Some Lessons from Simulations of Scientific Disagreements

Published in Synthese, DOI 10.1007/s11229-019-02182-0

Dunja Šešelja

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

Abstract

This paper examines lessons obtained by means of simulations in the form of agent-based models (ABMs) about the norms that are to guide disagreeing scientists. I focus on two types of epistemic and methodological norms: (i) norms that guide one's attitude towards one's own theory, and (ii) norms that guide one's attitude towards the opponent's theory. Concerning (i) I look into ABMs that have been designed to examine the context of peer disagreement. Here I challenge the conclusion that the given ABMs provide a support for the so-called Steadfast Norm, according to which one is epistemically justified in remaining steadfast in their beliefs in face of disagreeing peers. I argue that the proposed models at best provide evidence for a weaker norm, which concerns methodological steadfastness. Concerning (ii) I look into ABMs aimed at examining epistemic effects of scientific interaction. Here I argue that the models provide diverging suggestions and that the link between each ABM and the type of represented inquiry is still missing. Moreover, I examine alternative strategies of arguing in favor of the benefits of scientific interaction, relevant for contemporary discussions on scientific pluralism.

Keyowords: agent-based models, scientific disagreement, rational endorsement, scientific interaction, epistemic toleration, scientific pluralism.

1 Introduction

Scientific disagreements are usually considered one of the key conditions of scientific progress (Kuhn, 1977; Longino, 2002; Solomon, 2006). Nevertheless, as many have noted, an inadequate response to disagreements can lead to premature rejections of fruitful inquiries (Chang, 2012; Šešelja and Weber, 2012), to fragmentation of scientific domains (Barnes and Sheppard, 2010; Rolin, 2011), and hence to consequences that are counterproductive for the progress of science. In light of this, determining epistemic and methodological norms that are to guide scientists in their inquiry in face of disagreements is an important philosophical challenge. Moreover, it is a challenge that may not be easy to address from a philosophical armchair: estimating how different norms guiding individual scientists impact the efficiency of inquiry as a collective enterprise is a complex issue, depending on a variety of factors and their mutual interaction. While historical case studies can help in pointing out examples of disagreements and their possible beneficial or harmful effects, they are a weak basis for formulating general norms. As a result, the normative aspect of scientific disagreements has attracted the attention of scholars employing computer simulations in the form of agent-based models (ABMs). The primary value of such simulations is that they help in examining how different norms guiding individual scientists affect the performance of the given scientific community.[1]

In this paper I look into lessons obtained from such models, focusing on two types of norms: on the one hand, those that guide a scientist's attitude towards her own theory and on the other hand, those that guide a scientist's attitude towards her opponents' theory, after she has recognized she is involved in a peer disagreement. While the discussion on these norms is inspired by the peer disagreement debate in contemporary epistemology, it comes with certain peculiarities of scientific inquiry. Let's take a closer look at these issues.

The primary focus of the peer disagreement debate is the adequate doxastic attitude one should hold towards p upon recognizing that one's peer disagrees on p. At one end of the spectrum there is the Consilliatory Norm (CN), demanding one lower the confidence in p, split the difference in the

[1] Beside the case-study approach and computational methods, empirical studies of scientific disagreements are also becoming employed by the philosophical community (see e.g. Beebe et al., 2018). Such studies are important not only for the understanding of the descriptive state of affairs, but they may also serve as a basis for the formulation of normative conclusions, e.g. if combined with computer simulations that examine counterfactual scenarios.

sense of taking the middle ground between the opponent's belief and one's own belief on the issue, or entirely suspend their judgment on p. At the other end of the spectrum we have the Steadfast Norm (SN), according to which one should stick to one's guns and keep the same belief with the same confidence as before encountering a disagreeing peer.[2]

Similarly, in the context of scientific inquiry we can ask whether a scientist who has recognized she is involved in a peer disagreement should strive towards ameliorating the differences with her peer by means of (one of the versions of) CN, or whether she should remain steadfast. Nevertheless, the context of scientific inquiry comes with two caveats.

First, attitudes that guide scientists in their inquiry do not concern only scientists' beliefs about phenomena in the given scientific domain, but also their methodological assessments, commitments and preferences (see e.g. Elgin, 2010). As such they are not subject only to epistemic norms, but also to instrumental or methodological norms. While epistemic norms are traditionally understood as addressing the question what one is justified to believe in view of the available evidence, methodological norms address the question what one should do to attain the given goals – in this case: which actions one should perform to achieve scientific goals.[3] In view of such an interplay of epistemic and methodological aspects of inquiry, philosophers of science have introduced the notion of *cognitive attitude* when describing attitudes of scientists towards their theories, models, hypotheses, etc.[4] (see e.g. Elliott and Willmes, 2013; Lacey, 2015), aiming to cover both the epistemic and methodological (including the axiological) dimension of inquiry. For instance, the attitude of pursuit-worthiness towards a given hypothesis is a result of both: a retrospective epistemic evaluation, based on the available

[2]For arguments in favor of CN see e.g. Christensen, 2010; Elga, 2007; Feldman, 2007; Feldman, 2006; for arguments in favor of SN see e.g. Cruz and Smedt, 2013; Kelp and Douven, 2012; for reasons why norms are context-dependent see e.g. Christensen, 2010; Douven, 2010; Kelly, 2010a; Konigsberg, 2012.

[3]In the literature on (scientific) rationality we often find the third type of assessment: the axiological one. For instance, Rescher, 1988 distinguishes between epistemic, practical/instrumental and evaluative rationality. The latter concerns the assessment of our goals and their conduciveness to some more general ends. A similar point is made by Laudan (1984) in his 'reticulated model' of scientific rationality, according to which assessing scientific goals can be done in terms of their feasibility and the overall fit with the existing scientific practice. An important point made by both Rescher and Laudan is that all three types of assessments are interrelated since, e.g. what we believe (or accept) depends on what we do to obtain evidence, which depends on the goals we have; similarly, what we do and praise will depend on our beliefs, etc.

