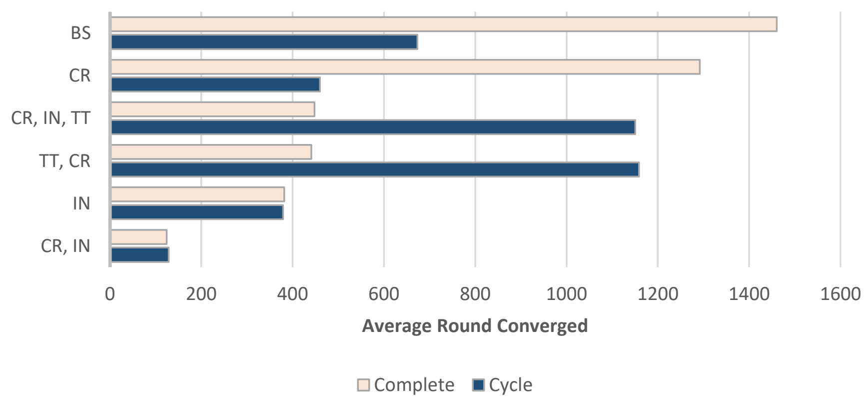

Figure 9: Time for successful convergence averaged over all population sizes presented in the previous sections. BS: basic setup (Section 3); CR: critical interaction (Section 4); IN: inertia (Section 5). The standalone theory threshold (TT) treatment (Section 6) has been omitted since its inclusion would require distortive scaling due to the very high time requirements in those runs.

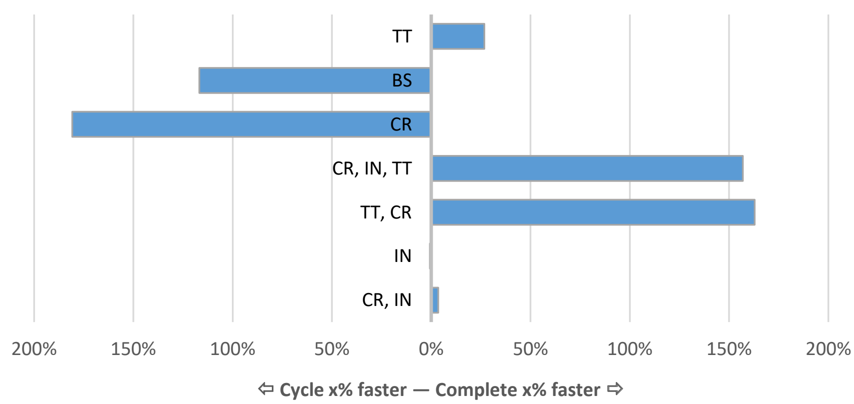

Figure 10: The influence of network structure on the efficiency of inquiry expressed in relative terms for each of the presented treatments. 'x% faster' refers to the given network structure needing x% less rounds than the alternative network structure for successful convergence on average. For the meaning of shortcuts used for the treatments see the caption below Fig. 9.

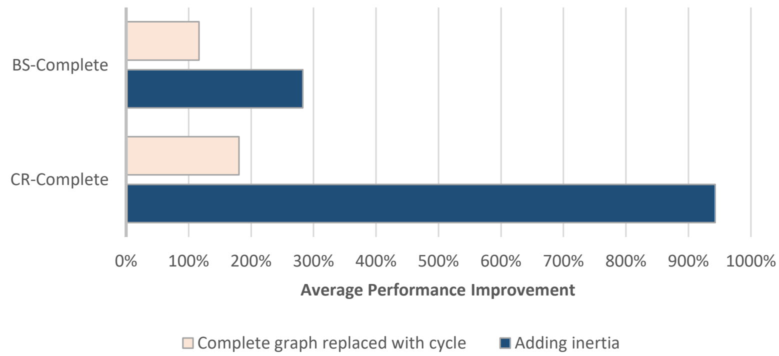

Figure 11: The impact of changing the network structure vs. adding inertia on the efficiency of two scenarios in which decreasing the degree of connectedness is beneficial (see Fig. 10): the complete network in the basic setup (BS-Complete), resp. the complete network in the setup with critical interaction (CR-Complete). An improvement of x% means that changing the given factor results in runs that on average need x% less rounds for successful convergence than BS-Complete, resp. CR-Complete.

