

# Truth Approximation, Social Epistemology, and Opinion Dynamics

Igor Douven · Christoph Kelp

Received: 30 March 2010 / Accepted: 4 November 2010 / Published online: 4 August 2011  
© The Author(s) 2011. This article is published with open access at Springerlink.com

**Abstract** This paper highlights some connections between work on truth approximation and work in social epistemology, in particular work on peer disagreement. In some of the literature on truth approximation, questions have been addressed concerning the efficiency of research strategies for approximating the truth. So far, social aspects of research strategies have not received any attention in this context. Recent findings in the field of opinion dynamics suggest that this is a mistake. How scientists exchange and take into account information about each others' beliefs may greatly influence the accuracy and speed with which the scientific community as a whole approximates the truth. On the other hand, social epistemologists concerned with peer disagreement have so far neglected the question of how practices of responding to disagreements with peers fare with respect to the goal of approximating the truth. Again, work on opinion dynamics shows that this may be a mistake, and that how we ought to respond to disagreements with our peers may depend on the specific purposes of our investigations.

This paper relates the topic of truth approximation to recent work in social epistemology.<sup>1</sup> In particular, it is shown that much of the work on opinion dynamics

---

<sup>1</sup> By “social epistemology” we mean *veritistic* social epistemology in the sense of Goldman (1999), that is, the systematic study of social practices with respect to their tendency to disseminate true beliefs in a society of truth-seeking agents. Thus understood, social epistemology is to be distinguished from social studies of knowledge and other fields that are also sometimes designated by the name of social epistemology; see Goldman (1999, Chap. 1).

---

I. Douven (✉)  
Faculty of Philosophy, University of Groningen, Groningen, The Netherlands  
e-mail: i.e.j.douven@rug.nl

C. Kelp  
Institute of Philosophy, University of Leuven, Leuven, Belgium  
e-mail: christoph.kelp@hiw.kuleuven.be

that has been carried out lately shows that social aspects of research strategies should be of direct interest to those working on the issue of truth approximation. Conversely, we argue that in the debate concerning peer disagreement, which social epistemologists have devoted much attention to in the past years, issues related to truth approximation have been unduly neglected, and that progress is to be expected from considering from a truth-approximation perspective the question of whether an overt disagreement among parties that take each other to be epistemically equally well positioned with respect to a given issue can be rationally sustained.

It is no exaggeration to say that the bulk of the work related to truth approximation has focused on getting the relevant *definitions* right: what exactly does it mean to say that a given theory is close to the truth, or that one theory is closer to the truth than another? This project turned out to be much harder than it at first appeared. The literature now contains various sophisticated proposals for defining the notion of truth-closeness as well as for defining related notions (e.g., estimated truth-closeness). We will have nothing new to say on these, or indeed on how to best define truth-closeness. Whenever we assume a particular definition of truth-closeness in this paper, the context will be simple enough to make the assumed definition entirely uncontroversial.<sup>2</sup>

Some of those interested in truth approximation have gone beyond the definitional work and investigated the question of which method or methods are most efficient for approximating the truth.<sup>3</sup> This related the issue of truth approximation to confirmation theory. We aim to address the question of efficiency in approximating the truth from a different angle, specifically, a socio-epistemic angle.

Modern science is fundamentally, perhaps even essentially, a social enterprise, in that most of its results not just happen to be due to, but *require* for their obtainment, the collaborative efforts of (often large) groups of scientists. At the same time, science is clearly a goal-directed enterprise. Specifically, it is plausible to assume that the aim of scientific practice is to approach the truth. From a social engineering perspective, then, the question arises how scientific collaboration is best to take place in order to achieve this goal: which forms of collaboration and interaction among scientists will be most conducive to approaching the truth. However, more specific questions will also crop up. Supposing that it is desirable to approach the truth both quickly and accurately, is there one way of organizing collaborative research that will ensure both maximal speed and accuracy or will there have to be trade-offs? And if the latter should turn out to be the case, how should collaborative research proceed when speed is important and how should it proceed when accuracy is?

Given the multifarious ways in which, in actual scientific practice, scientists collaborate and interact, one should not expect a general answer to these and kindred questions. In any event, it is not our purpose to provide such answers. Rather, we will concentrate on the efficiency (or otherwise) of what is likely to be a

---

<sup>2</sup> See Oddie (2008) for an up-to-date overview of the various proposals for defining the notion of truth-closeness that have been made in the literature.

