Social Epistemology, Debate Dynamics, and Truth Approximation

Gustavo Cevolani

1 Introduction

Social epistemology tackles the classical problems of knowledge, truth and rationality by exploring those social processes, methods and practices—like testimony, authority, belief aggregation or the search for consensus—by which "epistemic subjects" interact with other agents who influence, and are influenced by, their beliefs (Goldman 2010). In particular, "veristic" social epistemology (Goldman 1999; Kitcher 1993)—as opposed to constructivist approaches in the sociology of knowledge—aims at analyzing and evaluating such practices with respect to their tendency to disseminate true beliefs in communities of truth-seeking agents (both scientists and laymen). Within the veristic approach, the *truth conduciveness* of social practices—i.e., their effectiveness in promoting epistemic progress, construed as increasing approximation to the truth, within the relevant community—is thus the crucial issue.

In this paper, I address the issue of truth conduciveness by focusing on the "theory of dialectical structures" recently developed by Gregor Betz (2013). This theory aims at investigating the role of controversial argumentation in consensus formation and truth approximation. More specifically, the theory studies, via computer simulations, how the opinions of a group of agents change as arguments are introduced in the debate, and whether the group reaches a consensus or makes progress toward the truth as a result. As Betz (2013, pp. 39–40) notes, his approach is closely related to the philosophical accounts of verisimilitude or truthlikeness, and in particular to the "basic feature" approach of Cevolani et al. (2011). In

G. Cevolani
Department of Philosophy and Education, University of Turin, via S. Ottavio 20, 10124 Turin, Italy
e-mail: g.cevolani@gmail.com

U. Mäki et al. (eds.), *Recent Developments in the Philosophy of Science: EPSA13 Helsinki*, European Studies in Philosophy of Science 1,
DOI 10.1007/978-3-319-23015-3_5

Sects. 2 and 3, I briefly present the basic feature approach and the theory of dialectical structures, respectively; Sects. 4 and 5 focus on the formal and conceptual relationships between these two accounts; finally, in Sect. 6, some general remarks on the issue of truth conduciveness in veristic social epistemology are offered.

2 Truth Approximation as the Goal of Inquiry

Intuitively, theory A is verisimilar (or truthlike) if it is close or similar to the whole truth about a given target domain. Thus, A may be false but still a good approximation to the truth, and even a better approximation than another (true or false) theory. This notion was originally introduced by Karl Popper (1963, ch. 10) in order to defend the idea that progress can be explained in terms of the increasing verisimilitude of scientific theories. According to Popper, such theory-changes as that from Newton's to Einstein's theory are progressive because, although the new theory is, strictly speaking, presumably false, we have good reasons to believe that it is closer to the truth than the superseded one: increasing verisimilitude is the key ingredient for progress.[1]

The main idea underlying the basic feature approach (Cevolani et al. 2011, 2013) is that the verisimilitude of a theory can be defined in terms of the balance of the true and false information that it conveys about the "basic features" of the target domain. Suppose that such domain is described by a finite propositional language $\mathscr{L}_n$ with n logically independent atomic sentences $a_1, a_2, \ldots, a_n$. Then, its basic features are described by the so called basic sentences or literals of $\mathscr{L}_n$, i.e., by the atomic sentences and their negations. A non-contradictory conjunction of m basic sentences, with $m \leq n$, will be called a "conjunctive theory" ("c-theory", for short). The 3^n c-theories of $\mathscr{L}_n$ include (for the special case $m = n$) the so called constituents, which are consistent conjunctions of n basic sentences. Constituents are the most informative sentences in $\mathscr{L}_n$ (there are 2^n such sentences); intuitively, they are the descriptions of the possible worlds expressible within the language.

There is only one *true* constituent, denoted $C_\star$, which is the most informative true description of the actual world within $\mathscr{L}_n$. In this sense, $C_\star$ represents "the (whole) truth" about the world as expressible in $\mathscr{L}_n$. As c-theory A is compared with $C_\star$, its degree of verisimilitude can be simply defined in terms of the number of its true and false basic sentences, i.e., conjuncts that A shares or not, respectively, with $C_\star$. While this definition is restricted to the rather special case of conjunctive theories,

[1] After Popper's own definition of verisimilitude was shown to be untenable, authors like Oddie (1986), Niiniluoto (1987), Kuipers (2000), Schurz and Weingartner (2010), and Zamora Bonilla (1996) developed a number of post-Popperian theories of verisimilitude; see Niiniluoto (1998) for an historical survey, and Cevolani and Tambolo (2013) for an introduction to the verisimilitudinarian approach to scientific progress.

it will be sufficient for my purposes, since, as explained in Sect. 3, also the theory of dialectical structures is limited to such kind of theories.