[4]I will henceforth use these notions—theory, hypothesis, model—interchangeably since the discussion applies to all of them.

evidence which provides indices that the hypothesis is supported by one's background knowledge, preliminary evidence, etc., as well as a prospective methodological evaluation, based on the insight that there are open lines of inquiry, that the hypothesis has a programmatic character, etc. (Šešelja and Straßer, 2014; Whitt, 1992).

The second issue specific for the context of scientific disagreements is that beside addressing scientists' attitudes towards their own theories, it is also important to address their attitude towards their opponents' theories. Should a rivaling theory be epistemically tolerated, engaged with or rather ignored? These questions lie at the heart of philosophical discussions on scientific controversies and methodological puzzles arising from them.

As we will see, these two caveats will be of direct relevance for the discussion making the bulk of this paper. Here is how I will proceed. In Section 2 I will look into the lessons we can draw from ABMs concerning norms guiding one's attitude towards one's own theory in face of a peer disagreement. In Section 3 I will do the same for norms concerning one's attitude towards one's opponent's theory. I conclude the paper in Section 4 by turning to prospects and limitations of employing ABMs for providing answers to normative questions about scientific disagreements and controversies.

2 The attitude towards one's own theory

As mentioned in Section 1, one way to examine whether, and if so, how scientists should adjust their attitude towards their current theory when encountering a peer who disagrees with them, is to evaluate the impact of different norms on the efficiency of the given scientific community. This is where the methodology of ABMs comes into play: by varying the behavior of individual agents, we can study emergent properties of the given community. In this section I will look into two ABMs – by Douven, 2010 and De Langhe, 2013 – designed to tackle this question. After providing a brief overview of each, I will discuss conclusions that have been drawn from them. Here I will focus on two issues: first, whether these conclusions can be considered a robust property in view of results obtained by structurally different ABMs of scientific inquiry, and second, whether the drawn message is the best interpretation of the given findings.

2.1 ABMs of peer disagreement

Both Douven's and De Langhe's ABMs are enhanced versions of the well-known Hegselmann and Krause's (Hegselmann and Krause, 2002, 2005,

2006) ABM of opinion dynamics. The aim of Hegselmann and Krause's (H&K's) model was to examine the process of consensus formation in a group of agents who adjust their beliefs by means of different processes of opinion aggregation.[5] Building on H&K's (2006) model, Douven and De Langhe represent scientists as truth seeking agents who are trying to determine the value of a certain parameter τ for which they only know that it lies in the interval $]0,1]$. While at the beginning of the simulation each agent is given a certain random value for τ within the interval, throughout the run of the model agents adjust their beliefs in view of their own research and by receiving information from some other agents.[6] According to Douven and De Langhe, this makes the model suitable for examining the impact of CN as the difference-splitting norm, or the lack thereof, representing SN. More precisely, we can examine the performance of agents in their attempts to determine the true value of τ a) if they update their beliefs both in view of their own research and in view of other agents' opinions (difference-splitting populations), or b) if they update their beliefs only in view of their own research (steadfast populations).

Douven's model suggests that whether difference-splitting is conducive to efficient inquiry or not is highly context dependent. For instance, when inquiry is easy (in the sense that agents get information that directly points at the true value of τ), difference-splitting populations converge on the true value of τ faster than steadfast populations. Nevertheless, if agents start receiving noisy data through their own research,[7] representing e.g. measure-

[5]This is the so-called bounded-confidence model in the sense that when adjusting their opinions agents take into account only those opinions of agents which are sufficiently similar to their own.

[6]The update of information is modeled in terms of the following function:

$$x_i(u+1) = \alpha \frac{1}{|X_i(u)|} \sum_{j \in X_i(u)} x_j(u) + (1-\alpha)\tau$$

where $x_i(u)$ is the opinion of agent x_i after the u-th update, $\alpha \in]0,1]$ is the weighting factor determining how much the opinions of others and one's own research influence the change of one's belief, $\tau \in]0,1]$ is the objective value of the parameter, $X_i(u) := \{j : |x_i(u) - x_j(u)| \leq \varepsilon\}$ with $\varepsilon \in [0,1]$ being the confidence interval determining the agents whose opinions are taken into account, and $|X_i(u)|$ the cardinality of $X_i(u)$.

[7]This is done by slightly adjusting the process of updating:

$$x_i(u+1) = \alpha \frac{1}{|X_i(u)|} \sum_{j \in X_i(u)} x_j(u) + (1-\alpha)(\tau + \mathrm{rnd}(\zeta))$$

where $\mathrm{rnd}(\zeta)$ is a function that gives a unique uniformly distributed real number in the interval $[-\zeta, +\zeta]$, where $\zeta \in [0,1]$.

ment errors, there appears to be a trade-off between accuracy and speed. While steadfast populations get within a moderate distance of the true value of τ relatively quickly, they don't improve their accuracy in the subsequent rounds of the simulation. In contrast, difference-splitting populations end up closer to the true value of τ but it takes them relatively longer to do so. In view of this Douven concludes that determining a rational response to peer disagreement largely depends on empirical issues, underlying the context of the given disagreement. Moreover, he takes his simulations to provide an argument against the claim that SN is necessarily irrational, since under some circumstances (e.g. when time and resources are scarce so that the speed of inquiry is more important than accuracy) it might be the right norm to follow.

