Rainer Hegselmann and
Ulrich Krause

Deliberative Exchange, Truth, and Cognitive Division of Labour: A Low-Resolution Modeling Approach

ABSTRACT

This paper develops a formal framework to model a process in which the formation of individual opinions is embedded in a deliberative exchange with others. The paper opts for a *low-resolution modeling approach* and abstracts away from most of the details of the social-epistemic process. Taking a bird's eye view allows us to analyze the chances for the truth to be found and broadly accepted under conditions of cognitive division of labour combined with a social exchange process. *Cognitive division of labour* means that only some individuals are active truth seekers, possibly with different capacities. Both mathematical tools and computer simulations are used to investigate the model. As an analytical result, the *Funnel Theorem* states that under rather weak conditions on the social process, a consensus on the truth will be reached if all individuals possess an arbitrarily small capacity to go for the truth. The *Leading the pack Theorem* states that under certain conditions even a single truth seeker may lead all individuals to the truth. Systematic simulations analyze how close agents can get to the truth depending upon the frequency of truth seekers, their capacities as truth seekers, the position of the truth (more to the extreme or more in the centre of an opinion space), and the willingness to take into account the opinions of others when exchanging and updating opinions.

1. INTRODUCTION

The idea that we gain knowledge through an interwoven individual and social epistemic process has a long history. It is already present in Aristotle's *Topics*; and it is found explicitly in Kant and Mill, both of whom believed that we acquire knowledge by a process of individual reasoning *and* deliberative exchange with others.[1] *Social epistemology* is the recently founded discipline that develops this old insight.

As yet, we do not have a *well elaborated, systematic, detailed, and unified theory* of the process by which the formation of individual judgments is embedded

 DOI: 10.3366/E1742360009000604

in a deliberative exchange with others. What we have instead is a bundle of contributions, some of which are formal and some informal; and some are concerned with global processes and some with special details of the process. Typically, if the concern is more global, then the analysis is either informal or, if formal, only at a programmatic stage.[2] If the analysis is formal (being based on formal logic, probability calculus, and Bayesianism) *and* more than only programmatic, then the concern is typically very specific. The challenge now is the formal analysis of the global process. This is a formidable task because by definition formal modeling requires a high degree of precision and explicitness and some may believe that it is impossible.

Nevertheless, in what follows we want to demonstrate that it is indeed possible to formally model the globally interwoven individual and social-epistemic process. The trick is to opt for a *low-resolution modeling approach* and abstract away from most of the details by taking a bird's eye view of the individual and social-epistemic processes. As in some other approaches to social epistemology (cf. Goldman 1999, List 2005), truth seeking *and* aggregating of individual opinions is the central focus of the following analysis. We offer an integrated view.

2. LOW-RESOLUTION MODELS OF DELIBERATIVE EXCHANGE

Consider a group of knowledge seeking agents, say, a group of experts in a field. Each group member has an opinion on the topic under discussion, for instance, the probability of a certain type of accident. Nobody is fully sure that he is totally right. To some degree, everybody is willing to revise his opinion when informed about the opinions of others, especially the opinions of 'competent' others. The revisions generate a new distribution of opinions, further revisions, and so on and so on. How to model such a process?

As a *first step,* we introduce some notation: Let $I = \{1, 2, \ldots, n\}$ be the set of n agents in the group under consideration. We think of *time* as an infinite sequence of periods $t = 0, 1, 2, \ldots$. *Opinions* are represented as real valued numbers from the interval $[0,1]$. The opinion of agent i in period t is denoted as $x_i(t)$. The *profile* of all opinions at time t is given by

$$x(t) = (x_1(t), x_2(t), \ldots, x_i(t), \ldots, x_n(t)).$$

(Note: It is for *convenience* only that we restrict ourselves to a one-dimensional opinion space and the unit interval therein. Both restrictions are *not* inherent to the approach!)

The decisive *second step* is to characterize the *social process* that generates for all agents i their updated opinion $x_i(t+1)$ by a deliberative exchange in period t. At this point, a *radical* move is made: We do not even try to model explicitly the processes and actions of deliberative exchange (questions, answers, speech acts of all sorts, clarifications, inferences, consistency checks, weighing pieces of evidence,

https://doi.org/10.3366/E1742360009000604 Published online by Cambridge University Press

reorganizing beliefs, etc.). Instead, we assume for each agent i a function f_i that simply 'delivers' the updated opinion $x_i(t+1)$ based on $x(t)$, i.e. the opinion profile at time t. By that step, the processes and the results of deliberative exchange is compressed into the functions f_i, the *social process function* as we will call it. f_i will be specified later.

This type of approach to model opinion dynamics is *not new*. It has a history of more than 50 years (see Hegselmann and Krause 2002, sect. 1–3). There are *two* social process functions that are especially interesting from the philosophical point of view that guide us.