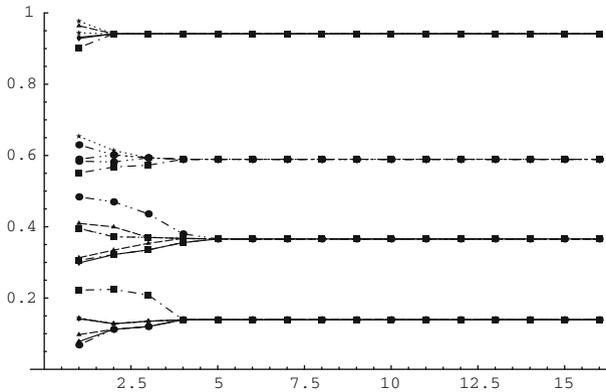
<sup>3</sup> See, for instance, Kuipers (1984, 2000, Chaps. 6 and 7).

consequence of collaboration and interaction among scientists, to wit, that a scientist's belief on a given issue tends to be influenced to at least some extent by his or her colleagues' beliefs on the same issue. While this kind of mutual influencing may be inevitable, it is not a priori that it helps to attain the scientific goal of truth approximation. Perhaps this goal would be served better by scientists' going strictly by the data and ignoring altogether their colleagues' opinions on the matter of interest. Or perhaps in the end it makes no difference whether or not there is this mutual influencing in scientific communities. The first part of the paper surveys recent work in social epistemology that bears on the question of what contribution to the achievement of the designated goal (if any) is made by the fact that scientists frequently allow their beliefs to be affected by those of their colleagues, or at least by those of some of their colleagues. Even though this is a narrowly focused question, we nevertheless hope that it will bring into relief the relevance of social epistemology to the topic of truth approximation.

As said above, we believe that the converse holds as well, in that issues relating to truth approximation are immediately relevant to one of the currently central debates in social epistemology, namely, the debate concerning the possibility of rational sustained disagreement among peers. So far, this issue has been mainly studied from a static or synchronic perspective: only single cases of disagreement have been considered, and authors have queried what the best response in those cases would be. We take instead a more dynamic or diachronic perspective on the matter by considering *strategies* for responding to disagreements and querying how these strategies fare, in terms of truth approximation, in the longer run, when they are applied over and over again to respond to arising cases of disagreement. This approach is reasonable: If, as most participants to the debate on peer disagreement hold, there is an a priori answer to the question of whether or not it is rational to stick to one's opinion in the face of disagreement with a person one regards as one's peer, the answer will imply how one should respond to an overt case of peer disagreement *whenever* it arises. In other words, the answer will imply an epistemic policy, and it makes sense to assess such policies in terms of their conduciveness to the scientific goal of approximating the truth.

## 1 Truth Approximation in Models of Opinion Dynamics

In the past decade, social epistemologists but also computer scientists and physicists have developed models to study the epistemic behavior of communities of agents who update their beliefs at least partly as a function of the beliefs of some or all other agents in the community. Questions that have been asked about such communities concern the conditions under which the beliefs of the initially disagreeing agents converge or tend to converge, partly or fully, as well as those under which the beliefs tend to polarize. Some of these models postulate a truth which the agents can try to determine via evidence they receive. In those models, belief updates are functions either of that evidence alone or of that evidence in combination with the beliefs of other agents. Here the main questions asked all concern the conditions under which the agents' beliefs converge, or can be expected



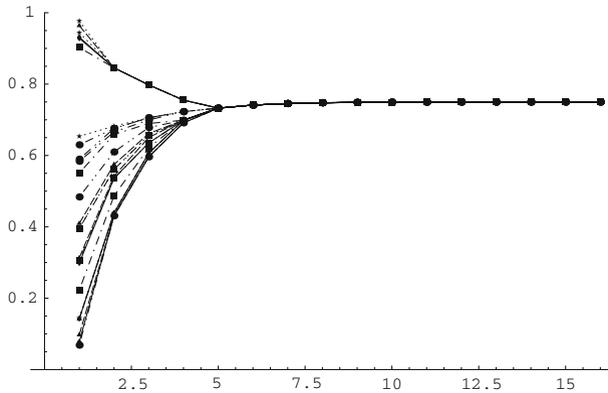
**Fig. 1** Repeated updating by averaging

to converge, to the truth. While some analytical results have been obtained in this field, by far most results come from computer simulations. Models for studying opinion dynamics in this fashion have been developed by a number of researchers, including Deffuant et al. (2000), Weisbuch et al. (2002), and Ramirez-Cano and Pitt (2006). But the arguably most elegant and in any case best-known model of this type was developed by Hegselmann and Krause in a series of seminal papers.<sup>4</sup> In the following, we begin by describing the original Hegselmann–Krause (HK) model and then go on to survey some recent extensions of it, which aim to make the model more realistic in a number of respects. These extended models will be seen to offer at least a beginning for studying the effect that various forms of socio-epistemic interaction among scientists may have on how well they will do with respect to the scientific goal mentioned earlier.