De Langhe comes to a similar conclusion. His extension of the H&K model aims to represent longstanding scientific disagreements, typical for the context of theoretical diversity where there are multiple epistemic systems[8] in the given scientific domain. He represents such a diversity by making the objective value of τ relative to the agent's epistemic system. In addition, agents in the model are able to distinguish between (i) beliefs of peers within their epistemic system and (ii) beliefs of peers outside of their epistemic system (for technical details see De Langhe, 2013, p. 2552.) The simulations suggest that there is a trade-off between difference-splitting within one's epistemic system and difference-splitting between epistemic systems. More precisely, in order for each group of agents to converge on the true value of their τ, it is not beneficial to split the difference with agents from other epistemic systems.[9]

2.2 Epistemic vs. methodological steadfastness

It seems then that both of these models provide scenarios in which SN may be epistemically preferable to CN. Moreover, the robustness of these results appears to be supported by results obtained by structurally different models representing theoretical diversity. For instance, the findings of Zollman's

[8] De Langhe employs Goldman's (2010) idea that even though disagreeing peers may share the evidence concerning the given issue in question, they may not share the evidence for the epistemic system within which they evaluate the former (object-level) evidence.

[9] This is however not surprising since difference-splitting with agents from other epistemic systems bears information on *their own respective* τ, while the success of each agent is measured by how close she gets to τ *in her own system.* Note that this issue may pose a more general conceptual problem for De Langhe's model since it is not clear which epistemic benefits are included in the representation of interaction between agents belonging to different epistemic systems.

(2010) suggest that extreme initial opinions of agents, making them steadfast in their inquiries, can help in preventing the community from prematurely abandoning the better of two theories. A recent enhancement of Zollman's model by Frey and Šešelja, 2018a suggests that equipping agents with a doze of inertia towards their theories improves the efficiency of their inquiry. Similar results have been obtained by the argumentation-based ABM proposed by Borg et al., 2019, at least under some conditions underlying scientific inquiry.

Hence, simulations seem to provide a strong argument in favor of SN at least to the extent that sometimes it may be a better option than CN. Or do they? In what follows I argue that they do not since the behavior of agents in all of these models does not represent SN understood as an epistemic norm, but steadfastness as a methodological norm, compatible with both SN and CN.

Recall that according to SN one should stick to one's belief upon discovering a peer who disagrees on the given issue. In Douven's and De Langhe's simulations a non-difference splitting agent will update her hypothesis concerning τ in view of her own evidence, ignoring the beliefs of other agents. That means that she is steadfast in the sense of sticking to her hypothesis as a premise guiding her further inquiry. Now, if we assume that agents' beliefs are correlated with their pursuit-related attitudes, then it indeed follows that non-difference splitting agents employ SN. The problem is, however, that one's beliefs concerning a given hypothesis do not need to be correlated with one's attitude as for whether this hypothesis if worthy of pursuit. As often discussed in the literature on pursuit worthiness (e.g. Nickles, 2006; Šešelja, Kosolosky, and Straßer, 2012; Whitt, 1990, 1992), one may suspend the judgment on whether hypothesis h is true and yet consider it highly worthy of pursuit. And the other way around: one may believe h and at the same time consider it a closed question, not worthy of further pursuit. Hence, what the non-difference splitting agents in the above models employ is not steadfastness as an epistemic norm, but as a *methodological norm*, according to which they should not abandon their current hypothesis in face of a peer disagreement. Instead, they are supposed to retain the methodological attitude of pursuit worthiness towards their current hypothesis, which Fleisher, 2017 has recently dubbed 'rational endorsement'.[10] Such an attitude expresses one's strong commitment towards the given hypothesis, but it does not require from one to believe the given hypothesis, nor to remain

[10] Not to be confused with Hugh Lacey's notion of endorsement which refers to the acceptance of hypotheses in the context of application, see Lacey, 2015.

steadfast in her beliefs towards it. To the contrary, as Fleisher points out, one may not believe her current hypothesis at all, suspecting it may be false, or believe the opponent's hypothesis to be superior, and nonetheless stick with the former as long as it is a viable option, worthy of further pursuit.

Note, however, that such an attitude is perfectly compatible with CN as an epistemic norm: one may indeed lower one's confidence in the belief that the given hypothesis is true, suspend the judgment on it, or even form her belief by splitting the difference, and yet not include these updated beliefs as premises that guide her further inquiry. In order to argue that the results of the above discussed models provide an argument for SN as an epistemic norm, one would have to show that the epistemic and the methodological steadfastness are correlated, that is, that agents employing the (epistemic) SN are more likely to stick with their current hypothesis than if they don't employ this norm. This is however an empirical question, neither addressed in the context of these ABMs, nor is it self-evident.[11] In other words, the methodological impact of the epistemic SN is absent from the models.[12]

Altogether, results of various simulations suggest that in some circumstances one should not immediately abandon one's current inquiry upon encountering a disagreeing peer. Moreover, as we have seen, the benefits of such methodological steadfastness have been confirmed a robust finding across various models, at least in certain contexts of inquiry. While this is not a very surprising result, it remains to be seen whether more can be said about the conditions under which methodological steadfastness increases efficiency of inquiry, and those under which it could perhaps be harmful. For instance, our recent study via an argumentation-based model (Borg et al., 2019, see below Section 3) suggests that the methodological steadfastness, represented in the model as 'cautious decision making', is beneficial only if

[11] Even though philosophers have sometimes conjectured what such a relationship may look like (e.g. Magnus, 2014 suggests that "scientists who cultivate agnosticism might not pursue their chosen research program with the necessary vigor. The community would then do better if those individuals fully embraced the presuppositions of their approach." (p. 132)), as explained above, the situation is not so simple, and only via a proper empirical study can we obtain reliable information about this relationship.