(a) The Lehrer/Wagner-model

In 1981 Lehrer and Wagner published the book *Rational Consensus in Science and Society* (cf. Lehrer 1975, 1976, 1977, 1981a, and 1981b). The book received major attention, especially among philosophers (cf. Loewer 1985; Bogdan 1981). It presents a mechanism that *can be interpreted* as a social process driven by *iterated weighted averaging*. The weights reflect the *respect* agents assign to other agents. The social process is given by

$$x_i(t+1) = f_i^{LW}(x(t)) \tag{1}$$

where

$$f_i^{LW}(x) = w_{i1}x_1 + w_{i2}x_2 + \ldots + w_{in}x_n \tag{2}$$

The weights w_{ij} are *fixed* nonnegative values adding up to 1. We will refer to (1) as the *Lehrer/Wagner-model* (*LW*-model). Formally it is a *dynamical* system, though – and this is important to notice – Lehrer and Wagner *do not interpret it as a process over time*. Their *starting* point is a "*dialectical equilibrium*", i.e. a situation *after* "the group has engaged in extended discussion of the issue so that all empirical data and theoretical ratiocination has been communicated. … the discussion has sufficiently exhausted the scientific information available so that further discussion would not change the opinion of any member of the group" (Lehrer and Wagner 1981, 19). The central question for Lehrer and Wagner then is: Once the dialectical equilibrium is reached, is there a *rational* procedure to *aggregate* the normally still divergent opinions in the group (cf. Lehrer 1981b, 229)? Their answer is "Yes". The basic idea for the procedure is to make use of the fact that normally we all not only have opinions but also judgements on the expertise or reliability of others. These judgements can be used to assign weights to other individuals. The whole aggregation procedure is then *iterated weighted* averaging with $t \to \infty$ and based on *constant* weights.[3] It is shown that for many weight matrices the individuals reach a consensus whatever the initial opinions might be – if they only were willing to apply the proposed aggregation procedure.[4] We interpret (1) in this article in a different way: We assume (1) to define a *social process of iterated deliberative exchange*.

(b) The Bounded-confidence-model

Lehrer and Wagner do not say very much on the question *"How to assign weights?"* The second model of deliberative exchange, however, does address this issue. In this model, agents only take seriously, i.e. assign positive weights to, opinions that are '*not too far away*' from their own opinion. More formally: Each individual i takes into account only those individuals j for which $|x_i(t)-x_j(t)| \leq \varepsilon$. We refer to ε as the *confidence level*. The set of all individuals that i takes seriously is

$$I(i, x(t)) = \{1 \leq j \leq n \,|\, |x_i(t) - x_j(t)| \leq \varepsilon\}. \tag{3}$$

Individuals *update* their opinions. The next period's opinion of individual i is the *average* opinion of all those that i takes seriously, i.e.:

$$x_i(t+1) = f_i^{BC}(x(t)), \tag{4}$$

where

$$f_i^{BC}(x) = |I(i,x)|^{-1} \sum_{j \in I(i,x)} x_j. \tag{5}$$

(Note: The expression "$|I(i,x)|$" is the *number* of elements of $I(i,x)$.) Thus the very essence of the model is '*averaging over all opinions within one's confidence interval*'. The model in (4) is called the *bounded confidence* model (BC-model)[5]. A detailed, thoroughgoing and rigorous analysis of that model is given in Hegselmann and Krause (2002). Extensions of all sorts are described in Hegselmann (2004) and Hegselmann and Krause (2005). Here are some of the main results: Exchange processes that follow the BC-model always stabilize in *finite* time. The final pattern of opinions crucially depends upon the size of the confidence interval ε: For a small ε, the final pattern consists of many different opinions (*plurality*). If ε is increased up to a certain size, then the agents tend to end up in two camps (*polarization*). Above a certain threshold, the final result is unanimity (*consensus*).

Both the LW and the BC models assign weights. But they do that in *different* ways: In the LW-model, weights are assigned *in*dependently of the distance to one's own opinion; in the BC-model, weights crucially *depend* upon that distance. In the LW-model, weights are *fixed* and remain constant over time; in the BC-model, weights *vary* over time according to

$$w_{ij}(x(t)) = |I(i, x(t))|^{-1} \text{ for } j \in I(i, x(t)) \text{ and } w_{ij}(x(t)) = 0 \quad \text{otherwise.} \tag{6}$$

Additionally, one can easily check that both models, i.e. the social processes according (1) and (4), share the following property (which will become important later in section 4): The *range* of opinions is *preserved over time* and never expands. Stated more formally:

$$\min\{x_j \mid j \in I\} \leq f_i(x) \leq \max\{x_j \mid j \in I\} \tag{7}$$

for all $i \in I$ and all possible opinion profiles x.[6]

3. MODELING TRUTH SEEKING AGENTS INVOLVED IN DELIBERATIVE EXCHANGE

There is a fundamental objection against (1) and (4) as models of deliberative exchange: *"The models forget about the most decisive point of deliberative exchange: the truth. At best, they can be applied to areas only where truth does not play any role."* Admittedly, so far, truth, justification, etc. don't play any explicit role in the processes. But we can extend and modify the models in a way that they cover a process in which more or less successful truth seeking agents go for the truth, while at the same time, the agents are engaged in a process of deliberative exchange.