All versions of the HK model to be considered here study the opinion dynamics of communities of agents who are individually trying to determine the value of a certain parameter, where the agents know antecedently that the true value lies in the half-open interval  $(0, 1]$ . In the simplest version of the model, the agents simultaneously update their beliefs as to the value of the parameter by averaging over the beliefs of those other agents in the community whose beliefs are within a distance of  $\varepsilon$  from their own, for some given  $\varepsilon \in [0, 1]$  (henceforth to be called “neighbors”; note that every agent counts as its own neighbor). To illustrate, Fig. 1 shows the evolution of the beliefs of twenty-five agents who, starting from different initial beliefs, repeatedly update in the said way, and where  $\varepsilon = .1$ . What we see are diverging converging groups.

But most of Hegselmann and Krause’s work concerns a more interesting version of their model. In this version, the agents update their beliefs not just on the basis of the beliefs of their neighbors, but also on the basis of direct information—say, experimental evidence—they acquire about the value of the parameter they are trying to determine. Specifically, they update their beliefs by taking a weighted average of, on the one hand, the (straight) average of the beliefs of their neighbors

<sup>4</sup> See Hegselmann and Krause (2002, 2005, 2006, 2009).



**Fig. 2** Repeated updating with evidential input

and, on the other hand, of the reported value of the targeted parameter. To put this still more formally, let  $\tau \in \langle 0,1 \rangle$  be the value of the parameter,  $x_i(t)$  the opinion of agent  $x_i$  at time  $t$ , and  $\alpha \in [0,1]$  the weighting factor. Further let  $X_i(t)$  be the set of agent  $x_i$ 's neighbors at  $t$ —that is,  $X_i(t) := \{j : |x_i(t) - x_j(t)| \leq \varepsilon\}$ —and let  $|X_i(t)|$  be the cardinality of that set. Then the belief of agent  $x_i$  at time  $t + 1$  is given by the following equation:

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

For purposes of illustration, we set  $\tau = .75$ ,  $\alpha = .5$ , and  $\varepsilon = .1$ . Then if the agents considered in Fig. 1 updated their beliefs by (1), these would evolve over time as shown in Fig. 2. We now see convergence of all opinions.<sup>5</sup>

To clarify the idea of evidential input, we note that it is not an assumption of this model that the agents *know* the value of the parameter. Rather, the idea is that the agents get information from the world which gives some indication of the value of that parameter. Exactly how the agents evaluate this information and bring it to bear on their beliefs is left implicit in the model, except that the impact on each agent's new belief is supposed to be captured by Eq. 1.<sup>6</sup>

Hegselmann and Krause have investigated almost the complete parameter space of the above model by means of computer simulations; they also offer some analytical results. Though the model is still rather simple, they obtain a batch of

<sup>5</sup> Rainer Hegselmann brought to our attention that the dynamics of the version of the HK model considered here is, abstractly put, a convex combination of a social component and an objective component, and that the social component need not be defined in terms of averaging over the beliefs of neighbors as understood here. One alternative he suggested uses the repeated weighted averaging as proposed by Lehrer and Wagner (1981). Questions regarding conduciveness to the scientific goal could be asked about this and any other instantiation of the just-mentioned abstract dynamics, as well as about similar generalizations of the extensions of the HK model to be discussed below. This would give rise to a large and very general research program on socio-epistemic policies.

<sup>6</sup> For more on the interpretation of this aspect of the HK model, see Hegselmann and Krause (2006, Sect. 1).

interesting results, relating, for instance, speed of convergence to the value of  $\varepsilon$  and also to the value of  $\tau$ . Furthermore, in some of their computer experiments, they allow different agents to have different values for  $\alpha$ . These show that it is not a necessary condition for convergence to  $\tau$  that all agents have access to direct information concerning  $\tau$ . In fact, Hegselmann and Krause show analytically that, under certain (rather weak) circumstances, it is enough that *one* agent has a value for  $\alpha$  strictly less than 1 in order for *all* agents to eventually end up at the truth.<sup>7</sup>

