Synthese (2014) 191:2383–2401
DOI 10.1007/s11229-014-0486-2

Truth approximation, belief merging, and peer disagreement

Gustavo Cevolani

Received: 15 January 2014 / Accepted: 15 January 2014 / Published online: 20 May 2014

Abstract In this paper, we investigate the problem of truth approximation via belief merging, i.e., we ask whether, and under what conditions, a group of inquirers merging together their beliefs makes progress toward the truth about the underlying domain. We answer this question by proving some formal results on how belief merging operators perform with respect to the task of truth approximation, construed as increasing verisimilitude or truthlikeness. Our results shed new light on the issue of how rational (dis)agreement affects the inquirers' quest for truth. In particular, they vindicate the intuition that scientific inquiry, and rational discussion in general, benefits from some heterogeneity in opinion and interaction among different viewpoints. The links between our approach and related analyses of truth tracking, judgment aggregation, and opinion dynamics, are also highlighted.

Keywords Verisimilitude · Truthlikeness · Truth approximation · Belief merging · Peer disagreement · Theory change · Marketplace of ideas

1 Introduction

When, and to what degree, may rational disagreement and discussion among inquirers help them in approaching the truth about the target domain? In his classical essay *On liberty*, John Stuart Mill defends freedom of expression by emphasizing the truth-conducive character of discussion and interaction among different opinions or theories (Mill 1859, pp. 41 and 95, italics added):

G. Cevolani (✉)
Department of Philosophy and Education, University of Turin, Turin, Italy
e-mail: g.cevolani@gmail.com

> [...T]he only way in which a human being can make some *approach to knowing the whole of a subject*, is by hearing what can be said about it by persons of every variety of opinion, and studying all modes in which it can be looked at by every character of mind. [...S]ince the general or prevailing opinion on any subject is rarely or never *the whole truth*, it is only by the collision of adverse opinions that the remainder of the truth has any chance of being supplied.

Already in his *Principles of political economy* (Mill 1848, III.17.5, p. 594), describing the "intellectual and moral advantages of commerce"—whose importance, according to him, is even greater than the economical ones—Mill defends diversity of opinion as a fundamental force driving cognitive and moral progress:

> It is hardly possible to overrate the value, in the present low state of human improvement, of placing human beings in contact with persons dissimilar to themselves, and with modes of thought and action unlike those with which they are familiar. [...] Such communication has always been, and is peculiarly in the present age, one of the primary sources of progress.

Such Millian insights are arguably the main modern source of the metaphor depicting scientific inquiry, and rational discussion in general, as a "marketplace of ideas"—a metaphor widely discussed also in the philosophy and the sociology of science (Zamora Bonilla 2012).

In this paper, we focus on Mill's idea that "variety of opinion" favors truth approximation since a community of inquirers benefits from (the outcome of) rational disagreement and discussion. When does a group of rational agents, having different and maybe incompatible opinions or theories, make progress toward the truth by merging together their individual theories into an unique, "collective" theory representing the opinion of their (scientific) community?

This question raises two different but related problems. The first is how a group of "recognized epistemic peers"—i.e., of agents who trust each other and judge each other as reliable as themselves on the relevant matter—should combine their beliefs in order to obtain a theory which is consistent and represents their individual opinions in the most faithful way. The second problem is how to assess whether the "new", collective theory is closer to the truth, conceived as the main aim of inquiry, than each of the "old", individual theories. This latter problem has been thoroughly discussed by philosophers of science interested in the analysis of truth approximation as increasing verisimilitude or truthlikeness (Oddie 2008; Niiniluoto 1987; Kuipers 2000). The former problem has attracted much attention both in the fields of logic and artificial intelligence (AI), and in (formal) social epistemology and philosophy of science. Recent approaches include at least formal theories of "belief merging" (Konieczny and Pino Pérez 2011) in AI, analyses of "judgment aggregation" (List 2012; Zamora Bonilla 2007; Pigozzi 2006) in economics and formal epistemology, simulation-based accounts of "debate dynamics" (Betz 2013) and of "opinion dynamics" (Riegler and Douven 2009) in philosophy of science, and epistemological accounts of "peer disagreement" (Frances 2010). However, not much work has been done, so far, on the connection between the two problems just described.

In the following, we investigate the problem of truth approximation via belief merging, in the attempt of reconstructing the missing link between these two notions and vindicating Mill's intuitions about cognitive progress through rational disagreement. In Sect. 2, we introduce the essential elements of the theory of belief merging, that aims at specifying how different inquirers should resolve the disagreement among them. The basic idea is that the agents will merge together their beliefs by agreeing on a collectively accepted theory which is as "close" or "similar" as possible to each of the original individual theories. In Sect. 3, we briefly present the post-Popperian theories of verisimilitude, specifying what does it mean, for a given theory, to be closer to the whole truth about the target domain than another theory. The resulting notion of truth approximation allows us to formally investigate the question whether belief merging promotes cognitive progress about the domain. We offer a negative answer to this question, proving that, in general, belief merging does not track truth approximation. In order to obtain more positive results, in Sect. 4 we focus on the so-called "basic feature" approach to verisimilitude (Cevolani et al. 2011, 2013), which allows for a simple definition of both the verisimilitude and the merging of "conjunctive" theories, construed as conjunctions of basic statements (atomic and logically independent sentences, and their negations) about the world. Within such approach, straightforward definitions of agreement and (weak and strong) disagreement can be given, and the conditions under which belief merging tracks truth approximation can be rigorously analyzed (Sect. 5). We conclude, in Sect. 6, by hinting at some natural extensions and generalizations of our analysis, and discussing its links with other related accounts in the literature.