To realize that, a *second* radical abstraction is made: We stick to a macroscopic approach, follow the KISS heuristic, "*Keep it simple, stupid!*", and do not even try to model explicit actions in the quest for truth. Instead, we cover all that might be important by the following assumption: There is a true value T somewhere in the opinion space [0,1] and that value T somehow 'attracts' the agents – at least some and to some degree. The decisive equation is

$$x_i(t+1) = \alpha_i T + (1 - \alpha_i) f_i(x(t)) \tag{8}$$

We will refer to (8) as the *truth dynamics*. In equation (8), the opinion of individual i in period $t + 1$ is given by a convex combination with $0 \leq \alpha_i \leq 1$. The first part, $\alpha_i T$, is the *objective* component. α_i controls the *strength* of the attraction of truth. The second part, with the weight $(1 - \alpha_i)$, is the *social* component, with f_i being a social process as given by $f_i^{LW}(x)$ or $f_i^{BC}(x)$. For $\alpha = 0$, for all i we get again our original dynamics governed by (1) or (4), i.e. a situation in which only a social process is at work and the truth does not play any explicit role. The framework given by (8) offers a fairly natural (though again very abstract) understanding of *cognitive division of labour*; by cognitive division of labour, we refer to a situation in which *only some* individuals $i \in I$ have an $\alpha_i > 0$.

Since the truth dynamics defined by (8) is threatened by *misunderstandings* of all sorts, some clarifying remarks on the interpretation and status of (8) will be helpful:

(a) *Could* (8) *be a mechanism intentionally followed by our individuals?* No! The simple reason is that whoever understands the concept of truth and somehow

https://doi.org/10.3366/E1742360009000604 Published online by Cambridge University Press

knows T would immediately believe that T, i.e. update $x_i(t + 1) = T$. The interpretation of (8) should be about the following: Individuals with a positive α have access to or generate new data (arguments, evidence, test results, etc.) that point in the direction of T. Nevertheless, their and their fellows' prior opinions are in their mental or cognitive background and influence them as well.

(b) *What is assumed regarding the nature of truth?* It is assumed that the truth is one and only one, that it does not change over time, that it is somewhere in the opinion space, and that it is independent of the opinions the individuals hold. The truth influences opinions, but the *opinions don't have any impact on the truth*. Ever since the beginning of philosophy, there has been a discussion on the nature of truth, its general concept, status, definition, criteria, indicators, etc. The commonsensical assumptions regarding truth made above are *not* compatible with *all* known conceptions of truth. They are compatible with a *correspondence theory of truth*, but so-called *deflationists* shouldn't have any problem accepting them as well.

(c) *How to make sense of an 'attraction of truth'?* For $\alpha_i > 0$, truth attracts in a *technical* sense. But this technical feature of the equation is used to generate a process that *must not* be interpreted in such a way that truth *attracts* anybody in a literal sense. The *technical* attraction of truth is used to model individuals that to a certain degree *successfully aim at the truth*. Thus, a positive α_i could be interpreted as the combined effect of education, training, profession, and interest, and some epistemic success is based on these qualities of the individual. A problem is that according to (8), the 'attraction of truth' works smoothly in the direction of T. But new evidence might be ambiguous. It may point in different directions or only indicate that the truth is *not* in a certain region of the opinion space. Therefore, (8) should be taken as a *starting point*. Later, we (or others) can focus on more complicated epistemic situations.

Under our interpretation, (8) provides a very general and simple formal framework for *the study of truth seeking agents embedded in a community and therein engaged in a process of deliberative exchange*.[7]

4. TWO ANALYTICAL RESULTS: THE *FUNNEL THEOREM* AND THE *LEADING THE PACK THEOREM*

First, we address the question of whether all agents will approach a *consensus on the truth* if all go for the truth, i.e. if $\alpha_i > 0$ for all $i \in I$.

Funnel Theorem (FT)

For the truth dynamics (8) with a range preserving social process that satisfies (7), the agents will approach a consensus on the truth, provided all agents go for the truth.

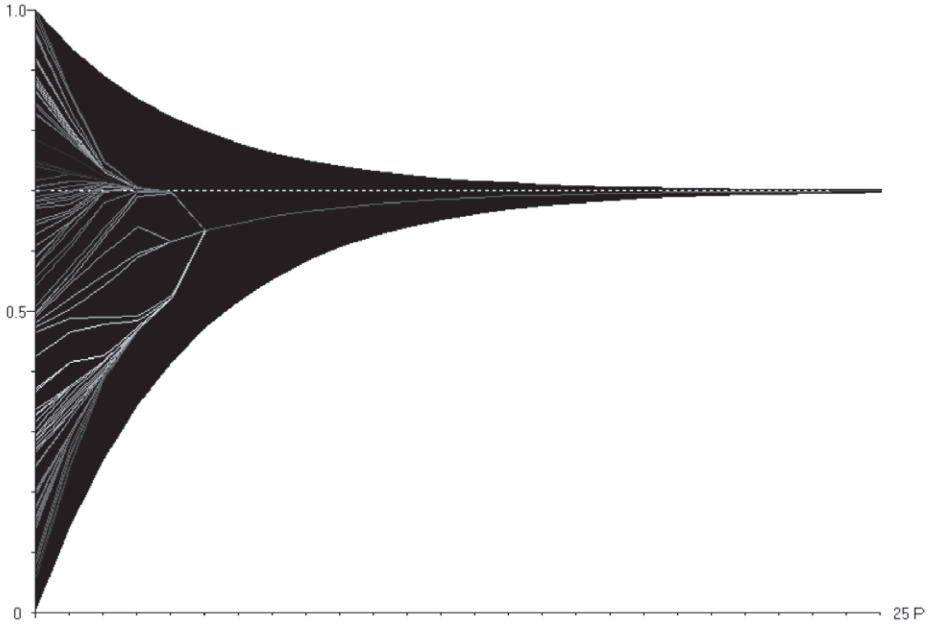

Figure 1. The Funnel Theorem at work: consensus on the truth.