Let us call each true conjunct of c-theory A a *match* of A, and each false conjunct a *mistake* of A. Moreover, let t_A, and f_A denote, respectively, the number of matches and the number of mistakes of A. Finally, let define the "degree of true (basic) content" $cont_t(A, C_\star)$ and "degree of false (basic) content" $cont_f(A, C_\star)$ of A as the normalized number of its matches and mistakes, respectively:

$$cont_t(A, C_\star) \stackrel{df}{=} \frac{t_A}{n} \quad \text{and} \quad cont_f(A, C_\star) \stackrel{df}{=} \frac{f_A}{n} \tag{1}$$

As said before, A is (highly) verisimilar if A makes many matches and few mistakes about $C_\star$. This intuition is captured by the following "contrast measure" of the verisimilitude of A (Cevolani et al. 2011, p. 188):

$$Vs_\phi(A) \stackrel{df}{=} cont_t(A, C_\star) - \phi cont_f(A, C_\star) \tag{2}$$

where $\phi > 0$.[2] Intuitively, different values of ϕ express the relative weight assigned to truth and falsity: the greater ϕ, the farther from the truth A will be due to its mistakes.

It is important to appreciate the difference between verisimilitude and other related notions. For instance, the (degree of) "accuracy" or "approximate truth" $acc(A)$ of c-theory A is arguably definable as the number of matches of A divided by the total number m_A of its conjuncts (cf. also Cevolani 2014a):

$$acc(A) \stackrel{df}{=} \frac{t_A}{m_A} \tag{3}$$

Note that, for any true c-theory A, $acc(A) = 1$. This fact alone shows that accuracy is only one "ingredient" of verisimilitude: A may be highly accurate (for instance because it is true) without being highly verisimilar. As an example, if $C_\star \equiv a_1 \wedge \cdots \wedge a_n$, then c-theories $A = a_1$ and $B = a_1 \wedge a_2$ are equally accurate, since they are both true and hence $acc(A) = acc(B) = 1$, but B is more verisimilar than A, since it conveys more true information about the world. For the same reason, the *false* c-theory $B' = \neg a_1 \wedge a_2 \wedge a_3 \wedge \cdots \wedge a_n$ may well be more verisimilar, although less accurate, than A. In Popper's words, verisimilitude "represents the idea of approaching comprehensive truth. It thus combines truth and content" (Popper 1963, p. 237): it follows that informative falsehoods may be more valuable, in terms of verisimilitude, than uninformative truths (e.g., tautologies). Among truths, however, verisimilitude covaries with logical strength: in the above example, since A and B are both true, and B entails A, then B is more verisimilar

[2] One may note that measure Vs_ϕ is not normalized, and varies between $-\phi$ and 1. A normalized measure of the verisimilitude of A is $(Vs_\phi(A) + \phi)/(1 + \phi)$, which varies between 0 and 1.

than A (cf. Niiniluoto 1987, sec. 6.6 for a discussion of this and other properties of verisimilitude measures).

3 Debate Dynamics: Evolving Dialectical Structures

The theory of dialectical structures (Betz 2013) aims at reconstructing multi-agent debates and studying, via computer simulations, the role of controversial argumentation in consensus formation and truth approximation. A debate is characterized by the "proponents" (agents) engaged in a given controversy, by a "pool" of relevant sentences, and by the arguments recognized by the participants in the debate. At any stage of the debate, each proponent adopts a given "position" (theory) concerning the sentences in question; as arguments are introduced in the debate, agents may be forced to change their positions, since these may become untenable in view of the proposed arguments.

With reference to the formal framework introduced in Sect. 2, the theory of dialectical structures can be presented as follows (cf. Betz 2013, ch. 2). The pool of sentences under discussion is the set of basic sentences (literals) of $\mathscr{L}_n$. At any moment of time, each participant adopts a position represented by a c-theory of $\mathscr{L}_n$: "complete" positions correspond to constituents (conjunctions of n literals) and "partial" positions to "proper" c-theories (conjunctions of m literals, with $0 < m < n$). So far, the study is limited to agents adopting complete positions or constituents as their theories (Betz 2013, p. 10; but see Sect. 5 below).