[12] Given that the ABMs in question were originally developed to address issues discussed in the literature on peer disagreement around 2005, this is not surprising. The whole setup of the peer disagreement debate revolved around doxastic attitudes, and the importance of alternative cognitive attitudes has only recently entered the discussion (see e.g. Fleisher, 2018). At the same time, ABMs of scientific inquiry are still typically based on the assumption that an agent's beliefs and pursuit-related attitudes are mutually correlated. While sometimes this may be a harmless idealization, in case differentiating between the two could have an impact on conclusions we draw from the model, it is important to keep this distinction in mind.

the given scientific community has a high degree of connectedness, and if furthermore, scientists base their pursuit decisions on a specific type of evaluation (e.g., when they prefer theories with a larger scope than their rivals, rather than, for example, theories that are less anomalous than their rivals). We add, however, that the methodological steadfastness can be additionally beneficial if we take into account costs related to changing one's pursued theory, which can be avoided if scientists preserve a dose of inertia. This indicates that the impact of methodological steadfastness on the efficiency of inquiry is still an open question, where both empirical and computational methods can help to address it.

3 The attitude towards one's opponent's theory

While the literature on the epistemology of peer disagreement has been primarily focused on the norms guiding one's attitude towards one's own beliefs upon encountering a disagreeing peer, philosophers of science discussing scientific disagreements have primarily focused on how one should treat the opponent's views. In particular, scientific pluralists[13] have argued that scientists should interact with their opponents or at least tolerate their views (see e.g. Chang, 2012; Lacey, 2009; Longino, 2002). In addition, my collaborators and I have proposed a normative account of epistemic toleration in the context of scientific disagreements, arguing that toleration implies interaction (Straßer, Šešelja, and Wieland, 2015). The basic idea of epistemic toleration is as follows: upon recognizing indications that the stance of my opponent is a result of a rational deliberation, I have (i) a duty to treat her stance in a charitable way, as potentially rational, potentially non-futile and thus potentially fertile and (ii) a duty to consider my opponent's stance as a potentially serious competition and a challenge to my own stance and a duty to stay in critical correspondence with my opponent (p. 131).

While (critical) interaction among scientists may indeed be beneficial for their impartiality and the reliability of their knowledge about the current state of their field, its impact on the efficiency of inquiry (in the sense of making the greatest progress with as little resources as possible) is not so straightforward. For example, is interaction among scientists beneficial at all times, or are there situations where restricting the information flow would lead to a more efficient inquiry? Ideally, we would like to have an empirically informed answer to this question, but this may be hard to come by. This is where simulations of scientific inquiry may once again come in

[13]For a recent discussion on different types of scientific pluralism see Šešelja, 2017.

handy. And indeed, throughout the last decade a variety of ABMs of scientific interaction has entered the philosophical literature. In this section I will examine results of three classes of such ABMs (some of which have already been mentioned in the previous section): those proposed by Zollman and inspired by his work (Frey and Šešelja, 2018a; Zollman, 2007, 2010), those proposed by Grim, Singer and collaborators (Grim, 2009; Grim et al., 2013), and our argumentation-based ABMs (Borg et al., 2017a,b, 2018, 2019). After briefly presenting the basic idea of each of these models, I discuss lessons that can be drawn from them concerning the effects of interaction on the efficiency of inquiry, focusing on the reliability of current conclusions and future prospects of this line of research.

3.1 ABMs of scientific interaction

A common feature of ABMs of scientific interaction is that interaction is represented in terms of different types of social networks. Different social networks capture different degrees of connectedness among the members of a given community (see Figure 1). The representation of other aspects of scientific inquiry and the information flow among scientists varies from one model to another.

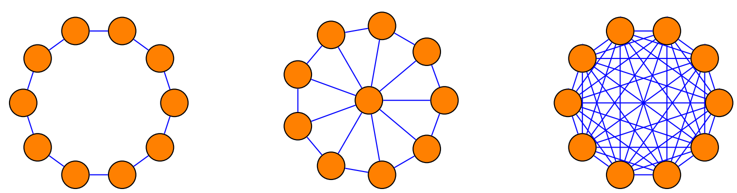

Figure 1: Three commonly employed social networks, representing an increasing degree of connectedness: a cycle, a wheel and a complete graph. The nodes in each of the graphs represent agents (or groups of agents), while edges that connect the nodes represent transmission of information between two agents (or two groups) (the figure is taken from Borg et al., 2017b, p. 397).

Social networks and bandit problems. Zollman's models are based on the analogy between scientific inquiry and the so-called bandit problems, well-known in economics and statistics. The latter concern a situation in which a gambler (or a group of gamblers) is confronted with multiple slot machines (bandits), which have different probabilities of success, where a gambler is trying to determine which machine will give her a better payoff. This type of uncertainty is similar to the one scientists find themselves in when confronted with multiple rivaling hypotheses, where they try to determine which hypothesis is the best one.

At the beginning of the simulation scientists are assigned random prior probabilities for two rivaling hypotheses, each of which has a designated objective probability of success, unknown to the agents-scientists. Agents always choose to pursue a theory which they consider to be better. During the simulation they update their beliefs in view of their own findings and by receiving information from other agents with whom they are connected in a social network.[14] Zollman employs three types of social networks from Figure 1. His results suggest that the degree of connectedness of the scientific community is inversely proportional to the success of scientists in converging on the objectively better hypothesis. The reason why a fully connected community often fails in converging on the better hypothesis is that initial findings by scientists may be misleading, and due to the full connectedness, they may spread quickly throughout the whole community, resulting in a premature abandonment of the objectively better hypothesis.

As subsequent studies have shown, however, this result holds only for a small portion of the relevant parameter space (see Rosenstock, O'Connor, and Bruner, 2017) and only under specific idealizing assumptions concerning the decision-making by scientists (see Frey and Šešelja, 2018a). In particular, if scientists are represented as employing a dose of caution when deciding if they should abandon their current theory and start pursuing the rivaling one, the cycle is not anymore superior to the complete graph.[15] Interestingly, we also show that adding the assumption that *critical* interaction is epistemically beneficial doesn't on its own help the complete graph to catch up with the cycle: cautious decision-making appears necessary for such an

[14]Every round an agent makes 1,000 pulls, each of which can be a success or a failure, where the probability of success is given by the objective probability of success of the respective hypothesis. Agents then update their beliefs via Bayesian reasoning (modeled by means of beta distributions). Note that the model described here is Zollman's (2010) model, which is a generalized version of his (2007) one.