For a proof, see Hegselmann and Krause (2006, sect. 3). The name of the theorem is due to the original proof strategy. In the *first* step, a 2-dimensional funnel is defined whose upper and lower 'walls' approach T for $t \to \infty$. In the *second* step, it is shown by induction that all opinions are caught in that funnel. The Funnel Theorem holds in particular for the social processes of the LW-models as well as for the BC-model. Actually, in the Funnel Theorem, the consensus on the truth is mainly due to the objective component in (8) – though normally the social process *shortens* the time that it takes to get to a consensus. Figure 1 shows a computer simulation where all agents have a $\alpha_i > 0$ and interact according to the BC-model. The dotted line depicts the value of T. The grey shaded curves within the black funnel show the agents' opinions converging to a consensus that approaches T.

Next, we address the question of what will happen if not all the agents go for the truth, including the extreme case where just one single agent goes for the truth. Now the social process plays a decisive role for a consensus on the truth. The crucial point is the extent to which agents are connected by chains of other agents such that each agent takes into account the opinion of the next one. More precisely, for the LW-model (2) *agent i is connected to agent j* if there exists a chain of agents i_1, $i_2, \ldots, i_k$ such that

$$w_{ii_1} > 0, w_{i_1 i_2} > 0, \ldots, w_{i_k j} > 0.$$

https://doi.org/10.3366/E1742360009000604 Published online by Cambridge University Press

The matrix **W** of the weights w_{ij} is called *irreducible* if any two agents are connected. Connectedness for the BC-model is a little bit more involved notion due to the state dependence of the weights. An *agent i is connected to agent j on the time interval (s,t)* (for $s < t$ and $\varepsilon > 0$ given) if there exists a chain of agents $i_1, i_2, \ldots, i_k$ such that

$$|x_i(t-1) - x_{i_1}(t-1)| \leq \varepsilon, |x_{i_1}(t-2) - x_{i_2}(t-2)| \leq \varepsilon, \ldots, |x_{i_k}(s) - x_j(s)| \leq \varepsilon.$$

Now we can state our second analytical result.

Leading-the-pack Theorem (LPT)

For the truth dynamics given by (8), consensus on the truth holds for every initial profile $x(0)$ under the following conditions.

(i) For the LW-model, if each agent not going for the truth is connected to some agent going for the truth.
(ii) For the BC-model, if there exist some r and a sequence of time periods t_m with $2 \leq t_m - t_{m-1} \leq r$ for $m = 1, 2, 3, \ldots$ such that each agent not going for the truth is on each time interval $(t_{m-1}+1, t_m)$ connected to some agent going for the truth.

For a proof, see the appendix in Hegselmann and Krause (2006).

Some comments on the theorem: In both parts, the theorem states in particular that consensus on the truth can be achieved even when *just one single agent* goes for the truth – *provided* all other agents are *connected* to this one. That's where the name of the theorem comes from. Interestingly enough, in this case, the leading agent who approaches the truth though his own opinion may have been quite distant from the truth at the beginning.

By part (i), for the LW-model, consensus on the truth holds in particular if at least one agent is going for the truth and the matrix of the weights is irreducible. If the latter is not the case, it may happen that the agent going for the truth does not approach the truth; all the other agents can approach a consensus among themselves, also different from the truth.

By part (ii), for the BC-model, consensus on the truth holds if every agent not going for the truth connects at every time interval to some agent going for the truth – this truth-seeking agent, however, may change from one time interval to another. Very different from what may happen in the LW-model, in the BC-model every agent going for the truth does approach the truth as has been shown in Kurz and Rambau (2006). In the BC-model, two extreme cases can be settled directly without relying on the LPT. For if $\varepsilon = 0$ or $\varepsilon = \infty$ (meaning ε is big enough), the BC-model can be explicitly solved. In the first case, no two agents are connected and one obtains that exactly the agents going for the truth will approach it. In the second case, all agents are connected and it is not difficult to verify that consensus on the truth holds as long as one agent goes for it.

https://doi.org/10.3366/E1742360009000604 Published online by Cambridge University Press

More interesting, of course, is part (ii) for a confidence level ε that is not extreme. This is the case in figures 2a and 2b below, both results of computer simulations. A dark grey path represents an agent going for the truth; a light grey path shows an agent not going for the truth. The dotted line is the true value T. Figure 2a shows the first 20 periods and figure 2b shows the first 200 periods of a truth dynamics. In figure 2a, 50% of the agents are randomly distributed truth seekers. Nevertheless, at period 20 about 75% of the agents are already very close to the truth. The society ends up with two polarized camps, the upper one far from the truth. In figure 2b, one single truth seeker will finally lead the whole society to a consensus on the truth – though that will take many more periods than the 200 that are visible in figure 2b.

5. IF NOT ALL ARE TRUTH SEEKERS: A CASE-STUDY

The *Leading-the-pack Theorem* implies that under certain conditions, a few truth seekers, or even one, may lead the pack to the truth. But when can we expect those conditions to be fulfilled? And where do the dynamics tend to end if such conditions are not fulfilled? We will address these questions by simulations, thereby doing a kind of **c**omputer **a**ided **s**ocial **e**pistemology, i.e. a CASE-study.