An argument is a deductively valid inference from some premises to a conclusion; all sentences in an argument are drawn from the pool of basic sentences. Thus, as arguments are proposed, previously unseen inferential relationships among the basic sentences of the underlying language are established.[3] The arguments in a debate are dialectically interconnected in the sense that each of them can "support" or "attack" other arguments.[4] More precisely, an argument attacks another argument when the conclusion of the former contradicts a premise of the latter; and an argument supports another argument when the conclusion of the former is equivalent to a premise of the latter. A *dialectical structure* represents a "snapshot" of the current state of the debate, displaying the set of arguments and the corresponding support and attack relations. For example, in the simple dialectical structure

[3] This implies, as noted by an anonymous referee, that the basic sentences appearing as premises or conclusions of some argument do acquire a richer "internal structure" to account for the inferential relation captured by the argument itself. At the present stage of Betz's theory, however, such internal structure plays no role, and both premises and conclusions are treated as unanalyzed basic sentences, especially as far as the definition of the verisimilitude proponents' positions is concerned.

[4] There are also other possible relationships among arguments (cf. Betz 2013, sec. 1.3), that however play no role in what follows. Even the support and attack relations are introduced here just to illustrate Betz's basic framework.

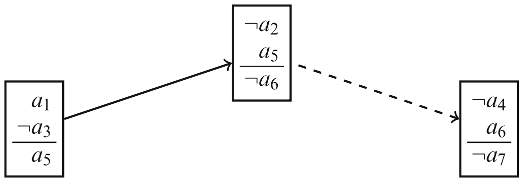

Fig. 1 A (very simple) dialectical structure; *continuous* and *dashed arrows* indicate the support and attack relation, respectively

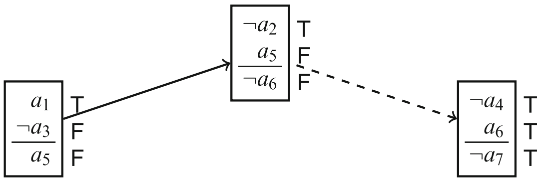

Fig. 2 A coherent complete position on the dialectical structure of Fig. 1. A "T" or "F" placed on the right of a sentence means that the corresponding sentence is, respectively, true or false

displayed in Fig. 1, the first argument (on the left) supports the second (middle) argument, which attacks the third argument (on the right).

At each step of a controversy, a proponent's position is "dialectically coherent" if it respects the argumentative structure of the debate, i.e., the inferential relationships so far established among the basic sentences of $\mathscr{L}_n$. More precisely, a complete position is dialectically coherent when, for any argument in a given structure, if its premises are true according to the position, also the conclusion is deemed true. As an example, Fig. 2 displays a coherent complete position—corresponding to constituent $a_1 \wedge \neg a_2 \wedge a_3 \wedge \neg a_4 \wedge \neg a_5 \wedge a_6 \wedge \neg a_7$ in $\mathscr{L}_7$—on the dialectical structure of Fig. 1.

How is belief dynamics modeled within the theory of dialectical structures? The central idea is that a debate is driven by the arguments which are proposed in the course of a controversy.[5] For instance, suppose that, with reference to the above dialectical structure, an argument is introduced to the effect that a_1 and a_6 jointly entail a_2. Then, the position displayed in Fig. 2 becomes incoherent: in fact, the proponent accepts both premises a_1 and a_6 but rejects the conclusion a_2 (since $\neg a_2$ is true according to the displayed position). Thus, the position has to be changed, on pain of inconsistency. Position dynamics is governed by an instance of the principle of conservatism or informational economy (Gärdenfors 1988): the proponent adopting a position which is made incoherent will shift to one of the *closest* coherent positions available. Such closest coherent positions are represented by the constituents which minimize the normalized Hamming distance from (the constituent representing) the current position (Betz 2013, p. 39). The Hamming distance between two constituents C_i and C_j is the number of atomic sentences on which the two constituents disagree; the normalized Hamming distance $\Delta(C_i, C_j)$

[5] The simplest method of introducing arguments in the simulation of a debate is randomly selecting some literals as premises and another as the conclusion; a good part of the theory, however, is devoted to studying more complex, and more interesting, argument construction mechanisms (see Betz 2013, pp. 8 ff. for a overview).