Interesting though these and further results are, it is not hard to see that the model has some important limitations. For one, the agents are supposed to always receive *entirely accurate* information concerning  $\tau$ . For another, in the updating process the agents' beliefs all weigh equally heavily. Douven (2010) and Douven and Riegler (2010) argue that neither assumption is particularly realistic. Scientists have to live with measurement errors and other factors that make the information they receive noisy, and it is an undeniable fact of both everyday and scientific life that the beliefs of some people are given more weight than those of others. To remove the first limitation, Douven (2010) proposed to replace (1) by an update rule which permits the receipt of information about  $\tau$  that is "off" a bit, meaning that it does not necessarily point in the direction of precisely  $\tau$  but possibly only to some value close to it. To remove the second limitation, Douven and Riegler (2010) proposed a further variant of (1) which allows the assignment of weights to the agents, where each individual agent may receive a different weight. Taken together, these proposals amount to an extension of the HK model which replaces (1) by this update rule:

$$x_i(t+1) = \alpha \frac{\sum_{j \in X_i(t)} x_j(t) w_j}{\sum_{j \in X_i(t)} w_j} + (1 - \alpha)(\tau + \text{rnd}(\zeta)). \quad (2)$$

In this formula,  $w_j \geq 0$  is the weight assigned to agent  $j$ , and  $\text{rnd}(\zeta)$  is a function which outputs a unique uniformly distributed random real number in the interval  $[-\zeta, +\zeta]$ , with  $\zeta \in [0, 1]$ , each time the function is invoked.<sup>8</sup>

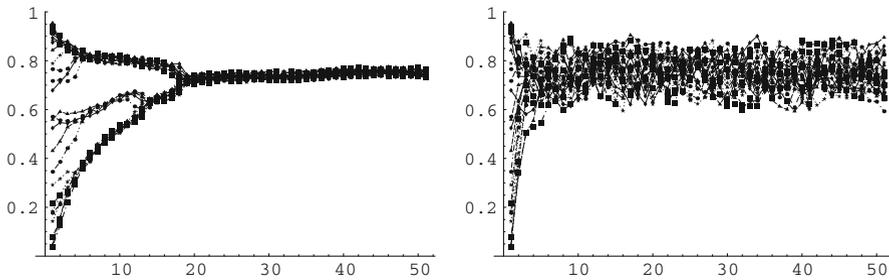
There were two main results to report about this model. The first is that communities of agents who may obtain imprecise information about  $\tau$  end up on average being closer to  $\tau$  for higher values of both  $\alpha$  and  $\varepsilon$ . On the other hand, with lower values for these parameters, they on average tend to get faster to a value that is at least moderately close to  $\tau$ . For an illustration, see the graphs in Fig. 3, which show the developments of the beliefs of twenty-five agents who receive imprecise information about  $\tau$ , but who in the left and right graph give different weight to the beliefs of their neighbors each time they update. Simulations showed the result depicted here to be entirely typical.<sup>9</sup>

The second main result is that, contrary to what one might initially expect, varying the assignments of reputations has neither a significant influence on average

<sup>7</sup> This is a consequence of their Leading-the-pack Theorem; see, for instance, Hegselmann and Krause (2009, 137).

<sup>8</sup> One could consider a further extension of this model by allowing each agent to assign a weight to every agent. In the current model, each agent has a fixed weight that is assigned to it by all agents.

<sup>9</sup> Here, distance from  $\tau$  is measured in the most straightforward way: the distance of agent  $x_i$ 's belief from  $\tau$  at time  $t$  equals  $|\tau - x_i(t)|$ .



**Fig. 3** Updating with imprecise information ( $\zeta = .2$ ); *left*:  $\alpha = .9$ ,  $\varepsilon = .1$ ; *right*:  $\alpha = .5$ ,  $\varepsilon = 0$

speed of convergence to  $\tau$  nor on average accuracy of convergence. Presumably, this just goes to show that (2) is still too poor as a model to simulate realistically the effects of reputation in opinion dynamics. As of this writing, Rainer Hegselmann is completing a new, much richer model—the Epistemic Network Simulator, as he has dubbed it—which allows agents to move about in a two-dimensional environment and to associate with, as well as dissociate from, others on the basis of a number of criteria, including reputation. More significant results concerning the role of reputation in opinion dynamics, in particular with respect to speed and accuracy of convergence to the truth, are to be expected from experiments that are scheduled to be conducted in this model.