Intuition and what we have considered so far suggest that at least *four* parameters are crucial: the frequency of α-positives, the position of the truth, the confidence level ε, and the strength α of the truth attraction. In a huge number of systematic simulations, we will vary the parameters as follows:

1. The position of the truth T: 0.1, 0.3, 0.5.
2. The frequency F of agents with $\alpha = 0$: 10%, 50%, 90%.
3. The confidence level: $\varepsilon = 0.01, 0.02, \ldots, 0.4$; i.e. 40 steps.
4. The strength of the truth directedness: $\alpha = 0.01, 0.02, \ldots, 1.0$; i.e. 100 steps.

As to T and F, the simulation strategy is a *scenario* approach. The truth may be in the centre of the opinion space, the truth may be an extreme, or T lies somewhere in between. Almost all, half, or only some of the agents may be *non*-truth seekers. Together, that is a total of 3×3 scenarios. As to ε and α, we follow a kind of '*grid* approach' with 40×100 points.

The guiding research question will be: *What is the final 'distance to the truth', what is the final 'truth deviation' after the dynamics has stabilized?* Truth deviation could be measured in different ways. We will do it in a way that closely follows the definition of the *standard deviation*. We simply substitute the mean by T. Thus *truth deviation* δ (t) is defined as follows:

$$\delta(t) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(T - x_i(t))^2} \qquad (9)$$

https://doi.org/10.3366/E1742360009000604 Published online by Cambridge University Press

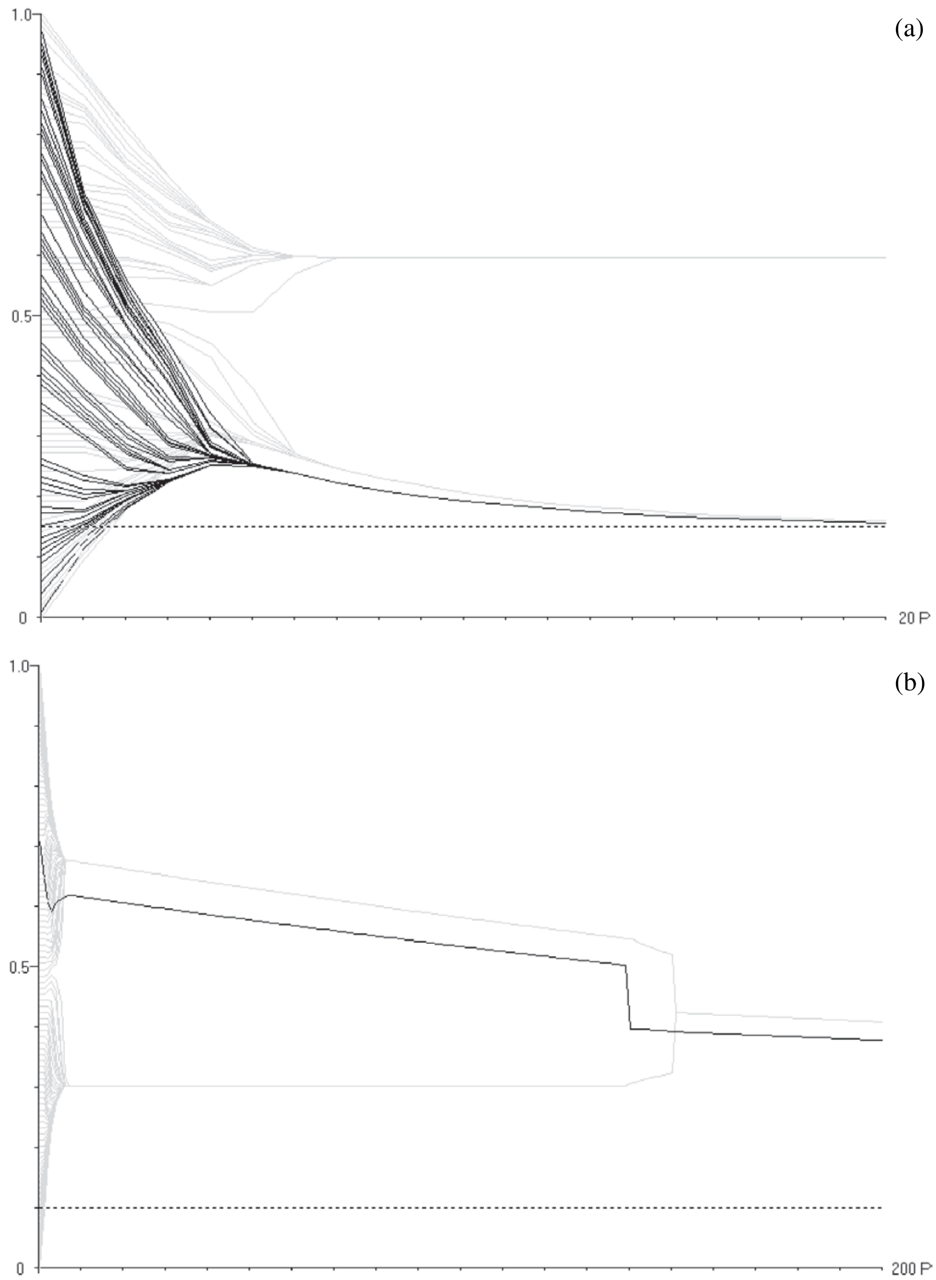

Figure 2. (a) 50% truth seekers with $\alpha_i = 0.25$ (b) One single truth seeker with $\alpha_i = 0.25$. Light grey: $\alpha_i = 0$, i.e. not interested in the truth. Dark grey: $\alpha_i = 0.25$, i.e. a truth seeker. The confidence level is always $\varepsilon = 0.2$. The dotted line is the true value T.