2 Belief change, belief merging, and peer disagreement

The problem of consistently aggregating, combining or merging pieces of information coming from different sources arises in many areas. Suppose that a policy maker has to take an important and urgent decision, and asks for advice a panel of (scientific) experts. Apparently, all experts are equally reliable and well-informed but still they disagree about some crucial aspects of the problem under discussion. How should the policy maker combine the experts' contradictory opinions into one consistent piece of information, on which the decision can be based? Problems of this kind motivated, already from the early Sixties, the development of a number of methods for aggregating the opinions of different individuals; one of the oldest and most widely used, especially in forecasting in the social sciences and in futures studies, is the so-called Delphi method, which is based on the iterative comparison and revision of individual opinions until a consensus emerges (Linstone and Turoff 1975). More recently, similar problems have arisen in many applications in AI, for instance the design of expert and multi-agent systems, when different sets of data have to be aggregated into an unique, consistent database.

The theory of belief merging is a logic-based account of how two or more rational agents should combine their beliefs or theories without giving priority to any of them and in such a way that the resulting theory is consistent. In this account, belief merging is construed as a special kind of belief change in the tradition of the AGM theory of belief revision—so called after Alchourrón et al. (1985)—which is the dominant account of belief change in AI and formal epistemology. The AGM theory studies how the beliefs of an ideally rational agent should change in response to certain inputs coming from some information source. The source is assumed to be reliable, in the sense that, if an input contradicts the agent's beliefs, the agent will always give up some of them in order to consistently incorporate the input in his belief state. In other words, "new" information has always priority over the "old" one; for this reason, AGM belief change is known as "prioritized" belief change (Hansson 2011, sections 6.3–6.6). Belief merging can be seen as the case of belief change where each source is equally reliable, and hence no piece of information has priority over any other (Konieczny and Pino Pérez 2011, pp. 239–240).

Both belief revision and belief merging follow a basic methodological principle, according to which the agent should always perform a "minimal change" of his beliefs. In the case of belief revision, this means that, after the change has occurred, his new theory should be as close as possible to the old one. In belief merging, minimal change amounts to requiring that the collective theory should be as close as possible to each of the individual ones. This intuitive idea can be made precise as follows.

Let us consider k different agents or rational inquirers (with $k \geq 2$) and suppose that the domain under inquiry ("the world") is described by a finite propositional language $\mathscr{L}_n$, with n logically independent atomic sentences $a_1, \ldots, a_n$. The beliefs or theory of each agent at any given time will be represented by the strongest proposition about the world accepted by the agent at that time. A "basic sentence" or "literal" of $\mathscr{L}_n$ is an atomic sentence or its negation (hence there are $2n$ basic sentences). Constituents $c_1, \ldots, c_q$ are the maximally informative conjunctions of $\mathscr{L}_n$; i.e., each constituent is a conjunction of n basic sentences, one for each atomic sentence. There are $q = 2^n$ constituents, which, intuitively, describe the possible worlds expressible in $\mathscr{L}_n$. It is well known that each non-contradictory statement T in $\mathscr{L}_n$ can be expressed, in normal form, as the disjunction of the constituents entailing it, or, equivalently, as the set of possible worlds in which the statement is true. This set of constituents is called the "range" of T, and will be denoted by $\mathscr{R}_T$; it represents the class of "possibilities" compatible with the agent's theory T. Finally, we shall assume that an adequate *distance measure* Δ is defined on the class of constituents. In most applications, the natural choice is to take Δ as the Hamming (also known as Dalal) distance between constituents: i.e., $\Delta(c_i, c_j)$ is the number of basic sentences on which c_i and c_j differ. The minimum distance $\Delta_{min}(T, c_i)$ between theory T and constituent c_i can then be defined as $\min_{c_j \in \mathscr{R}_T} \Delta(c_j, c_i)$; note that $\Delta_{min}(T, c_i) = 0$ if and only if c_i is in the range of T.

When an agent with theory T receives an input A, represented by a sentence of $\mathscr{L}_n$, T has to be revised by A in the most conservative way possible. This amounts to define the *revision* of T by A as the theory identified by the constituents (in the range) of A which are the closest to T. Formally:

$$T * A \overset{df}{=} \bigvee \{c_i \in \mathscr{R}_A : \Delta_{min}(T, c_i) \text{ is minimal}\} \quad (1)$$

i.e., such that $\Delta_{min}(T, c_i) \leq \Delta_{min}(T, c_j)$ for all $c_j \in \mathscr{R}_A$ (Niiniluoto 2011, p. 171). Note that, if T and A are logically compatible, then their ranges have a non-empty intersection, which obviously contains the closest constituents of A to T; thus, in this case $T * A$ is simply the conjunction of T with A, and is called the *expansion* of T by A.