Zollman (2007, 2010) presents another model for network simulations, this one operating on the basis of a Bayesian update mechanism. Zollman's intriguing work shows that, from a socio-epistemic perspective, it may be important to maintain epistemic diversity in a community of agents, at least for a while, and that, for that reason, it is not always best if all agents have access to all information available in their community; for the same reason, a certain dogmatism on the part of the agents—understood in terms of their prior probability distributions—may be beneficial. In particular, it is shown that well-connected networks of relatively undogmatic agents are more vulnerable to the receipt of misleading evidence than less well-connected networks or networks with more dogmatic agents; in the former, misleading evidence spreads rapidly through the community, and is thereby more likely to have a lasting effect on their convergence behavior. This suggests that performance of networks in terms of their capacity to approximate the truth can be improved either by ensuring that different agents have access to different parts of the information that is somewhere available in the network or by endowing the agents with extreme priors, thereby building in a kind of dogmatism. As Zollman's (2010) simulations also show, however, doing *both* of the foregoing will typically make the network perform quite *poorly*, in as much as dogmatic agents with limited access to the available information are likely to stick with their initial extreme views, and so likely not to converge to the truth.<sup>10</sup>

<sup>10</sup> Olsson (2011) describes yet another simulation environment (which was developed by Staffan Angere in collaboration with Erik Olsson). This one is still richer and more flexible than the one used by Zollman. However, results of systematic simulations within this environment are still to be reported.

Finally, Riegler and Douven (2009) present a model that, while clearly in the spirit of the HK model—it still involves interacting truth-seeking agents who receive information about the truth—departs from the HK model in that the agents' belief states no longer concern a single parameter but are characterizable as *theories*. The theories are axiomatizable in a finitary propositional language, so that the agents' belief states can be semantically represented as finite sets of possible worlds and hence also—assuming some ordering of the worlds—as finite strings of 0's and 1's. Agents' being neighbors is then defined in terms of a metric on binary strings known in the literature as the *Hamming distance*, which is defined as the number of digits in which the strings differ. Specifically, an agent *i*'s neighbors are said to be those agents *j* such that *i* and *j*'s belief states are within a given Hamming distance from one another. Just as in the HK model the agents are trying to determine the value of a parameter, in the current model they are trying to determine the *truth*, which is some theory in their language. They again do this by repeatedly updating their belief states both on the basis of their neighbors' belief states and on the basis of incoming information about the truth (which comes in the form of theories entailed by the truth). Also as in the HK model, the updating proceeds by a specific way of averaging over the neighbors' belief states together with the evidence.

In simulations carried out in this model, both the value of the minimal Hamming distance required for inclusion in the neighborhood and the strength (in the logical sense of the word) of the information about the truth the agents received at each update were varied. The main conclusion that could be drawn from these simulations was qualitatively similar to one of the conclusions of the simulations carried out in the extended version of the HK model that assumes (2) as an update rule, to wit, that being open to interaction with other agents helps agents to approximate the truth more closely, but it also slows down convergence in that it takes longer to get within a moderately close distance from the truth as compared to agents that ignore the belief states of other agents and go strictly by the information about the truth.<sup>11</sup>

This is a far from complete survey of what has been going on in the area of opinion dynamics in the past ten years or so. Nevertheless, we hope it is enough to show that the area of research is of immediate interest to the topic of truth approximation. An exclusive focus on the individual scientist, and which confirmation-theoretic principles he or she might deploy in order to approximate the truth as expeditiously and/or accurately as he or she can, is too narrow. To complete the picture, we must attend to communication practices, ways of sharing information, cooperation networks, and other features of the social architecture. For the studies discussed above give ample reason to believe that how accurate and how fast, on average, a scientist will approximate the truth depends on these features, too.

---

<sup>11</sup> In the simulations, the distance from the truth of a given belief state was simply taken to be the Hamming distance between that belief state and the true theory.

## 2 The Rationality of Peer Disagreement and Truth Approximation

While it is a fairly uncontroversial descriptive fact that people disagree in a variety of areas of thought, there is an interesting normative question concerning the epistemic status of such disagreements. More specifically, the question has been raised whether there can ever be rational disagreements among agents who are and/or take each other to be epistemic peers on a certain proposition. To say that a number of agents are epistemic peers concerning a question,  $Q$ , is to say, first, that they are equally well positioned evidentially with respect to  $Q$ , second, that they have considered  $Q$  with equal care, and, third, that they are equally gifted with intellectual virtues, cognitive skills, and the like, at least insofar as these virtues, skills, and so on, are relevant to their capacity of answering  $Q$ .<sup>12</sup> For epistemic peers to rationally disagree on  $Q$  is for them to justifiably hold different doxastic attitudes concerning the issue of what is the correct answer to  $Q$ . Finally, of special interest in the debate on peer disagreement, and also the question we will focus on, is whether or not epistemic peers can *sustain* a rational disagreement after full disclosure, that is, once they have shared all their relevant evidence as well as the respective doxastic attitudes arrived at.