https://doi.org/10.3366/E1742360009000604 Published online by Cambridge University Press

Analysing a 4-dimensional parameter space is hard enough a task. Therefore all other things are kept as simple as possible. We assume, first, confidence levels ε that are the *same* for all agents and remain *constant* over time, second, a *constant* α that is the same for all α-positives, third, a *fixed number* of 100 individuals, fourth, a *uniform* start distribution, and, finally, *simultaneous* updating.

The *average* truth deviations after the dynamics has stabilized are summarized for a total of 9 scenarios in figure 3. Along the two horizontal axes of each 3-D plot, α and ε increase in 100 and 40 steps, respectively, each step of size 0.01. For each parameter constellation, we ran 50 simulations. Thus figure 3 summarizes the results of a total of 1.8 million simulation runs, each one continued until – within the limits of numerical accuracy of a computer – nothing changed any longer. The vertical axis gives the final *average* truth deviation δ for the 50 runs. Figure 4 shows the coefficients of variation for all the results in figure 3.

The most striking result of our simulations is that in all scenarios for a huge set of ⟨α,ε⟩-values the final truth deviation is *zero. Obviously, a whole society might arrive at a consensus on the truth without every member being an active and successful truth seeker.* This effect is more distinctive if the truth is in the centre of the opinion space and the proportion of agents that are not interested in the truth is small. However, even if the truth is extreme and the fraction of truth seekers is small, the effect exists for many ⟨α,ε⟩–values. Additionally, the effect is remarkably robust. In all scenarios in figure 3, we see *two* plateaus. On the 0-level plateau, all end up at the truth. That does not hold for the higher, second plateau. As figure 4 shows, a lot of dispersion (measured by the coefficient of variation) is to be found only at the edges between the two plateaus – but nowhere else. Consequently, for almost all circumstances, the final truth deviation in repeated runs is pretty much the same.

It does not take a whole society of truth seekers to finally end up with a consensus on the truth. However, for that to be the case, α and ε have to have the right relation. For a given ε interval, the truth attraction may be too strong. Or the other way round: For a given truth attraction a confidence interval may be too small. In both cases the α-positives end up at the truth, but on their way to the truth, they loose their influence on major parts of the population. The avant-garde may be 'too good' in the sense that for major parts of the population, the positions of the avant-garde is 'out of range' in a very literal sense, i.e. out of their confidence intervals. Then, what one might call a *leaving-behind-effect* generates a situation in which the truth seekers end up at the truth, but only among themselves, while all others live somewhere else in the opinion space, possibly polarized and quite distant from the truth.

6. PERSPECTIVES

The framework we presented here allows the study of social and epistemic processes in which more or less successful truth seekers are at the same time