equals this number divided by n, the total number of atomic sentences. Since $\Delta(C_i, C_j)$ varies between 0 and 1, the (Hamming) similarity or closeness $s(C_i, C_j)$ between C_i and C_j can be defined as:

$$s(C_i, C_j) \stackrel{df}{=} 1 - \Delta(C_i, C_j) \tag{4}$$

i.e., as the normalized number of matches between C_i and C_j. Betz (2013, p. 39) takes $s(C_i, C_j)$ to measure the agreement between the two complete positions represented by C_i and C_j. Moreover, if $C_\star$ represents the true position, then $s(C, C_\star)$ is construed as the degree of verisimilitude of position C. As Betz (2013, p. 40, fn. 7) notes, this is an extreme special case of the Vs_ϕ measure of verisimilitude introduced in Sect. 2, since it is easy to prove that for any constituent C[6]:

$$\text{for } \phi = 0, \text{ then } Vs_0(C) = cont_t(C, C_\star) = s(C, C_\star) \tag{5}$$

Finally, Betz (2013, p. 40) introduces the idea of "debate-wide mean verisimilitude". This is simply defined as the average degree of verisimilitude over all individual positions—a value which intuitively expresses how close to the truth is the group of proponents taken as a whole. This mean verisimilitude value is used by Betz to assess the truth conduciveness of different epistemic practices as dialectical structures evolve during a debate (Betz 2013, part II).

4 Truth Approximation in Debate Dynamics

When does belief dynamics lead a group of epistemic agents closer to the truth, i.e., increases the verisimilitude of their individual and/or aggregated theories? As recent work has shown, it is very hard to specify suitably general conditions under which revising or merging individual beliefs effectively tracks truth approximation.[7] The same is true in the present case, since the introduction of new valid arguments in a debate can both increase and decrease, depending on the specific circumstances, the verisimilitude of the individual positions, and hence the debate-wide mean verisimilitude. The following example is a case in point.

Example 1. Let us consider a simple pool of sentences, given by the eight basic sentences of $\mathscr{L}_4$: $a_1, \neg a_1, \ldots, a_4, \neg a_4$. Suppose that, at the initial stage of some debate, three agents adopt the following complete positions:

[6]More generally, if C is a constituent then $Vs_\phi(C)$ only depends on its degree of truth content $cont_t(C, C_\star)$, i.e., on $s(C, C_\star)$. In fact, since for constituents $cont_t(C, C_\star) = 1 - cont_f(C, C_\star)$, then $Vs_\phi(C) = cont_t(C, C_\star) - \phi cont_f(C, C_\star) = (\phi + 1)cont_t(C, C_\star) - \phi$.

[7]See, in particular, Niiniluoto (1999, 2011) and the papers collected in Kuipers and Schurz (2011); for the relations between the basic feature approach and different accounts of belief dynamics see Cevolani et al. (2011, 2013), and Cevolani (2013, 2014b).

$$\begin{array}{l} X: \ a_1 \ \wedge \ a_2 \ \wedge \neg a_3 \wedge \ a_4 \\ Y: \ a_1 \ \wedge \neg a_2 \wedge \ a_3 \ \wedge \neg a_4 \\ Z: \neg a_1 \wedge \ a_2 \ \wedge \neg a_3 \wedge \neg a_4 \end{array}$$

Assuming that $a_1 \wedge a_2 \wedge a_3 \wedge a_4$ is the true position, is it easy to calculate that $Vs_0(X) = \frac{3}{4}$, $Vs_0(Y) = \frac{1}{2}$, and $Vs_0(Z) = \frac{1}{4}$, and that the debate-wide mean verisimilitude is $\frac{1}{2}$. Now suppose that an argument is introduced in the debate to the effect that a_1 and a_3 jointly entail a_4. While positions X and Z remain coherent, position Y is now dialectically incoherent, since the agent adopting Y deems both premises a_1 and a_3 true, but the conclusion a_4 false. As a consequence, position Y has to be changed to one of the closest positions, i.e., to one of the following constituents:

$$\begin{array}{l} Y': \ a_1 \ \wedge \neg a_2 \wedge \neg a_3 \wedge \neg a_4 \\ Y'': \neg a_1 \wedge \neg a_2 \wedge \ a_3 \ \wedge \neg a_4 \\ Y''': \ a_1 \ \wedge \neg a_2 \wedge \ a_3 \ \wedge \ a_4 \end{array}$$

As one can check, $Vs_0(Y') = Vs_0(Y'') = \frac{1}{4}$ and $Vs_0(Y''') = \frac{3}{4}$. In two cases out of three, corresponding to the choice of either Y' or Y'' as the new position, this will be less verisimilar than the old position Y; accordingly, the introduction of the argument above will decrease the debate-wide mean verisimilitude (from 0.5 to about 0.42). Only in the third case, corresponding to the change from Y to Y''', does the individual verisimilitude and hence the debate-wide mean verisimilitude increase (the latter from 0.5 to about 0.58).