A number of philosophers have recently argued that it is impossible for peers to rationally sustain a disagreement, at least after full disclosure. Suppose PRO and CON are both leading experts in a given field. While they both take each other to be peers on a certain question,  $Q$ , they disagree on what the correct answer is; PRO thinks it is  $P$ , CON thinks it is the negation of  $P$ . There is a powerful intuition that the rational thing to do here is to “split the difference,” as it has come to be called, which at least in the kind of case where peers hold contradictory beliefs is commonly taken to amount to suspending judgment on the issue of what the correct answer is to  $Q$ . This intuition is backed by the observation that if the parties to the disagreement were entitled to hold on to their respective beliefs, they would also be entitled to discount their opponent’s opinion simply on the grounds that a disagreement has occurred. Certainly, however, there exists no such entitlement. One cannot discount the opinion of someone one takes to be a peer on some question simply on the basis that a disagreement has occurred.<sup>13</sup>

In some cases in which peers hold contrary and not just contradictory beliefs on a given matter, “splitting the difference” has another natural interpretation, besides “suspension of belief.” Suppose, for example, that OPT holds that recovery from the current (2010) financial crisis will begin to occur in 2012 whereas PES holds that it will not begin before 2016. Then one way in which they could settle their disagreement is by quite literally splitting the difference and adopting, both of them, the belief that the recovery from the financial crisis will begin to occur in 2014.

<sup>12</sup> This characterization of “epistemic peer” can be found in nearly all recent contributions to the debate. See, e.g., Kelly (2005), Feldman (2006, 2007), Christensen (2007), Douven (2009, 2010), and Goldman (2010).

<sup>13</sup> This, we take it, is the core point of Elga’s (2007, Sect. 10) bootstrapping argument. Proponents of the view that peers cannot rationally sustain disagreements further include Feldman (2006, 2007) and Christensen (2007).

Lehrer and Wagner's early work on opinion dynamics can in effect be interpreted as an argument in support of this way of resolving disagreements among peers.<sup>14</sup>

Now, while we did not use the terminology "splitting the difference" in the previous section, it will be clear that, as a matter of fact, the agents in the HK models and its variants do exactly that: they split the difference (in the second sense mentioned above) with their "neighbors" (as we called them), that is, with other agents whose relevant beliefs are sufficiently close to their own. We are free, of course, to stipulate that an agent's neighbors at a given time are precisely those agents he or she regards as his or her peers at that time. And, as is argued in Douven (2010), if we do make that stipulation, then the results concerning truth approximation that were obtained in the said models can be swiftly brought to bear on the debate about peer disagreement.<sup>15</sup>

The central claim of that paper is that there is no a priori answer to the question how peers ought to respond to the discovery that they disagree: in some contexts, the rational thing to do for them may be to split the difference, in others, it may be to ignore each other's belief and purely go by whichever evidence they receive from the world. The claim is largely argued for on the basis of results concerning the variant of the HK model which assumes (2) as an update rule. As we said, and as can be readily gleaned from Fig. 3, if the information the agents receive is "noisy" (which as a rule it is in scientific practice), then if they adhere to the practice of splitting the difference, they will tend to converge to the truth more accurately, albeit that in this case convergence to a value at least moderately close to the truth takes, on average, longer than if they go strictly by the (noisy) information. The claim then follows from the fact that there is no general answer to the question which is better: more accurate convergence, even if this takes a while, or more rapid convergence which, however, is less accurate, even in the long run.<sup>16</sup>

That in the peer disagreement debate considerations related to truth approximation have been neglected so far is, we believe, because authors are strongly inclined to concentrate on single-case resolutions of disagreements. However, if the answer to the question how peers should respond to the discovery that they disagree on some issue is a priori, as virtually all participants to the debate have claimed, then presumably peers ought to respond in a uniform way in each particular case of disagreement. And that makes it legitimate to compare *practices* of resolving disagreements. In particular, we may ask how these practices fare in the light of the goal of approximating the truth. If we do so, we said, the results from the computer simulations described in Sect. 1 suggest that no one practice may be *generally* best: splitting the difference, on the one hand, and ignoring one's peers beliefs and going purely by the evidence, on the other hand, can both be said to serve our epistemic goal, even if they do so in different ways.

<sup>14</sup> See Lehrer (1976, 1980), Wagner (1978), and Lehrer and Wagner (1981).

<sup>15</sup> For an argument that these same results also have some bearing on the issue of whether peers should split the difference in the sense of Christensen or Feldman, see Douven (2010, Sect. 3).

<sup>16</sup> That is to say, the claim follows given the assumption that what is most rational is what is most conducive to our epistemic goal, and the further assumption that our epistemic goal is to be understood in terms of truth approximation; see Douven (2010, Sect. 3) for a detailed exposition of the argument.