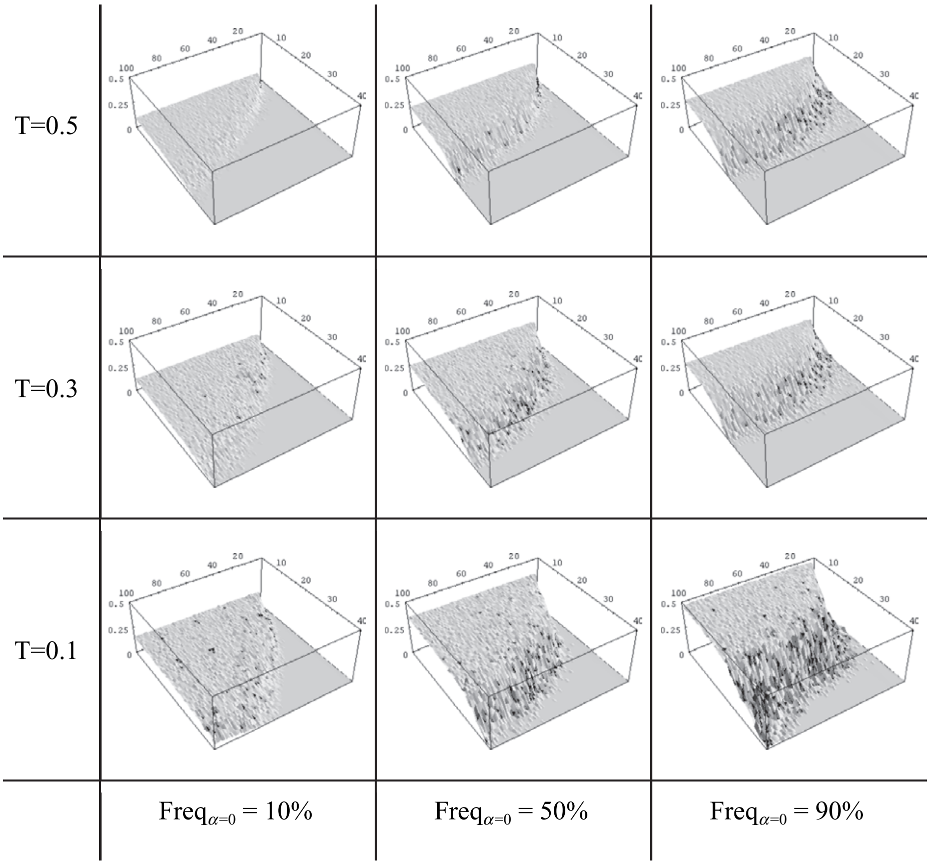

Figure 3. Average truth deviation after stabilization for three positions of the truth (T= 0.1, 0.3, 0.5) and three percentage values of agents not interested in the truth ($Freq_{\alpha=0}$= 10%, 50%, 90%). In each 3-D plot, the vertical axis is the average truth deviation. The strength of the truth attraction α is given in 100 steps from 0.01 to 1 on a horizontal axis. The size of the confidence interval ε is given on the second horizontal axis in 40 steps of size 0.01 from 0 to 0.4.

involved in social exchange processes. The framework allows us to study the dynamics of *stylized epistemic constellations*:

- One can focus on *difficult situations for spreading the truth*, for instance, opinion leaders with positions distant from the truth and not interested in the truth, asymmetric confidence intervals with a bias against the truth, or a peak far from the truth in the start distribution.
- The fairly optimistic *idealized assumptions on T could be given up*. For instance, new evidence may be noisy (cf. Douven 2008) or point in two or more different directions.

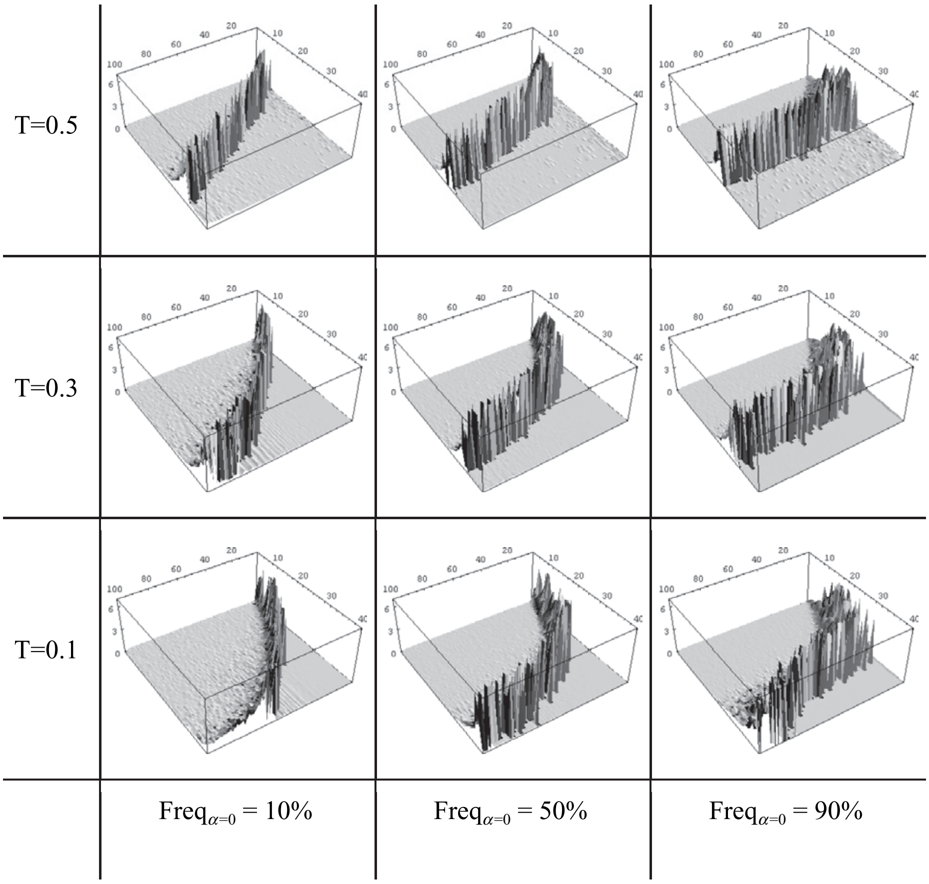

Figure 4. Coefficient of variation for the average truth deviation in figure 3. In each 3-D plot, the vertical axis is the coefficient. As in figure 3, the strength of the truth attraction α is given in 100 steps from 0.01 to 1 on a horizontal axis. The size of the confidence interval ε is given on the second horizontal axis in 40 steps of size 0.01 from 0 to 0.4.

- We took the strength of the attraction of truth as given. Another view could be to look at α as a parameter that can be influenced by intervention. Under such a view, one might start thinking about efficient *truth proliferation policies*: Which agents, holding what views, should have their attractions to truth modulated in order that all or at least a significant part of a society believes the truth? What if there is time pressure? What to do if the social exchange process has a network structure with primarily local interactions?
- In this article the LW-model and the BC-model played an important role. But the general framework is *not necessarily* tied to these social processes. There are other processes, for instance, processes that are *not* range preserving.

Thus the framework, the methods, and results suggest that there is a realistic perspective for a social epistemology that addresses theoretical and technical questions on the spreading of truths.