It is perhaps worth noting that, in the above example, the newly introduced argument is not only valid but also sound, in the sense that both premises are true. Thus, even accepting an inference from true premises to a true conclusion may actually lead proponents farther from the truth about the matter under discussion.

Cases like Example 1 show that debate dynamics does not track truth approximation in general, and can sometimes hinder, in fact, cognitive progress. In more specific situations, however, there may be other factors that counterbalance the negative effects of cases like these. For instance, the agents may already share some *true* information about the world, construed as common evidence or background knowledge (Betz 2013, p. 14 and ch. 12). This is modeled in Betz's framework by fixing, right at the beginning of a simulation, the truth values of some of the basic sentences to the correct value, hence helping the group both to attain consensus and to approach the truth about the domain. In any case, by repeatedly simulating a high number of abstract debates (evolving dialectical structures as those introduced in Sect. 3), one can identify some general patterns of how argumentation influences the group capability of approaching truth. In particular, Betz's results tend to confirm the idea that, under suitably specified circumstances, controversial argumentation is instrumental both for attaining a consensus and for approaching truth; moreover, consensus formation is a good, although fallible, indicator of truth approximation.[8] Without aiming at an even cursory presentation of these results (for which see Betz 2013, secs. 1.4–1.5), in the following section I focus on the link between the theory of dialectical structures and the notion of verisimilitude as introduced in Sect. 2.

5 Extending the Theory of Dialectical Structures

In the analysis of debate dynamics presented so far, I only considered agents maintaining complete positions (constituents). This means that all participants in a debate have a definite opinion on each element of the relevant pool of basic sentences. As Betz (2013, p. 10) notes, this is a quite limiting assumption: simulating debates among agents who adopt *partial* positions would allow for a more realistic representation of actual controversies, and may shed new light on the truth approximation issue.

An agent adopting a partial position (i.e., a proper c-theory) suspends the judgment on at least some of the basic sentences in the pool. As an example, Fig. 3 displays the partial position corresponding to c-theory $a_1 \wedge \neg a_2 \wedge \neg a_3 \wedge a_6$ in $\mathscr{L}_7$. Note that this position is (dialectically) incoherent, since it cannot be "extended" to a coherent complete position (Betz 2013, pp. 35–36). In fact, since according to the agent a_1, $\neg a_2$, $\neg a_3$ and a_6 are all true, given the first argument (on the left) also a_5 should be true; but since the second argument (in the middle) has a false conclusion ($\neg a_6$) and a true premise ($\neg a_2$), the remaining premise (a_5) should be false. Since the dialectical structure forces the proponent to assign opposite truth values to the same sentence (a_5), this position is incoherent.

Allowing for agents adopting partial positions requires two relevant modifications of the theory of dialectical structures. First, the mechanism governing belief dynamics, which guarantees that the agents' positions remain coherent in the course of the debate, has to be extended. Second, the verisimilitude of partial positions has to be adequately defined.

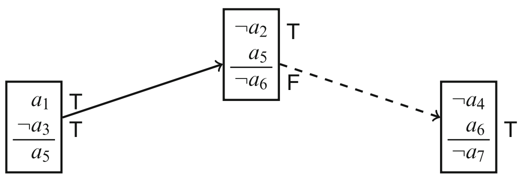

Fig. 3 An incoherent partial position on the dialectical structure of Fig. 1

[8]See Betz (2013, chs. 3 and 10) for a presentation of these results and of the underlying assumptions. Given space limitations, I cannot discuss the interesting relationships between truth approximation and consensus formation. As Betz (2013, pp. 13 ff.) notes, there is a clear asymmetry between the two: in fact, reaching the truth implies having reached a consensus, while the converse does not hold in general: cases of "spurious consensus"—i.e., agreement on some false position—are very well possible both in scientific and ordinary controversies.