We briefly want to point to some further considerations related to truth approximation that are pertinent to the peer disagreement debate, considerations which also suggest that under certain circumstances it would be wrong to adopt splitting the difference as one's practice for settling disagreements with peers. Historians of science have argued that scientists have a tendency to disregard evidence which sits badly or even goes against their preferred view. They may simply fail to see the evidence,<sup>17</sup> or they may reason that the evidence must be misleading, possibly due to some measurement error.<sup>18</sup> Even if this may not be what ideal scientists would do, given that it is what actual ones do, a certain diversity in the views held by scientists would seem to increase the chances of scientific progress by decreasing the chance that some—potentially crucial—evidence is missed by *all* groups of researchers. This is buttressed by the results of Zollman's (2010) simulations described in the previous section. Of course, as Zollman also notes, we do not want scientists' beliefs to diverge permanently; the hope is that, in the end, the scientific community as a whole *will* converge on the truth. Nevertheless, Zollman argues, we should aim for *transient* diversity. And, patently, even that would be prohibited if, on the discovery of peer disagreement, scientists were to suspend judgment on their theories, or were to adopt some "middle" position between their and their opponents' theories.<sup>19</sup>

We thus hope to have shown that observations concerning a community's capacity to approximate the truth are relevant to the debate about peer disagreement. Our main observation concerned the earlier reported results from computer simulations, showing that there can be a trade-off between accuracy and speed of convergence if the information that agents have access to and also take into account may be imprecise. An additional observation concerned situations in which, even though agents may all have access to the same information, they are prone to overlook, or perhaps even willingly ignore, information that is irreconcilable with

<sup>17</sup> As Hanson (1958) famously argued; see also Kuhn (1962/1996, Chaps. 6 and 10).

<sup>18</sup> Zollman (2010, Sect. 1) describes a nice case in point. It is currently accepted that peptic ulcer disease is caused by a bacterial infection. However, for a long time, evidence suggesting as much was simply dismissed, due to the fact that medical researchers had been largely convinced by a single, and as later turned out poorly conducted, study that seemed to indicate that no bacteria were to be found in the stomachs of patients suffering from the disease.

<sup>19</sup> Those who hold that peers cannot rationally sustain a disagreement might retort that suspension of judgment may have the effect of opening one's eyes to *all* evidence; after all, there is no longer evidence seemingly going against one's theory (for one will not regard as one's theory a theory that one is agnostic about). However, this may be to miss entirely the point of Hanson (1958) and others that there is no such thing as theory-neutral observation. For if that point is correct, then it is by no means obvious that suspension of judgment will open one's eyes to all evidence. To the contrary, suspension of judgment might then make it hard to interpret *any* evidence. Moreover, sometimes it is a *good thing* to ignore evidence, for instance, because the evidence is indeed misleading, or because, even though correct, the theory one is working on, and which is in conflict with the evidence, is, while false, still the one that is closest to the truth. Ignoring the evidence might then help to keep one motivated and do further work on the theory, perhaps with the result that, eventually, one comes up with a refined version of the theory that is no longer in conflict with the evidence. In connection to this, see Lakatos' (1978, 53–55) discussion of Prout's programme concerning the atomic weights of pure chemical elements. Prout and others kept working on this programme, knowing that it was contradicted by a great number of observations. In trying to explain away the observations as being due to experimental error, they developed important new experimental techniques.

their favored theory or theories. In such situations, a certain tenacity in one's holding onto a belief in the face of disagreeing peers may be more conducive to the goal of approximating the truth than the willingness to split the difference that Christensen, Feldman, and others hold to be appropriate here.

But what, then, remains of the powerful intuition that some epistemic response *is* called for when we discover that we disagree with a peer? The intuition is correct, we think; it just is to be noted that such a response can take many forms. What is correct is that it would be epistemically irresponsible behavior to simply pass by the discovery of disagreement with a peer. One way to take into account such disagreement is to suspend judgment on the disputed proposition, or to split the difference in the more literal sense discussed above. Another way is to scrutinize one's evidence for the said proposition a second time, carefully reconsider the argument or arguments that led one from the evidence to the proposition, checking the various steps involved in this argument (these arguments), and do the same with one's opponent's argument (arguments) in as open-minded a manner as possible. The latter type of response, we submit, is mandatory; the former, as we argued, is not, and must in some circumstances even be advised against.