In the case of belief merging, we need to combine k theories $T_1, \ldots, T_k$ representing the beliefs of the corresponding agents. Since two different agents may share exactly the same beliefs, it is useful to introduce the notion of a "profile", that is the so called multiset (or bag) $E = [T_1, \ldots, T_k]$ containing these theories.[1] The theory resulting from merging all theories in E will be denoted by E°; if $k = 2$, we use infix notation and write $T_1 \circ T_2$. Since E° must represent as faithfully as possible the beliefs of each agent, E° must be as close as possible to each T_i. Let us define the distance between a profile E and a constituent c_i as the sum of the minimum distances of all T_i from c_i:

$$\Delta(E, c_i) \overset{df}{=} \sum_{T_j \in E} \Delta_{min}(T_j, c_i) \tag{2}$$

Then, E° can be defined as the theory such that the constituents in its range minimize the distance from E:

$$E^\circ \overset{df}{=} \bigvee \{c_i : \Delta(E, c_i) \text{ is minimal}\} \tag{3}$$

i.e., such that $\Delta(E, c_i) \leq \Delta(E, c_j)$ for all c_j. The merging operator defined above, as based on the "sum" distance defined in (2), is a so-called "majority" merging operator, as opposed to so-called "arbitration" operators. The underlying idea is that majority merging tries to satisfy a maximum of agents, or to minimize "global" dissatisfaction, while arbitration merging tries to satisfy each agent to the best possible degree, or to minimize "individual" dissatisfaction (Konieczny and Pino Pérez 2002, p. 774). Definitions (2)–(3) give a majority operator since what is minimized is the sum of the distances of all the theories in the profile, considered as a whole, instead of the distance of each theory separately. Definitions of the distance between a profile and a constituent different from (2) provide alternative kinds of merging operators, with different formal features (for a survey, see Konieczny and Pino Pérez 2011, 2002).[2]

A couple of examples concerning only two theories T_1 and T_2 may illuminate the definition in (3). First, if the beliefs T_1 and T_2 of two agents are compatible, then $T_1 \circ T_2$ is just the conjunction of their beliefs. Second, even if T_1 is the *negation* of T_2, still the two agents may find an agreement on how to merge together their beliefs. For instance, suppose that p and q are basic sentences of $\mathscr{L}_n$. If T_1 is $p \wedge q$ and T_2 is its negation $\neg p \vee \neg q$, one can check that $T_1 \circ T_2$ is $p \vee q$. In this sense, belief merging can "resolve" even cases of serious disagreement. However, if T_1 is $p \wedge q$ and T_2 is its *reversal*—i.e., the conjunction $\neg p \wedge \neg q$ of the negations the basic sentences in T_1—then the result of merging these theories is just the disjunction of all constituents, i.e., the tautology.

[1] In a multiset, each element can appear more than once: thus, while the two *sets* $\{T_1, T_2, T_2\}$ and $\{T_1, T_2\}$ are the same, the two *multisets* $[T_1, T_2, T_2]$ and $[T_1, T_2]$ are different. In both cases the order of elements is irrelevant.

[2] The discussion of arbitration merging, and of other kinds of operators, in connection with truth approximation deserves further research, but has to be left for another occasion.

In other words, if two agents differ on the evaluation of every single basic sentence they believe, they cannot agree on any factual matter (see the "Appendix" for relevant definitions and proofs).

As noted by an anonymous referee, the examples above highlight a relevant assumption of the present model, i.e., the outspoken willingness of all agents involved to merge their beliefs in a way or another. This obviously appears as an unrealistic feature of the model, if this is meant to apply to concrete cases of discussion among real agents, since it implies that no agent will never refuse to change his mind. In any case, such an assumption is shared with several other accounts of belief merging, like the so-called Hegselmann–Krause model of opinion dynamics (cf. Riegler and Douven 2009, and Section 6 below). Thus, it seems that all such models are best seen as accounts of how to combine pieces of information together (regardless of their sources and motivations) rather than as realistic models of rational discussion (but see Betz 2013).

3 Truth approximation via belief change

Intuitively, theory T is verisimilar (or truthlike) if it is close or similar to the whole truth about a given target domain. Thus, T 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 in order to defend the idea that progress can be explained in terms of the increasing verisimilitude of scientific theories (Popper 1963, Ch. 10). 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.[3]

More formally, if $\mathscr{L}_n$ is used to describe the underlying domain, then "the (whole) truth" is represented by the only true constituent $c_\star$ of $\mathscr{L}_n$, which is the most informative true description of the actual world within $\mathscr{L}_n$. (We shall assume, without loss of generality, that $c_\star \equiv a_1 \wedge \cdots \wedge a_n$—i.e., the conjunction of the unnegated basic sentences of $\mathscr{L}_n$—is the whole truth about the world.) The (degree of) verisimilitude $Vs(T)$ of theory T can then be defined as the closeness of T to $c_\star$. The underlying idea is that increasing verisimilitude is a game of finding interesting truths and excluding serious falsehoods (cf. Niiniluoto 1987, p. 242): i.e., T is (highly) verisimilar if $\mathscr{R}_T$ contains possibilities which are close to $c_\star$ and excludes possibilities far from $c_\star$ (cf. Fig. 1). Given an adequate (normalized) measure $\Delta(T, c_\star)$ of the distance of T from the truth, $Vs(T)$ can then be defined as as $1 - \Delta(T, c_\star)$.[4] Most measures of

[3] After Miller (1974) and Tichý (1974) independently proved that, on the basis of Popper's own definition of verisimilitude, a false theory can never be closer to the truth than another (true or false) theory, such authors as Niiniluoto (1987, 1999b), Kuipers (1987, 2000), Oddie (1986), Festa (1987, 2007) and Schurz and Weingartner (1987, 2010) developed a number of post-Popperian theories of verisimilitude that succeed in avoiding the problems encountered by Popper's definition. A survey of the history of theories of verisimilitude is provided by Niiniluoto (1998); see also Oddie (2008, 2013) for a comparison and an assessment of different accounts, and Cevolani and Tambolo (2013) for an introduction to the verisimilitudinarian approach to scientific progress.