ture, we find the deployment of cautionary measures lacking. As mentioned in Section 1, philosophers of science and epistemologists have adopted the findings of his model (e.g. Goldman and Blanchard, 2016; Strevens, 2010; Wray, 2011). In addition, Zollman himself (in his (2010)) has used the results of his model to explain a concrete historical case study —the research on PUD— without showing that the given case falls under the context represented by the model.

Our findings suggest that Zollman's results hold only if it is neither the case that scientists make their decisions as for what to pursue cautiously, nor is it the case that they are confronted with the aggregation problem when making theory choice. As soon as one of these two conditions is satisfied, our ABM indicates that the cycle ceases to be superior to the complete graph, or it even becomes inferior to it.[23] Since each of these assumptions concerns a method conducive to the preservation of diversity, this conclusion is not very surprising given Zollman's own findings. Nevertheless, the novelty of our findings lies in the specification of certain conditions under which we may expect different degrees of connectedness to have (or to lack) a significant impact on the efficiency of inquiry in contrast to other related factors.

Having said that, it is important to add that even these results should be taken with a grain of salt when it comes to any real-world applications. The

[23]While our results still require a proper sensitivity analysis, which may point to additional restrictions of the application domain, it is worth mentioning that they are supported by similar findings obtained by Borg et al., 2017b, 2018, whose models also employ the *jump threshold* and the *theory threshold* as parameters of the models that can be interpreted along the similar lines as we have done in this paper.

primary challenge here is not just making sure that results are robust with respect to idealizing assumptions of the model, but also mapping the parameters used in the model to empirical information. For instance, the meaning of time steps in the model (relative to the pulls made by scientists, improvements in terms of their CPS, the time they need to converge on the right theory, etc.) requires further analysis if we are to make claims that are of relevance for actual scientific inquiry.[24] Similarly, different cases of scientific inquiry may require different representations of dynamic epistemic success, and our model may be applicable only to some of them.

Concerning the latter issue, even though in this paper we have presented results based on a specific form of dynamics underlying epistemic success of theories, we can make a few comments on some alternative forms. For example, if we assume that the epistemic success of the worse theory doesn't decrease but remains static or advances at a slower pace than the epistemic success of the better theory, and if there is no inertia or theory threshold, we observe that agents connected in the complete graph are more likely to converge on the wrong theory than those connected in the cycle.[25] Hence, in this scenario agents may very well remain stuck on the worse theory for good. Adding either inertia or theory threshold helps in avoiding the wrong convergence, but may also extend the time of the run. Thus, whether we observe Zollman's main result or not will primarily depend on the relative impact of inertia and theory threshold even if we replace the underlying dynamics of the notion of epistemic success.[26] This is another reason why mapping the parameter space in the model to the real world is essential in obtaining information that is relevant for actual scientific inquiry.

A more general take-home message of our findings is that simplicity of models, while possibly beneficial for their explanatory features (Batterman and Rice, 2014; Reutlinger, Hangleiter, and Hartmann, 2016), may nevertheless easily lead to a blind spot in determining the exact target phenomenon that the model represents. It seems then that simplicity comes at a price: a requirement to run an extensive robustness analysis in order to establish the link between the model and its real-world target phenomenon. Without examining whether changes in idealizing assumptions of the model lead to changes in its results it might be impossible to say whether the model is adequately minimal with respect to the given target phenomenon.

[24] Some preliminary work in this direction has been conducted in our Frey and Šešelja, 2018, where we propose a way in which the above mentioned case study on PUD could be used to calibrate the model presented in the current article.

[25] An interested reader can easily examine each of these scenarios by means of our model (see Footnote 10 for the link to the code).

[26] Similarly, we expect that introducing exploratory agents from (Kummerfeld and Zollman, 2016) would be another method of diversity preservation, which would help in avoiding the wrong convergence in the above scenarios.