As far as the former problem is concerned, the example in Fig. 3 makes clear that a proponent adopting an incoherent partial position has many different options to make it coherent. In fact, while a complete position can only be modified by replacing one or more of the conjuncts of the corresponding constituent with their negation (cf. Example 1), an agent adopting a partial position can also withhold judgment on a previously accepted sentence to restore coherence. As an example, the position in Fig. 3 can be made coherent by replacing, say, $\neg a_2$ with its negation (thus shifting to the new position $a_1 \wedge a_2 \wedge \neg a_3 \wedge a_6$), but also by suspending the judgment, e.g., on a_1 (thus weakening the agent's position to $\neg a_2 \wedge \neg a_3 \wedge a_6$); and indeed in many other ways. This raises the problem of how the new, coherent position should be selected. Since the Hamming distance is only defined for constituents (cf. Eq. 4), the distance minimization method introduced in Sect. 3 can not be applied in this case. In short, one needs an extended definition of the distance (or, equivalently, closeness) between any two proper c-theories in order to identify the closest partial positions to a given, incoherent partial position.[9]

Concerning the definition of the degree of verisimilitude of a partial position A, Betz (2013, pp. 39–40, especially fn. 7) suggests to use the number t_A of the matches (true basic sentences) of A divided by the total number m_A of the sentences in A. This proposal amounts to defining the verisimilitude of a c-theory A as its accuracy or approximate truth $acc(A)$ (cf. Eq. 3 in Sect. 2). It is worth noting that, as far as complete positions (i.e., constituents) are concerned, this definition indeed agrees with the one given in Sect. 3 in terms of the Hamming distance (see Eq. 4). In fact, it is easy to check that, for any constituent C:

$$acc(C) = cont_t(C, C_\star) = \frac{t_C}{n} = Vs_0(C) \tag{6}$$

In other words, the accuracy of a constituent equals its degree of true content, which is taken by Betz to express the verisimilitude of the corresponding complete position. On the other hand, this can be accepted as an adequate definition of verisimilitude only because, for constituents, verisimilitude and accuracy covary: since any two constituents are equally (and maximally) informative, the greater the accuracy, the greater the verisimilitude (cf. also Eq. 5).

This strict relation between verisimilitude and accuracy, however, does not hold when proper c-theories are considered (cf. Sect. 2). This implies, among other things, that two agents may adopt two different highly accurate (or even true) c-

[9]Such an extended definition is also needed if other notions employed by Betz have to be generalized to agents adopting partial positions. These include, for instance, the agreement between two positions, the debate-wide mean normalized agreement, and the resulting notion of consensus (see Betz 2013, pp. 39–40). In this connection, the analysis of "theory distance" developed by verisimilitude scholars (see, e.g., Niiniluoto 1987, sec. 6.7 and the corresponding references) may provide valuable suggestions for such a definition. In particular, the feature contrast measures of verisimilitude introduced in Sect. 2 are arguably applicable to a generalized definition of the similarity between c-theories. The discussion of such a generalization, however, has to be left to another occasion.

theories as their partial positions, but still these may be very far from the truth and also far from each other. As an example, again assuming that $C_\star \equiv a_1 \wedge \cdots \wedge a_n$ is the truth, suppose that the two agents adopt $A = a_1 \wedge a_2$ and $B = a_3 \wedge a_4$, respectively, as their partial positions. Then, the accuracy of their positions is maximal ($acc(A) = acc(B) = 1$) but their verisimilitude may well be low, if n is sufficiently high.

In conclusion, in order to adequately generalize the theory of dialectical structures to agents adopting partial positions, one needs to introduce an appropriate measure of their verisimilitude. To this purpose, the feature contrast measure Vs_ϕ defined in Sect. 2 suggests itself, being equally applicable to proper c-theories and to constituents. Allowing for agents to maintain partial positions, and using a measure like Vs_ϕ to assess their verisimilitude, would predictably lead to a richer and more realistic account of truth conduciveness within debate dynamics. It is admittedly difficult, however, to anticipate what implications this extension would have for Betz's results concerning truth approximation and consensus formation. Indeed, it is well possible that the introduction of partial positions makes debate dynamics too complex to replicate the positive tendency toward reciprocal agreement and truth displayed by Betz's simulations. Interesting as it may be, an exploration of how this extension would affect the results of the corresponding simulations exceeds the limits of the present paper.

6 Conclusions: Truth Tracking vs. Truth Approximation

Starting at least with the work of Keith Lehrer in the 1970s of the past century (Lehrer and Wagner 1981), philosophers of science have developed a number of models of belief dynamics and consensus formation in order to evaluate the truth conduciveness of different social practices. Examples are Kitcher's account of consensus practices and of the division of cognitive labor (Kitcher 1993, ch. 8) and the simulation-based approaches to so called opinion dynamics (Hegselmann 2004; Lehrer and Wagner 1981). Other relevant approaches are the theory of judgment aggregation (List 2012), developed at the interface between political theory and social choice theory (see also Zamora Bonilla 2007, for a philosophy of science perspective), and the theories of belief revision (Gärdenfors 1988) and of belief merging (Konieczny and Pino Pérez 2011) as studied in logic and AI.