[4] Two examples of such distance measures are the "average" measure proposed by Oddie (1986)—who defines $\Delta(T, c_\star)$ as the average distance of the constituents in $\mathscr{R}_T$ from $c_\star$—and the "min-sum" measure

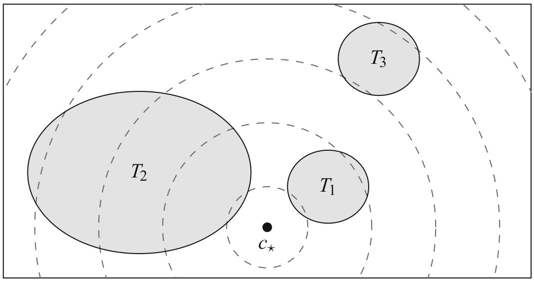

Fig. 1 Verisimilitude as approximation to the truth. The *rectangle* represents the set of all constituents (possible worlds); the dashed concentric "spheres" represent the constituents' distances from the true one, $c_\star$. The *gray areas* represent the ranges of three false theories T_1, T_2 and T_3; intuitively, T_1 is the most verisimilar theory, including possibilities close to the truth and excluding possibilities far from the truth

verisimilitude (with the notable exception of Oddie's favored measure) satisfy the Popperian requirement that verisimilitude co-varies with logical strength among true theories (cf. Niiniluoto 1987, pp. 186–187, 233–236):

$$\text{If } T_1 \text{ and } T_2 \text{ are true and } T_1 \text{ entails } T_2, \text{ then } Vs(T_1) \geq Vs(T_2) \tag{4}$$

In other words, among true theories, the stronger the better. This condition does not hold, of course, among false theories, since logically stronger falsities may well lead us farther from the truth: if T_1 and T_2 are both false, the more verisimilar theory will be the one making less or less serious errors.

Under what conditions does belief change track truth approximation, in the sense that it leads our theories closer to the truth? This problem was first raised by Niiniluoto (1999a) with respect to AGM belief revision, and has been recently discussed by a number of scholars. The preliminary answer is mainly negative, since apparently no general conditions guarantee that the revision of T even by *true* inputs A leads to a theory $T * A$ which is more verisimilar than T.[5] In particular, Niiniluoto (1999a) proved that the only "safe case" of truth approximation through belief revision is when a true theory is revised by a true input:

$$\text{If both } T \text{ and } A \text{ are true, then } Vs(T * A) \geq Vs(T) \tag{5}$$

Note that, since both T and A are true, they are compatible, and hence $T * A$ is just $T \wedge A$, which is true and stronger than T. Thus, result (5) holds for all verisimilitude measures which satisfy the Popperian requirement (4) according to which, among truths, verisimilitude increases with logical strength. Unfortunately, expanding true

Footnote continued 4
proposed by Niiniluoto (1987)—defined as a weighted sum of the minimum distance of T and of the normalized sum of all distances of the constituents in $\mathscr{R}_T$ from $c_\star$.

[5] A survey of the main results obtained so far can be found in the introductory essay by Kuipers and Schurz (2011) to a special issue of *Erkenntnis* entirely devoted to the topic of *Belief Revision Aiming at Truth Approximation*; see in particular the contributions by Niiniluoto (2011), Cevolani et al. (2011), Kuipers (2011), and Schurz (2011). See also Cevolani et al. (2013) and Cevolani (2013) for further work in this direction.

theories by true inputs is a relatively uninteresting case of cognitive progress, corresponding to the very naïve view according to which an agent approaches the truth simply by accumulating more and more truths about the world—a view which is unacceptable as an account of belief change aiming at truth approximation (Cevolani and Tambolo 2013, section 5). Indeed, a real agent (like a scientist) will normally entertain also false beliefs; and one of the reasons for revising these beliefs is exactly the attempt to correct them and getting closer to the truth.

When the theories of different experts are merged together, one may hope to obtain a collective theory which is more verisimilar than each of the individual theories, and so constitutes an instance of cognitive progress. Thus, our question is: under what conditions does belief merging track truth approximation? That is: given a profile $E = [T_1, \ldots, T_k]$, under what conditions $Vs(E^\circ) \geq Vs(T_i)$ for all T_i in E? The answer is similar to that given above for the case of belief revision. In fact, one can easily check that:

$$\text{If } T_1, \ldots, T_k \text{ are all true, then } Vs(E^\circ) \geq Vs(T_i) \text{ for all } T_i \text{ in } E \qquad (6)$$

This means that if the agents have only true beliefs about the world, then merging them together leads the collective theory at least as closer to the truth as any of their individual theories. The proof of (6) is straightforward, since if $T_1, \ldots, T_k$ are all true, they are compatible and then E° is just their conjunction, which is true. It follows that E° will be at least as verisimilar as each T_i for all verisimilitude measures Vs satisfying the Popperian requirement (4).

However, even if just one agent holds some false beliefs, then merging his theory with the others in the profile can result in a collective theory which is much farther from the truth than many (or even all) of the individual theories. Still worse, this may happen even if all other agents, except the one accepting a false belief, only hold *true* beliefs. The reason is that when a false theory is combined with a true one, the resulting theory will be generally false and possibly less verisimilar than both. An example will clarify this point.[6]

Example 1 Suppose that two agents accept, respectively, theories $T_1 = a_1 \rightarrow \neg a_2 \wedge \cdots \wedge \neg a_n$ and $T_2 = a_1$. Note that T_1 is false while T_2 is true. Since T_1 and T_2 are compatible, their merging $T_1 \circ T_2$ will be their conjunction $a_1 \wedge (a_1 \rightarrow \neg a_2 \wedge \cdots \wedge \neg a_n)$, which amounts to a constituent which is very far from the true one $c_\star \equiv a_1 \wedge \cdots \wedge a_n$. Most verisimilitude theorists would agree that, in this case, $T_1 \circ T_2$ is *less* verisimilar than both T_1 and T_2. In fact, while both T_1 and $T_1 \circ T_2$ are false, the latter entails more, and more "serious", falsehoods than the former. And while T_2 and $T_1 \circ T_2$ agree on the truth of a_1, the latter adds to this true belief a number of false beliefs on the remaining basic sentences of $\mathscr{L}_n$.