**Acknowledgments** We are greatly indebted to Rainer Hegselmann and an anonymous referee for very valuable comments on a previous version of this paper.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Non-commercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Christensen, D. (2007). Epistemology of disagreement: The good news. *Philosophical Review*, *116*, 187–217.
- Deffuant, G., Neau, D., Amblard, F., & Weisbuch, G. (2000). Mixing beliefs among interacting agents. *Advances in Complex Systems*, *3*, 87–98.
- Douven, I. (2009). Uniqueness revisited. *American Philosophical Quarterly*, *46*, 347–361
- Douven, I. (2010). Simulating peer disagreements. *Studies in History and Philosophy of Science*, *41*, 148–157.
- Douven, I., & Riegler, A. (2010). Extending the Hegselmann–Krause Model I. *Logic Journal of the IGPL*, *18*, 323–335.
- Elga, A. (2007). Reflection and disagreement. *Noués*, *41*, 478–502.
- Feldman, R. (2006). Epistemological puzzles about disagreement. In S. Hetherington (Ed.), *Epistemology futures* (pp. 216–236). Oxford: Oxford University Press.
- Feldman, R. (2007). Reasonable religious disagreements. In L. Antony (Ed.), *Philosophers without gods: Meditations on atheism and the secular life* (pp. 194–214). Oxford: Oxford University Press.
- Goldman, A. I. (1999). *Knowledge in a social world*. Oxford: Oxford University Press.
- Goldman, A. I. (2010). Epistemic relativism and reasonable disagreement. In R. Feldman, & T. Warfield (Eds.), *Disagreement*. Oxford: Oxford University Press (available now).
- Hanson, N. R. (1958). *Patterns of discovery*. Cambridge: Cambridge University Press.
- Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence: Models, analysis, and simulations. *Journal of Artificial Societies and Social Simulation*, *5* (available at <http://jasss.soc.surrey.ac.uk/5/3/2.html>).
- Hegselmann, R., & Krause, U. (2005). Opinion dynamics driven by various ways of averaging. *Computational Economics*, *25*, 381–405.

- Hegselmann, R., & Krause, U. (2006). Truth and cognitive division of labor: First steps towards a computer aided social epistemology. *Journal of Artificial Societies and Social Simulation*, 9 (available at <http://jasss.soc.surrey.ac.uk/9/3/10.html>).
- Hegselmann, R., & Krause, U. (2009). Deliberative exchange, truth, and the cognitive division of labour: A low-resolution modeling approach. *Episteme*, 6, 130–144.
- Kelly, T. (2005). The epistemic significance of disagreement. In J. Hawthorne, & T. Szabó Gendler (Eds.), *Oxford studies in epistemology* (Vol. 1, pp. 167–197).
- Kuhn, T. (1962/1996). *The structure of scientific revolutions* (3rd ed.). Chicago: Chicago University Press.
- Kuipers, T. (1984). Approaching the truth with the rule of success. *Philosophia Naturalis*, 21, 244–253.
- Kuipers, T. (2000). *From instrumentalism to constructive realism*. Dordrecht: Kluwer.
- Lakatos, I. (1978). The methodology of scientific research programmes. In J. Worrall, G. Currie (Eds.) *Philosophical papers* (Vol. I). Cambridge: Cambridge University Press.
- Lehrer, K. (1976). When rational disagreement is impossible. *Nous*, 10, 327–332.
- Lehrer, K. (1980). A model of rational consensus in science. In R. Hilpinen (Ed.), *Rationality in science* (pp. 51–62). Dordrecht: Reidel.
- Lehrer, K., & Wagner, C. (1981). *Rational consensus in science and society*. Dordrecht: Reidel.
- Oddie, G. (2008). Truthlikeness. In S. Psillos, & M. Curd (Eds.), *The Routledge companion to philosophy of science* (pp. 478–488). London: Routledge.
- Olsson, E. J. (2011). A simulation approach to veritistic social epistemology. *Episteme* (in press).
- Ramirez-Cano, D., & Pitt, J. (2006). Follow the leader: Profiling agents in an opinion formation model of dynamic confidence and individual mind-sets. *Proceedings of the 2006 IEEE/WIC/ACM international conference on intelligent agent technology*, pp. 660–667.
- Riegler, A., & Douven, I. (2009). Extending the Hegselmann–Krause Model III: From single beliefs to complex belief states. *Episteme*, 6, 145–163.
- Wagner, C. (1978). Consensus through respect. *Philosophical Studies*, 34, 335–349.
- Weisbuch, G., Deffuant, G., Amblard, F., & Nadal, J. P. (2002). Meet, discuss and segregate! *Complexity*, 7, 55–63.
- Zollman, K. (2007). The communication structure of epistemic communities. *Philosophy of Science*, 74, 574–587.
- Zollman, K. (2010). The epistemic benefit of transient diversity. *Erkenntnis*, 72, 17–35.