[15]Our enhancement of Zollman's ABM (Frey and Šešelja, 2018a) is based on the observation that, on the one hand, Zollman's result hinges on the parameter choices for the objective probability of success of two hypotheses, namely 0.499 and 0.5, and that, on the other hand, if scientists are considered successful only if they converge on the better hypothesis, then the difference between the two hypotheses should gradually increase. In other words, as scientists improve their methodology, they should have a better grasp of the difference between the rivaling approaches. The corollary of implementing this assumption is that scientists always converge on the better hypothesis, the only question is how long it takes them to do so. Consequently, efficiency in this model is measured in terms of time (needed for the successful convergence) rather than in terms of the percentage of successful runs.

effect to take place.

Social networks and epistemic landscapes. A result similar to the 'Zollman effect' (a superiority of the cycle to the complete graph in terms of a higher chance of successful convergence) was obtained by a set of models proposed by Grim, 2009; Grim et al., 2013. Grim and Singer's models employ a one-dimensional landscape which represents a range of rivaling hypotheses in a given domain. Each hypothesis is assigned a certain value representing its epistemic success, and agents are successful if they manage to discover the hypothesis with the highest value. At the beginning of the simulation agents are randomly positioned on the landscape. They learn the values of a hypothesis either by being positioned on it, or from other agents with whom they are connected in a social network. In case an agent learns that a hypothesis of another agent is more successful, she will start moving towards it.[16] The simulations were run for different types of social networks (such as the cycle and the complete graph) and for different types of epistemic landscapes.

The results of the model indicate that all the examined networks are equally successful in case of simpler epistemic landscapes, with smooth climbs to their peaks, though more connected groups reach their goal faster. However, in case of landscapes with a narrow hidden peak, the complete graph performs the worst, while less connected networks, such as the cycle, perform the best.[17] Similarly to Zollman's model, more connected networks tend to prematurely abandon exploration, converging onto a local maximum and leaving the global maximum undiscovered. This conclusion seems to support the thesis that the Zollman effect is more likely to occur in conditions in which discovering the optimal hypothesis is more difficult.

Social networks and argumentative dynamics. The robustness of the above results is challenged by the findings obtained with our argumentation-based ABM (ArgABM) (Borg et al., 2017b, 2018). ArgABM aims to capture the argumentative nature of scientific interaction, typical for the context of theoretical diversity and scientific disagreements. The model employs an 'argumentative landscape', representing rivaling research programs in a

[16] Agents move with the *speed* of 0.5, approximating her target halfway each round with the probability of 0.5 (the latter represents *inertia* of the agent). Each time an agent moves, she jumps to a random region within four points either side of the target spot. This represents a *shaking hand* phenomenon, namely, an attempt at replicating the target hypothesis which may give slightly different results (Grim, 2009, Section 3).

[17] Grim, 2009 shows that there is a threshold below which lower connected networks begin to perform worse.

given domain, which scientists gradually explore.[18]

Each theory (or a research program) is modeled as a (rooted) tree consisting of a number of arguments – represented abstractly as nodes in a directed graph, connected by a 'discovery relation'. The discovery relation stands for paths that agents take when moving on the landscape, from one argument to another. The argumentative dynamics in the model comes into play through the presence of 'argumentative attacks' that occur between different theories: an argument belonging to one theory can attack an argument of one of the rivaling theories. Roughly speaking, an argument in a theory is considered *defended* if it is not attacked, or if there is another defended argument in the same theory, which attacks the attacker-argument.[19] Throughout the simulation agents gather knowledge about the objective landscape by learning arguments in favor of each theory, as well as attacks on these arguments. In addition to gathering knowledge on their own, they also learn about the landscape from other agents with whom they are linked in a social network. The model employs networks with different degrees of connectedness, such as those in Figure 1. In view of this knowledge, agents evaluate the theories. Since the landscape is pre-defined in such a way that one theory is fully defended from all the attackers on it (in the objective landscape, unknown to the agents at the beginning of the simulation), the success of inquiry is measured in terms of the percentage of runs in which agents successfully converge on this theory[20].

In contrast to the previous models, the results of ArgABM don't indicate the Zollman effect. To the contrary, the degree of the connectedness of agents is directly proportional to their success in converging on the best theory, under a variety of conditions.

[18] Both the landscape and the decision-making of agents in the model are inspired by Abstract Argumentation Frameworks, pioneered by Dung, 1995 and previously used for the modeling of scientific debates in our (Šešelja and Straßer, 2013).

[19] More precisely, a subset of arguments A of a given theory T is *admissible* iff for each attacker b of some a in A there is an a' in A that attacks b. An argument a in T is said to be *defended in* T iff it is a member of the maximally admissible subset of T (note that each theory in the model is conflict-free in the sense that no two arguments in it attack one another).

[20] In (Borg et al., 2018) we also present an alternative, pluralist criterion of success, according to which a community is successful if at the end of the run the best theory doesn't have fewer agents than either of the rivaling theories (p. 295; the results of ArgABM simulations are usually obtained by employing a landscape consisting of three theories, though they are similar if the number of theories is reduced to two).

3.2 Different models, different types of inquiry?

What can we then conclude in view of ABMs of scientific interaction about the conduciveness of interaction to the efficiency of inquiry? The variety of results suggest not only that the answer may be highly context-dependent, but that at this point we cannot draw any reliable normative conclusion. While under certain conditions a high degree of interaction may increase the risk of a premature abandonment of fruitful hypotheses, under different conditions such a risk may be rather low. This means that in order to determine which results are relevant for actual scientific practice (in all its variety) it is essential to specify the type of inquiry represented by each of the given models.