8 Conclusion

The title of our paper is inspired by Muldoon and Weisberg, 2011 who argued that Kitcher's (1993) and Streven's (2003) model of the division of cognitive labor is not robust under changes of some of its idealizing assumptions.[27] We have presented a similar kind of investigation in the domain of ABMs of scientific interaction. To this end, we have developed an ABM, based on Zollman's (2010) model aimed at examining the impact of different assumptions about scientific inquiry on the results obtained by Zollman's model.

Since the context of difficult inquiry has been suggested as an adequate target phenomenon represented by Zollman's model (Rosenstock, O'Connor, and Bruner, 2017), we have focused on four idealizing assumptions, all of which may be relevant in the context of difficult inquiry. First, we have replaced the assumption that the epistemic success of given theories is static with the assumption that it is dynamic. Second, instead of assuming that all scientific interaction is epistemically equal, we have introduced the assumption that sometimes scientists criticize each other, where such interaction is epistemically beneficial. Third, instead of representing agents as easily swayed by new evidence, we have added the possibility that they have a rational inertia towards their current theories. Finally, rather than assuming that one always has a linear preference order over the rivaling theories, we have introduced the assumption that there is an interval within which rivaling theories count as equally good.

Our results suggest that whether and to which extent the degree of connectedness impacts the performance of a scientific community is an issue contingent on a number of factors that may be present in a given inquiry. In view of this we have also specified the contexts of inquiry in which Zollman's results are more likely to hold than in others.

The upshot of our investigation was showing the significance of robustness analysis under changes in idealizing assumptions, which has been neglected in the literature on ABMs of scientific interaction. To this end, we have paradigmatically focused on Zollman's model as the most prominent ABM of scientific interaction. Examining to which extent our changes in assumptions may affect results of other highly idealized ABMs of scientific interaction (such as those by Grim et al., 2013; Holman and Bruner, 2015, etc.) remains a task for future research.

Acknowledgements We are indebted to Christian Straßer for a number of valuable comments on the previous draft of this paper. We are also grateful to Cailin O'Connor and the two anonymous reviewers for fruitful comments and

[27] An interesting detail is that Weisberg and Muldoon's (W&M) epistemic landscape ABM (Weisberg and Muldoon, 2009), announced in their (2011) paper, turns out not to be robust under changes in idealizing assumptions either, as shown by Alexander, Himmelreich, and Thompson, 2015. See also Thoma, 2015 and Pöyhönen, 2017 for additional criticism of W&M's model.

suggestions, and to AnneMarie Borg for the latex code of a fancy illustration of a cycle, wheel and a complete graph.

References

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

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

Batterman, Robert W and Collin C Rice (2014). "Minimal model explanations". In: *Philosophy of Science* 81.3, pp. 349–376.

Betz, Gregor (2012). *Debate dynamics: How controversy improves our beliefs.* Vol. 357. Springer Science & Business Media.

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

— (2017b). "Examining Network Effects in an Argumentative Agent-Based Model of Scientific Inquiry". In: *Logic, Rationality, and Interaction: 6th International Workshop, LORI 2017, Sapporo, Japan, September 11-14, 2017, Proceedings.* Ed. by Alexandru Baltag, Jeremy Seligman, and Tomoyuki Yamada. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 391–406. ISBN: 978-3-662-55665-8. DOI: 10.1007/978-3-662-55665-8_27. URL: https://doi.org/10.1007/978-3-662-55665-8_27.

— (2018). "Epistemic effects of scientific interaction: approaching the question with an argumentative agent-based model". In: *Historical Social Research.* accepted for publication.

Bruch, Elizabeth and Jon Atwell (2015). "Agent-based models in empirical social research". In: *Sociological methods & research* 44.2, pp. 186–221.

Chavalarias, David (2017). "Whats wrong with Science?" In: *Scientometrics* 110.1, pp. 481–503.