As noted by an anonymous referee, results of this kind can be used to suggest that, in general, agents should not merge their beliefs too quickly, if they aim at truth

[6] Essentially the same kind of examples can be used to show why the revision of a false theory by true inputs may be less verisimilar than the original theory; cf. Schurz (2011, p. 210, example 4) and Kuipers and Schurz (2011, p. 154).

approximation, since this may lead them farther from their cognitive goal. As noted in Sect. 2, however, in the present framework agents are simply assumed to always merge their beliefs, so the issue is whether merging is or not instrumental in approaching the truth. In this connection, belief merging, like AGM belief revision, does not track truth approximation in general. To be sure, it is well possible that merging together different *false* theories leads to a collective theory which is true, or closer to the truth than each of the original theories. However, there are apparently no general conditions under which this is guaranteed.

4 The basic feature approach to truth approximation and merging

In order to investigate the relationships between truth approximation and AGM belief revision, Cevolani et al. (2011) applied the so called "basic feature" approach to verisimilitude, henceforth "BF-approach" (see also Cevolani et al. 2013). They showed that, when attention is restricted to a specific kind of theories, labeled "conjunctive", the conditions under which belief revision tracks truth approximation can be made explicit in a fairly general way. In this section, we pursue this strategy and apply the BF-approach to belief merging in order to obtain more satisfactory results than the one in (6) in the previous section.

The main idea underlying the BF-approach is that the verisimilitude of theory T can be interpreted in terms of the balance of true and false information conveyed by T about the "basic features" of the target domain. When $\mathscr{L}_n$ is used to describe this domain, these basic features are described by the basic sentences (literals) of $\mathscr{L}_n$. This amounts to assume that the agents' beliefs are not represented by arbitrary theories (statements) of $\mathscr{L}_n$, but by the so called "conjunctive" or "basic" theories ("c-theories", for short) of $\mathscr{L}_n$, which are conjunctions of m basic sentences concerning m different atomic sentences (with $m \leq n$). Note that, if $m = 1$ then the c-theory is just a basic sentence, while if $m = n$ then it is a constituent. For this reason, c-theories can also be called the "quasi-constituents" of $\mathscr{L}_n$ (Oddie 1986, p. 86), since each c-theory is, so to speak, a "fragment" of a constituent of $\mathscr{L}_n$ (indeed, of all constituents compatible with that c-theory). One can check that there are exactly 3^n c-theories expressible in $\mathscr{L}_n$, including the 2^n constituents and the tautological c-theory with $m = 0$.

Each conjunct of a c-theory T will be called a "(basic) claim" of T (about the world). The key intuition underlying the BF-approach is that T is highly verisimilar if T makes many claims about the basic features of the world, as described by $c_\star$, and many of those claims are true. This is a way of making sense of Popper's original intuition that verisimilitude "represents the idea of approaching comprehensive truth. It thus combines truth and content" (Popper 1963, p. 237). Calling each true claim of T a "match", and each false claim a "mistake", of T, we may say that T is highly verisimilar if T makes many matches and few mistakes about $c_\star$. More formally, let $t(T, c_\star)$ and $f(T, c_\star)$ denote, respectively, the set of matches and of mistakes of T. Moreover, let $cont_t(T, c_\star) = |t(T, c_\star)|/n$ and $cont_f(T, c_\star) = |t(T, c_\star)|/n$ be the normalized number of matches and of mistakes of T. In order to assess the verisimilitude of T, $cont_t(T, c_\star)$ may be construed as the overall *reward* attributed to the matches of T and $-cont_f(T, c_\star)$ as the overall *penalty* attributed to the mistakes of T. A "contrast measure" of verisimilitude is a weighted average of the reward due to T's matches and of the penalty due to T's mistakes Cevolani et al. (2011, p. 188):

$$Vs_\phi(T) \stackrel{df}{=} cont_t(T, c_\star) - \phi cont_f(T, c_\star) \tag{7}$$

where $\phi > 0$. Intuitively, different values of ϕ reflect the relative weight assigned to truth and falsity, i.e., to the matches and mistakes of T. Note that, if $\phi = 0$ and T is a constituent (i.e., it makes n basic claims), then $Vs_\phi(T)$ reduces to the normalized Hamming (or Dalal) distance between propositional constituents. It is easy to check that Vs_ϕ satisfies the Popperian requirement (4) according to which, among truths, verisimilitude increases with logical strength.[7]

In order to apply the BF-approach to the analysis of belief merging, let us introduce some (more or less) standard terminology. We shall say that a given basic sentence b of $\mathscr{L}_n$ is "accepted" in T when b is a claim of T, is "rejected" in T when $\neg b$ is a claim of T, and is "indeterminate" in T when neither b nor $\neg b$ is a claim of T, i.e., when b is neither accepted nor rejected in T (cf. Gärdenfors 1988, p. 22). In these three cases, we shall also say that (the agent holding) T accepts, rejects, and "suspends the judgment on" b, respectively. When c-theory T_1 is compared with c-theory T_2, three "parts" of T_1 can be defined (Cevolani et al. 2011, p. 193):

(i) the "overlapping" part of T_1 w.r.t. T_2, i.e., the conjunction $O_{T_1T_2}$ of the claims of T_1 also accepted in T_2;
(ii) the "conflicting" part of T_1 w.r.t. T_2, i.e., the conjunction $C_{T_1T_2}$ of the claims of T_1 rejected in T_2; and
(iii) the "excess" part of T_1 w.r.t. T_2, i.e., the conjunction $X_{T_1T_2}$ of the claims of T_1 which are indeterminate in T_2.