Our approach disregards and abstracts away almost everything, the details of deliberative exchange, such as arguments, questions, inferences, systematization of beliefs, etc., all that does not come into focus and cannot be analyzed in our framework. The reason for that is obviously the *low* resolution of our approach. But it is exactly the low resolution that makes it possible to focus on macro effects that disappear under *high* resolution approaches. However, as research programs, both approaches are *compatible*. Their advantages and disadvantages are probably complimentary.

REFERENCES

Bogdan, Radu J. (ed.) 1981. *Keith Lehrer*. Dordrecht: D. Reidel.

Douven, Igor. 2008. *Simulating Peer Disagreements*. Manuscript.

Goldman, Alvin I. 1999. *Knowledge in a Social World*. Oxford: Oxford University Press.

Gustafsson, Johan and Martin Peterson. 2008. *A Computer Simulation of the Argument from Disagreement*. Manuscript.

Hamblin, Charles L. 1970. *Fallacies*. London: Methuen.

Hamblin, Charles L. 1971. "Mathematical Models of Dialogue." *Theoria* 37(2): 130–55.

Hegselmann, Rainer. 1985. *Formale Dialektik – Ein Beitrag zu einer Theorie des rationalen Argumentierens*. Hamburg: Felix Meiner.

Hegselmann, Rainer. 2004. "Opinion Dynamics – Insights by Radically Simplifying Models." In D. Gillies (ed.), *Laws and Models in Science*, pp. 19–46. London: King's College Publications.

Hegselmann, Rainer and Ulrich Krause. 2002. "Opinion Dynamics and Bounded Confidence – Models, Analysis, and Simulations." *Journal of Artificial Societies and Social Simulation (JASSS)* 5(3). http://jasss.soc.surrey.ac.uk/5/3/2.html

Hegselmann, Rainer and Ulrich Krause. 2005. "Opinion Dynamics Driven by Various Ways of Averaging." *Computational Economics* 25(4): 381–405.

Hegselmann, Rainer and Ulrich Krause. 2006. "Truth and Cognitive Division of Labor – First Steps Towards a Computer Aided Social Epistemology." *Journal of Artificial Societies and Social Simulation (JASSS)* 9(3). http://jasss.soc.surrey.ac.uk/9/3/10.html

Krause, Ulrich. 2009. "Compromise, Consensus, and the Iteration of Means." *Elemente der Mathematik* 64: 1–8.

Kurz, Sascha and Jörg Rambau. 2006. "On the Hegselmann-Krause Conjecture in Opinion Dynamics." Bayreuth: Working paper.

Lehrer, Keith. 1975. "Social Consensus and Rational Agnoiology." *Synthese* 31(1): 141–60.

Lehrer, Keith. 1976. "When Rational Disagreement is Impossible." *Nous* 10(3): 327–32.

Lehrer, Keith. 1977. "Social Information." *The Monist* 60: 473–87.

Lehrer, Keith. 1981a. "A Self Profile." In R. J. Bogdan (ed.), *Keith Lehrer*, pp. 3–104. Dordrecht: D. Reidel.

Lehrer, Keith. 1981b. "Replies." In R. J. Bogdan (ed.), *Keith Lehrer*, pp. 223–42. Dordrecht: D. Reidel.

Lehrer, Keith. 1985. "Consensus and the Ideal Observer." *Synthese* 62(1): 109–20.

Lehrer, Keith and Carl G. Wagner. 1981. *Rational Consensus in Science and Society: A Philosophical and Mathematical Study*. Dordrecht: D. Reidel.

List, Christian. 2005. "Group Knowledge and Group Rationality: A Judgement Aggregation Perspective." *Episteme: A Journal of Social Epistemology* 2(1): 25–38.

Loewer, Barry. (ed.) 1985. Consensus. *Synthese* 62(1) (special issue).

Malarz, Krzysztof. 2006. "Truth Seekers in Opinion Dynamic Models." *International Journal of Modern Physics C* 17(10): 1521–4.

NOTES

1 *Therefore* they stress the freedom of thought and discussion (Mill) or the right to make public use of reason in all matters (Kant).

2 Hamblin (1970, 1971) and Hegselmann (1985) are typical examples of ambitious formal approaches that so far did not deliver very much. That is probably due to their *high*-resolution modeling approach that contrasts sharply with the *low*-resolution approach favored in this article.

3 Note that under the interpretation of Lehrer and Wagner "*t*" in "$t \rightarrow \infty$" is *not* time but a step in the iterated aggregation procedure.

4 For analytical results, cf. Hegselmann and Krause (2002, sect. 2) as well.

5 For the history of the BC-model, see Hegselmann and Krause (2006, footnote 4).

6 For general range preserving social processes in higher dimensions, see Krause (2009).

7 For recent contributions taking up the BC-model with or without truth, see Douven (2008), Gustafsson and Peterson (2008), and Malarz (2006).

Rainer Hegselmann is professor of philosophy at the University of Bayreuth, Germany. He is chiefly working on problems of modeling and simulating complex social dynamics. Currently his main projects are modeling of opinion dynamics and a reconstruction of Hume's moral and political theory by means of computer simulations.

Ulrich Krause is professor emeritus of mathematics at the University of Bremen, Germany. He holds doctoral degrees in mathematics and economics. He published a few books and many articles in the fields of positive discrete dynamical systems and commutative algebra as well as in the fields of economics of production and social welfare theory.