This can be done by, on the one hand, analyzing structural differences between the models. For instance, the success of fully connected communities in ArgABM is due to the way information and knowledge are represented in the model (for a more detailed discussion see Borg et al., 2019). In particular, the accuracy of a scientist's assessment of the given theories depends on how much knowledge of the landscape she has, where larger gaps in knowledge can easily result in an erroneous theory assessment. Given that agents share only recently acquired information (rather than their full knowledge of the landscape), scientists in less connected communities may easily have a permanent information loss and end up with a 'patchy' knowledge of the theories. In contrast, in Zollman's model any shared information is representative of the whole theory, and as a result, information losses are much less harmful. Hence, these two scenarios correspond to different types of inquiry.

On the other hand, beside specifying the target the model seems to represent (in this case, a specific type of inquiry), it is important to empirically embed the model in order to provide further evidence that it is indeed informative of the given phenomena.[21] Such a process of empirically embedding ABMs of science includes at least one of the following methods: (i) empirical calibration – using empirical data to constrain the parameters in the model; (ii) theoretical embedding – using theoretical accounts from philosophy of science, sociology and psychology to inform the assumptions in the model; and (iii) empirical validation – the analysis and comparison

[21] This is important even in case of models that aim to provide a how-possibly explanation of the given target since not all possibilities are interesting in the sense that we can derive from them relevant information about real-world phenomena (see Frey and Šešelja, 2018b) and may instead amount only to 'just so stories' (Verreault-Julien, 2018) or 'model based story telling' (Arnold, 2006).

of a simulated macro behavior with the real-world macro behavior (Boero and Squazzoni, 2005; Casini and Manzo, 2016; Šešelja, 2018). For instance, a model of scientific interaction could be calibrated by taking into account a typical size of the given community of scientists (e.g. by using blibliometric data, such as the number of authors who have published in the given domain within a certain time period). Next, the behavior of scientists could be constrained in view of empirical studies suggesting how often scientists exchange information, what kind of information is exchanged, etc.[22] Finally, we could attempt to validate the model with respect to its predictive accuracy (Thicke, 2018) where we measure how well it reproduces certain patterns of group behavior (e.g. the dynamics of consensus formation in the relevant historical episodes). Another way to empirically validate ABMs is by means of experimental studies (such as those by Mason and Watts, 2012; Mason, Jones, and Goldstone, 2008)[23]

Hence, despite the tendency in the philosophical literature to consider ABMs as 'valid unless proven otherwise', what exactly such ABMs represent (which type of inquiry) and how relevant these conclusions are for actual scientific practice remains completely open in the absence of validation procedures. In the lack of validation, we should consider these models to be exploratory, having the status of theoretical abstractions (Boero and Squazzoni, 2005; Šešelja, 2018). As such, they can provide theoretical insights about scientific rationality (rather than providing explanations of cases from actual scientific practice), though the exact value of such theoretical insights remains a contested issue (see e.g. Arnold, 2013).

3.3 How to argue for scientific pluralism?

But where does this leave proponents of scientific pluralism, mentioned at the beginning of this section? What can they conclude, if anything, in view of the above discussed models? Let's note that there are two types of arguments that can be used in favor of the pluralist methodology, that is, for the norm that scientists should tolerate rivaling approaches and interact

[22]Clearly, beside descriptively adequate representation, a model can incorporate counterfactual assumptions if examining them is interesting from a normative perspective (e.g. we could use simulations to compare different types of information sharing and their relative impact on the efficiency of the given community).

[23]Experiments may however represent a different target phenomenon than the given model, and moreover, further studies may give conflicting results. For instance, the study by Mason, Jones, and Goldstone, 2008 suggesting that less connected networks have a better problem-solving performance than the fully connected ones was subsequently challenged by Mason and Watts, 2012 who found the opposite to be the case.

with their proponents:[24]

1. On the one hand, we can defend this norm in terms of its conduciveness to scientific (cognitive and non-cognitive) goals of *individual* scientists, such as their reliability, impartiality, (Lacey, 2013, 2014) or certain types of scientific objectivity (Douglas, 2009), etc.

2. On the other hand, we can defend the norm in terms of its conduciveness to the *communal* scientific (cognitive and non-cognitive) goals, such as the efficiency in collective knowledge acquisition.

Unfortunately, proponents of scientific pluralism (such as Chang, 2012 and Longino, 2013) have been rather vague when it comes to clarifying which of these two types of argumentative strategies they employ, leaving an impression of their conflation. Nevertheless, the above discussion on ABMs shows just how complex the latter strategy (arguing for the pluralist methodology in terms of its conduciveness to the communal scientific goals) is, at least if we are focusing on the efficiency of inquiry as the communal goal. As we have seen, it is hard to estimate whether unrestricted interaction among scientists is conducive to the efficiency of the given community, and if so, under which conditions. Whether we might get more reliable insights by means of empirical studies remains to be seen,[25] but at this point, there is no reliable information in view of which we could argue for the pluralist methodology via this path.

Consequently, pluralists would be better off employing the former strategy, showing the conduciveness of the pluralist methodology to scientific goals of individual scientists, and in extension to those community goals that directly result from the maximization of individual ones (e.g. there

[24] In (Straßer, Šešelja, and Wieland, 2015) we define the epistemic toleration as a conditional norm, which is triggered if "the tolerated stance is considered objectionable and in an important sense epistemically problematic" and if "there are reasons—namely, the indices of [rational disagreement]—in view of which it would be wrong not to tolerate an objectionable stance." Epistemic toleration, however, has its limits and it is not triggered if one has "reasons to consider the stance of the opponent as futile", for instance in case of "empirically backed up reasons to suppose bias or fraud on the side of the opposition, the refusal to take part in argumentative exchange, a systematic reluctance to put hypotheses under critical empirical tests, systematic self-immunization from empirical and argumentative scrutiny, etc." (p. 128-129).