In an interesting comparative survey, Betz (2013, sec. 1.7) argues that these approaches can be classified according to "the degree of logical competence which agents are assumed to possess according to the corresponding approach" (*ibidem*, p. 25). Some accounts (like standard belief revision theory) assume that epistemic agents are "logically omniscient", in the sense that their beliefs form consistent and logically closed sets with a complex internal inferential structure. Other accounts, on the contrary, depict their agents as "logically ignorant" subjects (Betz 2013, p. 27): these are exemplified by the Lehrer-Wagner and the Hegselmann-Krause models of opinion dynamics, where agents' beliefs are represented as point estimates of a real-valued parameter, with no inferential structure whatsoever (but see Riegler and Douven 2009).

As valuable as it may be, Betz's competence-based classification is better supplemented, for our present purposes, with a conceptually independent distinction concerning the issue of truth conduciveness. In Sect. 2, I introduced the notion of verisimilitude as characterizing the main epistemic goal of (a community of) rational truth-seeking agents. Such an approach naturally leads to what was called the *truth approximation* issue: i.e., exploring whether, and under what conditions, revising and merging the beliefs of a group of agents lead them closer to the truth about the relevant domain or, which is the same, increases the verisimilitude of their positions. A different question is the following: under what assumptions concerning the reliability of the agents in the group and the merging procedure, the whole truth about the domain (i.e., $C_\star$) is eventually identified? This is known in the literature as the *truth tracking* issue. The analysis of this problem was partly motivated by the so called Condorcet's jury theorem in political science, which roughly says that if the agents are sufficiently reliable—i.e., their individual probability of correctly judging the truth or falsity of a proposition is greater than 0.5—and form their opinions independently, then the probability that the majority of the agents correctly judges a proposition approaches 1 as the number of agents tends to infinity. This issue has been variously explored, by means of computer simulations, in a number of areas, including the logical theory of belief merging (Konieczny and Pino Pérez 2011), the Hegselmann-Krause model of opinion dynamics (Douven and Kelp 2011), and the theory of judgment aggregation (Hartmann and Sprenger 2012).

Truth approximation, on the one hand, and truth tracking, on the other, can be construed as different, but not incompatible, views of the truth conduciveness issue. In both cases the problem is how to gradually approach to the whole truth about the target domain, but the approach differs under a number of aspects. Without aiming at more than a partial and revisable characterization, the following main differences can be emphasized. When truth approximation is at issue,

(i) the focus is on the beliefs or theories of one or a few agents (with few exception, verisimilitude and related matters have been mainly studied, so far, within "traditional" or "individualistic" epistemology);
(ii) these beliefs are typically represented by incomplete or partial theories (like proper c-theories or partial positions) in some language, the issue being how to "complete" such theories in order to make them closer to the whole truth;
(iii) accordingly, the focus is on the *micro*dynamics of belief, and the revision or aggregating mechanism is usually sophisticated (like AGM revision rules, merging procedures, or the maximization of expected verisimilitude);
(iv) finally, this microdynamics is studied analytically, for instance by proving theorems showing how the verisimilitude of individual or collective theories is increased or maximized.

In comparison, the truth tracking issue raises different concerns:

(i) the focus is on the group or community which the epistemic agents belong to (the epistemological perspective is fully "social");

(ii) the agents' beliefs are typically represented by complete theories (like constituents or complete positions) and one then studies how such complete theories spread in the community, which may eventually converge toward the true complete position;
(iii) the *macro*dynamics of such evolution is the main concern, while the revision/aggregating procedure is comparatively less sophisticated (like the imitation of peers' average opinion in the Lehrer-Wagner and Hegselmann-Krause models);
(iv) due to the complex interaction among the beliefs of many different agents, computer simulations are usually the only available means to study the general truth tracking effectiveness of different procedures.

In the light of Betz's distinction between logically omniscient and logically ignorant approaches, theories of truth approximation tend to be associated with the former, and theories of truth tracking with the latter.

As the above comparison should make clear, truth approximation and truth tracking are strongly related ways of studying relevant problems in veristic social epistemology. Depending on the specific problem and on other circumstances, one approach may prove more useful than the other; more generally, each is characterized by strong and weak points. For instance, simulation-based accounts of truth tracking often enlighten the dynamics of entire communities of epistemic agents, but tend to obscure the relevant microdynamics at the level of individual theories; on the other hand, analyses of truth approximation often rely on a detailed account of individual belief change but quickly run into troubles with complex multi-agents interactions. As a mid-way approach combining aspects of both kinds of accounts, the theory of debate dynamics, especially when extended along the lines suggested in Sect. 5, seems able to provide valuable insights on both the truth approximation and the truth tracking issue.