The overlapping, conflicting, and excess parts of T_2 w.r.t. T_1 are defined in the same way (see Fig. 2). If T_1 and T_2 are the theories of two epistemic peers, their (dis)agreement can be construed in terms of the relevant parts of T_1 and T_2. In particular, we shall say that (the two agents holding) T_1 and T_2:

(i) "agree" on each atomic sentence appearing in $O_{T_1T_2}$ (or, equivalently, in $O_{T_2T_1}$);
(ii) "(strongly) disagree" on each atomic sentence appearing in $C_{T_1T_2}$ (or, equivalently, in $C_{T_2T_1}$); and
(iii) "weakly disagree" or "differ (in opinion)" on each atomic sentence appearing in $X_{T_1T_2}$ and in $X_{T_2T_1}$.

In words, T_1 and T_2 agree on their common claims, i.e., on each atomic sentence that both accept or reject; (strongly) disagree on their conflicting claims, i.e., on each atomic sentence that one accepts and the other rejects; and weakly disagree on each atomic sentence that is accepted or rejected in one theory, and indeterminate in the other. With reference to Fig. 2, for example, we shall then say that T_1 and T_2 agree on p_2, disagree on p_3 and p_4, and differ in opinion on p_1 and p_5.

[7] Indeed, Vs_ϕ satisfies the stronger condition that, among true theories, the one with the greater number of matches is more verisimilar than the other; i.e., if T_1 and T_2 are true and $cont_t(T_1, c_\star) > cont_t(T_2, c_\star)$ then $Vs_\phi(T_1) > Vs_\phi(T_2)$.

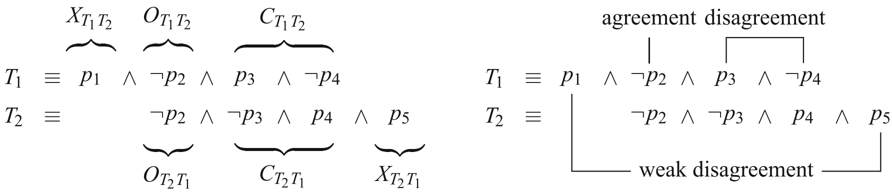

Fig. 2 Agreement, disagreement and weak disagreement (difference of opinion), defined in terms of the overlapping, conflicting, and excess parts of T_1 and T_2

The BF-approach provides a simple common framework to study both truth approximation and belief merging. In this connection, one may note that, in the new terminology just introduced, the verisimilitude of T is defined by (7) in terms of the accepted (basic) truths and of accepted (basic) falsehoods (or rejected truths) of T, or, equivalently, in terms of the agreement and (strong) disagreement between T and the truth $c_\star$. As far as belief merging is concerned, once can check that, given a profile $E = [T_1, \ldots, T_k]$ of k c-theories, the collective theory resulting by merging them is just (see the "Appendix" for a proof):

$$E^\circ = \bigwedge \{b_i : |\{T_j \in E : T_j \text{ accepts } b_i\}| > |\{T_j \in E : T_j \text{ rejects } b_i\}|\} \tag{8}$$

In words, E° is a new c-theory such that, for each of its claims, there are more agents accepting it than rejecting it. Note that the above result offers a straightforward interpretation of the process of belief merging in terms of a voting procedure. For each atomic sentence a_i of $\mathscr{L}_n$, each agent (in some order) is asked whether he accepts, rejects, or suspends the judgment on, that sentence. If some agent accepts (rejects) a_i, and if all other agents suspends the judgment on a_i, then a_i will be accepted (rejected) in E°. If some agent accepts (rejects) a_i and some other rejects (accepts) it, a_i will be accepted (rejected) in E° if and only if the majority of agents accepts (rejects) a_i (the vote of those who suspend the judgment is irrelevant). Finally, if all agents suspend the judgment on a_i, a_i will be excluded by the collective theory E°. In short, the merging of $T_1, \ldots, T_k$ is the conjunction of basic sentences which are accepted by some of the agents and on which the majority agrees.

Note that, if the number k of agents is even, and exactly half of them accept a_i and the other half rejects it, a_i will not be accepted in the collective theory, since no strict majority is available, so to speak, to solve the disagreement. It follows that, if there are only two agents, and they disagree on some a_i, a_i will be indeterminate in the collective theory. In other words, the merging of c-theories T_1 and T_2 is just:

$$T_1 \circ T_2 = O_{T_1T_2} \wedge X_{T_1T_2} \wedge X_{T_2T_1} \tag{9}$$

i.e., the conjunction of claims on which T_1 and T_2 agree or weakly disagree.[8] With reference to Fig. 2, for example, the merging of T_1 and T_2 is $T_1 \circ T_2 \equiv p_1 \wedge \neg p_2 \wedge p_5$.

[8] Note the difference between the merging and the AGM *revision* of c-theories T_1 and T_2, which amounts to $T_1 * T_2 = T_2 \wedge X_{T_1T_2}$ (Cevolani et al. 2011, p. 193).