[25] Historical case studies may also be helpful, especially if combined with the analysis of bibliometric data, in order to generate sufficiently broad evidence base. Such results could also be used for the empirical embedding of ABMs as suggested by Frey and Šešelja, 2018b and recently employed by Harnagel, 2018.

may be good reasons to consider the impartiality of individual scientists increasing the impartiality of the given community). In other words, pluralists would be better off showing how epistemic toleration and critical interaction benefit the achievement of scientific goals of individual scientists (*qua* scientists), irrespective of whether this results in a more efficient communal enterprise. After all, even if we obtained a convincing evidence (via empirical or computational methods) suggesting that pluralist norms are counterproductive for the efficiency of science as a collective enterprise, i.e. that there is a trade-off between efficiency and other scientific goals, it is not clear why pluralists would opt for the former.[26]

4 Conclusion

In this paper I have examined the current status of philosophical investigations into normative aspects of scientific disagreements by means of ABMs. To this end, I have focused on two types of norms that guide a scientist who has recognized she is involved in a peer disagreement: norms concerning one's attitude towards one's own inquiry, and norms concerning one's attitude towards one's opponent's inquiry. In both cases I have provided a survey of relevant ABMs and a discussion of conclusions that have been drawn in view of the simulations.

Taking stock of this literature, the primary limitation and the primary challenge to the employment of the ABM methodology appears to be their validation, which would help in linking these models to actual scientific inquiry. Whether we have a case of a robust property across structurally different models (as discussed in Section 2) or a case of different models providing different results (as discussed in Section 3), ABMs need to be empirically embedded in order to become useful for formulating normative suggestions in the context of scientific disagreements. In addition, this process could also benefit from a closer engagement with the literature on peer disagreement, which would help with a better theoretical embedding of the models. This especially concerns factors that are considered of epistemic significance, such as higher-order evidence (Christensen, 2010; Feldman, 2005; Kelly, 2010b; Straßer, Šešelja, and Wieland, 2015), which has so far largely been absent from ABMs of scientific interaction (a recent exception is Merdes, 2018),

[26] As we write in (Straßer, Šešelja, and Wieland, 2015): "To caricature it a bit ... do we rather want a science where the scientists are individually rational but the scientific machinery may sometimes move a bit slowlier than optimal, or do we want a scientific machinery that performs most efficiently but where the scientists may sometimes put on blinkers which make them suboptimal viz. slightly dogmatic epistemic agents?" (p. 145).

though it may play an important role in the formulation of epistemic and methodological norms.

Acknowledgments. I would like to thank Andrea Robitzsch for valuable discussions on epistemic and methodological norms, which inspired parts of this paper. I am also grateful to two anonymous referees, to Borut Trpin and to the audience of the MAP MCMP (Minorities and Philosophy at the Munich Center for Mathematical Philosophy) seminar where I first presented this paper, for valuable comments. Research for this paper was funded by the DFG (Research Grant HA 3000/9-1).

References

Arnold, Eckhart (2006). "The Dark Side of the Force: When computer simulations lead us astray and "model think" narrows our imagination—Pre conference draft for the Models and Simulation Conference, Paris, June 12-14". accessed on October 31, 2018. URL: `https://eckhartarnold.de/papers/2006_simulations/node10.html`.

— (2013). "Simulation Models of the Evolution of Cooperation as Proofs of Logical Possibilities. How Useful Are They?" In: *Etica & Politica/ Ethics & Politics* XV (2), pp. 101–138.

Barnes, Trevor J and Eric Sheppard (2010). "'Nothing includes everything': towards engaged pluralism in Anglophone economic geography". In: *Progress in Human Geography* 34.2, pp. 193–214.

Beebe, James R, Maria Baghramian, Luke O'C Drury, and Finnur Dellsen (2018). "Divergent Perspectives on Expert Disagreement: Preliminary Evidence from Climate Science, Climate Policy, Astrophysics, and Public Opinion". In: *arXiv preprint arXiv:1802.01889*.

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

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

Borg, AnneMarie, Daniel Frey, Dunja Šešelja, and Christian Straßer (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.

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

— (2019). "Theory-choice, transient diversity and the efficiency of scientific inquiry". In: *European Journal for Philosophy of Science*. http://doi.org/10.1007/s13194-019-0249-5.

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

Chang, Hasok (2012). *Is Water H2O? Evidence, Pluralism and Realism.* Springer.

Christensen, David (2010). "Higher-Order Evidence". In: *Philosophy and Phenomenological Research* 81.1, pp. 185–215.

Cruz, Helen De and Johan De Smedt (2013). "The value of epistemic disagreement in scientific practice. The case of Homo floresiensis". In: *Studies in History and Philosophy of Science Part A* 44.2, pp. 169–177.

De Langhe, Rogier (2013). "Peer disagreement under multiple epistemic systems". In: *Synthese* 190, pp. 2547–2556.

Douglas, Heather E. (2009). *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press.

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

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

Elga, Adam (2007). "Reflection and disagreement". In: *Noûs* 41.3, pp. 478–502.

Elgin, Catherin Z. (2010). "Persistent Disagreement". In: *Diasagreement*. Ed. by Richard Feldman and Ted A. Warfield. Oxford: Oxford University Press, pp. 53 –68.

Elliott, Kevin C and David Willmes (2013). "Cognitive attitudes and values in science". In: *Philosophy of Science* 80.5, pp. 807–817.