Acknowledgements I would like to thank Gregor Betz for useful discussion on the topics of the paper, and Luca Tambolo and two anonymous referees for very detailed comments and corrections on an earlier draft of the paper. Financial support from the priority program *New Frameworks of Rationality*, SPP 1516 (Deutsche Forshungsgemeinshaft, grant CR 409/1-2), and from the FIRB project *Structures and Dynamics of Knowledge and Cognition* (Italian Ministry of Scientific Research, Turin unit, D11J12000470001) is gratefully acknowledged.

References

Betz G (2013) Debate dynamics: how controversy improves our beliefs. Springer, Dordrecht
Cevolani G (2013) Truth approximation via abductive belief change. Logic J IGPL 21(6):999–1016
Cevolani G (2014a) Strongly semantic information as information about the truth. In: Ciuni R, Wansing H, Willkommen C (eds) Recent trends in philosophical logic. Springer, Cham, pp 59–74
Cevolani G (2014b) Truth approximation, belief merging, and peer disagreement. Synthese 191(11):2383–2401

g.cevolani@gmail.com

Cevolani G, Tambolo L (2013) Progress as approximation to the truth: a defence of the verisimilitudinarian approach. Erkenntnis 78(4):921–935
Cevolani G, Crupi V, Festa R (2011) Verisimilitude and belief change for conjunctive theories. Erkenntnis 75(2):183–202
Cevolani G, Festa R, Kuipers TAF (2013) Verisimilitude and belief change for nomic conjunctive theories. Synthese 190(16):3307–3324
Douven I, Kelp C (2011) Truth approximation, social epistemology, and opinion dynamics. Erkenntnis 75:271–283
Gärdenfors P (1988) Knowledge in flux: modeling the dynamics of epistemic states. MIT, Cambridge
Goldman A (1999) Knowledge in a social world. Oxford University Press, Oxford
Goldman A (2010) Social epistemology. In: Zalta EN (ed) The Stanford encyclopedia of philosophy, summer 2010 edn
Hartmann S, Sprenger J (2012) Judgment aggregation and the problem of tracking the truth. Synthese 187(1):209–221
Hegselmann R (2004) Opinion dynamics – insights by radically simplifying models. In: Gillies D (ed) Laws and models in science. King's College Publications, London, pp 19–46
Kitcher P (1993) The advancement of science: science without legend, objectivity without illusions. Oxford University Press, New York
Konieczny S, Pino Pérez R (2011) Logic based merging. J Philos Logic 40:239–270
Kuipers TAF (2000) From instrumentalism to constructive realism. Kluwer Academic, Dordrecht
Kuipers T, Schurz G (eds) (2011) Special issue on "Belief revision aiming at truth approximation". Erkenntnis 75(2):151–283
Lehrer K, Wagner C (1981) Rational consensus in science and society. Reidel, Boston
List C (2012) The theory of judgment aggregation: an introductory review. Synthese 187(1):179–207
Niiniluoto I (1987) Truthlikeness. Reidel, Dordrecht
Niiniluoto I (1998) Verisimilitude: the third period. Br J Philos Sci 49(1):1–29
Niiniluoto I (1999) Belief revision and truthlikeness. In: Hansson B, Halldén S, Sahlin NE, Rabinowicz W (eds) Internet Festschrift for Peter Gärdenfors, Department of Philosophy, Lund University, Lund. http://www.lucs.lu.se/spinning/
Niiniluoto I (2011) Revising beliefs towards the truth. Erkenntnis 75(2):165–181
Oddie G (1986) Likeness to truth. Reidel, Dordrecht
Popper KR (1963) Conjectures and refutations: the growth of scientific knowledge, 3rd edn. Routledge and Kegan Paul, London
Riegler A, Douven I (2009) Extending the Hegselmann-Krause model III: from single beliefs to complex belief states. Episteme 6(2):145–163
Schurz G, Weingartner P (2010) Zwart and Franssen's impossibility theorem holds for possible-world-accounts but not for consequence-accounts to verisimilitude. Synthese 172:415–436
Zamora Bonilla J (1996) Verisimilitude, structuralism, and scientific progress. Erkenntnis 44:25–47
Zamora Bonilla J (2007) Optimal judgment aggregation. Philos Sci 74(5):813–824