5 Truth approximation and peer (dis)agreement

Suppose that k agents merge their c-theories $T_1, \ldots, T_k$ into a collective theory E°. When is E° more verisimilar than each T_i in E, i.e., when have the agents as a group made a cognitive progress toward the truth about the world? A couple of toy examples will be sufficient to show that no general answer is readily available to this question. Indeed, it is well possible that E° is (much) more verisimilar than each T_i, as the following example shows:

Example 2 Let us consider three agents with theories $T_1 \equiv a_1 \wedge \neg a_2 \wedge a_3 \wedge \neg a_4$, $T_2 \equiv \neg a_2 \wedge \neg a_3 \wedge a_4 \wedge a_5$ (these are the two theories in Fig. 2), and $T_3 \equiv a_2 \wedge a_3 \wedge a_4 \wedge a_6$. By (8), their merging is the true c-theory $E^\circ = a_1 \wedge a_2 \wedge a_3 \wedge a_4 \wedge a_5 \wedge a_6$, which is more verisimilar than each of T_1, T_2, and T_3. In fact, one can easily check that $Vs_\phi(E^\circ) = \frac{6}{n}$ is greater than both $Vs_\phi(T_1) = Vs_\phi(T_2) = \frac{2-2\phi}{n}$ (since $\phi > 0$) and $Vs_\phi(T_3) = \frac{4}{n}$.

Thus, an adequate combination of agreement and disagreement among the agents may be instrumental to truth approximation. However, even a small degree of disagreement is sufficient, in some cases, to block cognitive progress, as the following extreme case shows.

Example 3 Suppose that $T_1 \equiv a_1 \wedge \neg a_2 \wedge a_3 \wedge \cdots \wedge a_n$, $T_2 \equiv \neg a_1 \wedge a_2 \wedge a_3 \wedge \cdots \wedge a_n$, and $T_3 \equiv \neg a_1 \wedge \neg a_2 \wedge a_3 \wedge \cdots \wedge a_n$. By (8), their merging is $E^\circ = T_3$, i.e., the least verisimilar among the three theories.

The example above shows that even if the agents agree on most of the basic truths about the domain, merging their theories may lead most of them farther from the truth.

In short, both agreement and disagreement may favor, but also hinder, cognitive progress, depending on the specific circumstances. In this connection, the only general conclusion one can draw is that a fair degree of heterogeneity in opinion—i.e., of *weak* disagreement—among the agents is essential for the purpose of truth approximation, both to trigger change and to resolve (strong) disagreement. This is particularly clear in the two-agents case. In fact, given (9) one can immediately prove that:

$$Vs_\phi(T_1 \circ T_2) > Vs_\phi(T_1), Vs_\phi(T_2) \text{ iff } \begin{cases} Vs_\phi(X_{T_2 T_1}) > Vs_\phi(C_{T_1 T_2}) \text{ and} \\ Vs_\phi(X_{T_1 T_2}) > Vs_\phi(C_{T_2 T_1}) \end{cases} \quad (10)$$

In words, $T_1 \circ T_2$ is more verisimilar than T_1 just in case weak disagreement outweighs strong disagreement, in the sense that the (possible) increase in verisimilitude due to the addition of the excess part of T_2 outweighs the (possible) decrease due to the elimination of the conflicting part of T_1 w.r.t. T_2 from the collective theory; and similarly for T_2.

Example 4 Assume that T_1 and T_2 are the c-theories in Fig. 2. Then one can easily check that $Vs_\phi(T_1 \circ T_2) = \frac{2-\phi}{n}$ is greater than $Vs_\phi(T_1) = Vs_\phi(T_2) = \frac{2-2\phi}{n}$.

Note that, given (9), if there is no difference of opinion between T_1 and T_2, then $T_1 \circ T_2$ is just the conjunction $O_{T_1 T_2}$ of the claims on which both agents agree. Thus, absence of weak disagreement leads to the extreme conservative policy of accepting only shared beliefs, that, in turn, tends to hinder cognitive progress. This is shown by the following result (see the "Appendix" for a proof):

$$\begin{aligned}&\text{If there is no weak disagreement between } T_1 \text{ and } T_2 \text{ then:}\\ &Vs_\phi(T_1 \circ T_2) > Vs_\phi(T_1) \text{ iff } Vs_\phi(T_1 \circ T_2) < Vs_\phi(T_2)\\ &\text{for all verisimilitude measures } Vs_\phi \text{ with } \phi \geq 1.\end{aligned} \tag{11}$$

Intuitively, this means that, in most cases (i.e., for $\phi \geq 1$) the collective theory will be closer to the truth than the less verisimilar theory, but farther from the truth than the more verisimilar one. In this sense, no overall cognitive progress is made in absence of weak disagreement. Two special cases are worth noting here. First, if T_1 and T_2 totally agree, i.e., share all their claims, then $T_1 \circ T_2 = T_1 = T_2$. In this case, no agent changes his beliefs, and hence no progress is made toward the truth. Second, if T_1 and T_2 totally disagree, i.e., the one rejects each claim of the other, then $T_1 \circ T_2$ is just the tautology and amounts to suspending the judgment on all factual matters.

Such results can be extended to the general case of k agents merging their c-theories $T_1, \ldots, T_k$, as follows. Let us say that an atomic sentence a_i is "at issue" if it appears as a conjunct of at least one of these theory. Then, if all agents agree on each sentence at issue (i.e., they all accept or reject it), the merging process will result in no change of opinion and so in no cognitive progress. Moreover, if k is even and the agents totally disagree in the sense that, for each sentence at issue, exactly half of them accept it and the other half reject it, merging is "annihilating", since it leads to suspend all possible beliefs. Thus, only if the agents weakly disagree on at least some matter, it is possible that some progress is made. In sum, weak disagreement appears as a necessary (but not sufficient) condition for truth approximation.