Feldman, R. (2007). “Reasonable Religious Disagreements”. In: *Philosophers without Gods*. Ed. by L. M. Antony. OUP, pp. 194–214.
Feldman, Richard (2005). “Respecting the Evidence”. In: *Philosophical Perspectives* 19.1, pp. 95–119.
— (2006). “Epistemological Puzzles About Disagreement”. In: *Epistemology Futures*. Oxford University Press, pp. 216–326.
Fleisher, Will (2017). “Rational endorsement”. In: *Philosophical Studies*, pp. 1–27.
— (2018). “How to Endorse Conciliationism”. In: https://doi.org/doi:10.7282/t3-z234-rj23.
Frey, Daniel and Dunja Šešelja (2018a). “Robustness and Idealization in Agent-Based Models of Scientific Interaction”. In: *British Journal for the Philosophy of Science* https://doi.org/10.1093/bjps/axy039.
— (2018b). “What is the Epistemic Function of Highly Idealized Agent-Based Models of Scientific Inquiry?” In: *Philosophy of the Social Sciences* https://doi.org/10.1177/0048393118767085.
Goldman, Alvin (2010). “Epistemic relativism and reasonable disagreement”. In: *Disagreement*, pp. 187–215.
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.
Harnagel, Audrey (2018). “A Mid-Level Approach to Modeling Scientific Communities”. In: *Studies in History and Philosophy of Science.* https://doi.org/10.1016/j.shpsa.2018.12.010.
Hegselmann, Rainer and Ulrich Krause (2002). “Opinion dynamics and bounded confidence models, analysis, and simulation”. In: *Journal of artificial societies and social simulation* 5.3.
— (2005). “Opinion dynamics driven by various ways of averaging”. In: *Computational Economics* 25.4, pp. 381–405.
— (2006). “Truth and cognitive division of labor: First steps towards a computer aided social epistemology”. In: *Journal of Artificial Societies and Social Simulation* 9.3, p. 10.
Kelly, Thomas (2010a). “Peer disagreement and higher order evidence”. In: *Social Epistemology: Essential Readings*, pp. 183–217.

Kelly, Thomas (2010b). “Peer Disagreement and Higher-Order Evidence”. In: *Disagreement*. Ed. by Richard Feldman and Ted A. Warfield. Oxford: Oxford University Press, pp. 111 –174.

Kelp, C. and I. Douven (2012). “Sustaining a Rational Disagreement”. In: *EPSA Philosophy of Science: Amsterdam 2009*, pp. 101–110.

Konigsberg, Amir (2012). “The Problem with Uniform Solutions to Peer Disagreement”. In: *Theoria*.

Kuhn, Thomas (1977). *The Essential Tension: selected studies in scientific tradition and change*. Chicago: University of Chicago press.

Lacey, Hugh (2009). “The Interplay of Scientific Activity, Worldviews and Value Outlooks”. In: *Science and Education* 18, 839–860.

— (2013). “Rehabilitating neutrality”. In: *Philosophical studies* 163.1, pp. 77–83.

— (2014). “Science, respect for nature, and human well-being: democratic values and the responsibilities of scientists today”. In: *Foundations of Science*, pp. 1–17.

— (2015). “‘Holding’and ‘endorsing’claims in the course of scientific activities”. In: *Studies in History and Philosophy of Science Part A* 53, pp. 89–95.

Laudan, Larry (1984). *Science and Values*. University of California Press.

Longino, Helen (2002). *The Fate of Knowledge*. Princeton: Princeton University Press.

Longino, Helen E (2013). *Studying human behavior: how scientists investigate aggression and sexuality*. University of Chicago Press.

Magnus, PD (2014). “Science and rationality for one and all”. In: *Ergo* 1.5. DOI: http://dx.doi.org/10.3998/ergo.12405314.0001.005, pp. 129–138.

Mason, Winter and Duncan J Watts (2012). “Collaborative learning in networks”. In: *Proceedings of the National Academy of Sciences* 109.3, pp. 764–769.

Mason, Winter A, Andy Jones, and Robert L Goldstone (2008). “Propagation of innovations in networked groups.” In: *Journal of Experimental Psychology: General* 137.3, p. 422.

Merdes, Christoph (2018). “Strategy and the pursuit of truth”. In: *Synthese*. https://doi.org/10.1007/s11229-018-01985-x.

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

Rescher, Nicholas (1988). *Rationality: A philosophical inquiry into the nature and the rationale of reason*. Oxford University Press.

Rolin, Kristina (2011). “Diversity and Dissent in the Social Sciences: The Case of Organization Studies”. In: *Philosophy of the Social Sciences* 41.4, pp. 470–494.

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.

Šešelja, Dunja (2017). “Scientific Pluralism and Inconsistency Toleration”. In: *Humana.Mente Journal of Philosophical Studies* 32, pp. 1–29.

Šešelja, Dunja (2018). “Exploring Scientific Inquiry via Agent-Based Modeling”. In: *Forthcoming*. URL: `http://philsci-archive.pitt.edu/15120/1/Exploratory_ABMs.pdf`.

Šešelja, Dunja, Laszlo Kosolosky, and Christian Straßer (2012). “Rationality of Scientific Reasoning in the Context of Pursuit: drawing appropriate distinctions”. In: *Philosophica* 86, pp. 51–82.

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

— (2014). “Epistemic justification in the context of pursuit: a coherentist approach”. In: *Synthese* 191.13, pp. 3111–3141.

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.

Solomon, Miriam (2006). “Groupthink versus The Wisdom of Crowds: The Social Epistemology of Deliberation and Dissent”. In: *The Southern Journal of Philosophy* 44, 28–42.

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

Thicke, Mike (2018). “Evaluating Formal Models of Science”. In: *Journal for General Philosophy of Science* accepted for publication.

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

Whitt, Laurie Anne (1990). “Theory Pursuit: Between Discovery and Acceptance”. In: *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*. Vol. 1, pp. 467–483.

— (1992). “Indices of Theory Promise”. In: *Philosophy of Science* 59, pp. 612–634.

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.