Frey, Daniel and Dunja Šešelja (2018). "What is the Epistemic Function of Highly Idealized Agent-Based Models of Scientific Inquiry?" In: *Philosophy of the Social Sciences* accepted for publication.

Goldman, Alvin and Thomas Blanchard (2016). "Social Epistemology". In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Summer 2016. Stanford University.

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

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

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

Hosking, S.W., S.C.S. Chung, M.Y. Yung, A.K.C. Li, J.J.Y. Sung, T.K.W. Ling, and A.F.B. Cheng (1994). "Duodenal ulcer healing by eradication of Helicobacter pylori without anti-acid treatment: randomised controlled trial". In: *The Lancet* 343.8896. Originally published as Volume 1, Issue 8896, pp. 508 –510. ISSN: 0140-6736. DOI: http://dx.doi.org/10.1016/S0140-6736(94)91460-5. URL: http://www.sciencedirect.com/science/article/pii/S0140673694914605.

Kelp, C. and I. Douven (2012). "Sustaining a Rational Disagreement". In: *EPSA Philosophy of Science: Amsterdam 2009*. Ed. by Stephan Hartmann Henk W. de Regt and Samir Okasha. Springer, pp. 101–110.

Kitcher, Philip (1993). *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford University Press.

Kummerfeld, Erich and Kevin JS Zollman (2016). "Conservatism and the scientific state of nature". In: *The British Journal for the Philosophy of Science* 67.4, pp. 1057–1076.

Laudan, Larry (1981). "Peirce and the trivialization of the self-corrective thesis". In: *Science and Hypothesis*. Springer, pp. 226–251.

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

Mayo, Deborah G (2005). "Peircean induction and the error-correcting thesis". In: *Transactions of the Charles S. Peirce Society* 41.2, pp. 299–319.

Moayyedi, Paul, Shelly Soo, Jonathan Deeks, David Forman, James Mason, Michael Innes, and Brendan Delaney (2000). "Systematic review and economic evaluation of Helicobacter pylori eradication treatment for non-ulcer dyspepsia". In: *BMJ* 321.7262, pp. 659–664. ISSN: 0959-8138. DOI: 10.1136/bmj.321.7262.659. eprint: http://www.bmj.com/content/321/7262/659.full.pdf. URL: http://www.bmj.com/content/321/7262/659.

Moffett, Marc (2007). "Reasonable disagreement and rational group inquiry". In: *Episteme: A Journal of Social Epistemology* 4.3, pp. 352–367.

Muldoon, Ryan and Michael Weisberg (2011). "Robustness and idealization in models of cognitive labor". In: *Synthese* 183.2, pp. 161–174.

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

Reutlinger, Alexander, Dominik Hangleiter, and Stephan Hartmann (2016). "Understanding (with) Toy Models". In: *The British Journal for the Philosophy of Science*.

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

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

Strevens, Michael (2003). “The role of the priority rule in science”. In: *The Journal of philosophy* 100.2, pp. 55–79.

— (2010). “Reconsidering Authority: Scientific Expertise, Bounded Rationality, and Epistemic Backtracking”. In: *Oxford studies in epistemology* 3, p. 294.

— (2013). “No understanding without explanation”. In: *Studies in history and philosophy of science Part A* 44.3, pp. 510–515.

Thoma, Johanna (2015). “The Epistemic Division of Labor Revisited”. In: *Philosophy of Science* 82.3, pp. 454–472.

Šešelja, Dunja and Erik Weber (2012). “Rationality and Irrationality in the History of Continental Drift: Was the Hypothesis of Continental Drift Worthy of Pursuit?” In: *Studies in History and Philosophy of Science* 43, pp. 147–159.

Weisberg, Michael (2007). “Three kinds of idealization”. In: *The journal of Philosophy* 104.12, pp. 639–659.

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

Wilensky, Uri (1999). “NetLogo.(http://ccl.northwestern.edu/netlogo/)”. In: *Center for Connected Learning and Computer Based Modeling, Northwestern University*.

Wray, K Brad (2011). *Kuhn’s evolutionary social epistemology*. Cambridge University Press.

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

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