6 Concluding remarks and further work

The results of the last section seem to vindicate the commonsense idea, going back at least to Mill, that disagreement among rational agents can trigger cognitive progress, construed as increasing approximation to the truth. In fact, both Mill's "variety of opinion" and "collision of adverse opinions"—i.e., what we called, respectively, weak and strong disagreement in Sect. 4—play a crucial role in the process of belief merging and hence of truth approximation.

In particular, when some agents combine together their beliefs (according to the belief merging model), weak disagreement or diversity of opinion between them is essential to guarantee at least the possibility of approaching the truth through belief merging. A too high degree of agreement, on the contrary, is, so to speak, paralyzing, since it may prevent the agents both from eliminating old falsehoods from their collective theory, and from accepting the new truths on which some of them disagree. In this connection, strong disagreement may be instrumental in eliminating from the final theory some falsehoods accepted in some of the individual theories, but can be annihilating if affecting too many of the agents' beliefs. While results like (10) and (11) are not easily extended to the many-agents case, they quite clearly highlight the role and relative importance of agreement and (weak) disagreement for the purpose of truth approximation.

As already noted in Sect. 1, the problems of belief merging and peer disagreement have recently received a great deal of attention. While a detailed comparison of our approach with other accounts in the literature is beyond the scope of this paper, in the following we highlight some of the main connections with related discussion in the areas of judgment aggregation, debate dynamics, and opinion dynamics. First, however, we suggest how to improve and extend the present approach in order to obtain a better understanding of more realistic cases of truth approximation through belief merging.

Estimated verisimilitude In the previous sections we have discussed the relationships between belief merging and truth approximation as if the truth values of the agents' theories were known. While this assumption is useful for the purposes of conceptual analysis, as emphasized by Niiniluoto (2011, p. 177) practical rules of belief change cannot be directly based upon absolute truth values, since in typical situations these are simply unknown. In other words, in real-life situations, including everyday reasoning and scientific inference, one has to operate under uncertainty and risk. Niiniluoto (1987, p. 269) has proposed to address this problem by defining a measure of "expected verisimilitude" based on an underlying epistemic probability distribution P defined over the set of the constituents $c_1, \ldots, c_q$ of $\mathscr{L}_n$ and expressing the rational degrees of belief in the truth of each alternative c_i given the available evidence e (see also Oddie 1986, p. 180). The expected degree of verisimilitude of T is then calculated by summing up the verisimilitude of T in each state multiplied by the corresponding probability given e:

$$EVs(T|e) \stackrel{df}{=} \sum_{c_i} Vs(T, c_i) P(c_i|e)$$

This definition provides a measure $EVs(T|e)$ of the degree of estimated closeness to the truth of theory T, given the evidence. It also allows us to compare two different theories T_1 and T_2 with regard to their estimated closeness to the truth, and to say, e.g., that theory T_1 seems, in view of e, more verisimilar than theory T_2, when $EVs(T_1|e) > EVs(T_2|e)$.

When approximation to the truth is construed as the aim of inquiry, expected verisimilitude appears as a suitable candidate notion in order to study and define operations of belief change (including belief merging) in the service of truth approximation (cf. Niiniluoto 2011, Sec. 7, in particular p. 178). In this connection, one might assume that all agents in the group share the same evidence e, and ask whether $EVs(E^\circ|e)$ is greater than $EVs(T_i|e)$ for each agent i, i.e., whether belief merging increases the expected verisimilitude of the collective theory. Another suggestion would be to treat the agents' beliefs as evidence, apply the merging operation in order to resolve disagreements among them, and, finally, accept the theory T which maximizes the value of $EVs(T|E^\circ)$, i.e., the theory which is estimated as the most verisimilar given the agent's opinions. This is perhaps the most sensible choice for a decision maker who has to act on the basis of some experts' advice. Further work is needed, however, in order to check under what conditions the theory so obtained is (expected to be) more verisimilar than the agents' individual theories, so guaranteeing cognitive progress about the domain.

Truth approximation vs. truth tracking In this paper, we investigated truth approximation via belief merging, asking whether combining the beliefs of different agents increases the (expected) verisimilitude of their final theory. A different question is the following: under what assumptions concerning the reliability of the agents 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 (cf. Konieczny and Pino Pérez 2011, pp. 259–262). Much research on the truth tracking problem has been motivated by Condorcet's jury theorem in political science, which 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 theorem has been recently studied, and generalized, in the field of belief merging (Everaere et al. 2010) in order to compare the truth tracking ability of different merging operators. Interestingly, the merging operator defined in (3) does not satisfy the truth tracking condition as defined by Everaere et al. (2010, see proposition 4); moreover, further work is needed to clarify the logical and conceptual relationships between the truth tracking and the truth approximation issue.

Belief merging, judgment aggregation, and truth tracking Belief merging and truth tracking are both related to the research program on "judgment aggregation", that is motivated by (the generalization of) another classical result due to Condorcet, the so-called "discursive dilemma" (in social choice theory) or "doctrinal paradox" (in jurisprudence) (List 2012, pp. 180 ff.). The problem arises with groups of decision makers (expert panels, legal courts, boards, etc.) who vote by majority on a set of logically related propositions aiming at producing a collective decision on them. It may then happen that, while all individual judgments concerning the propositions are consistent, the collective judgment obtained by majority voting is not. Pigozzi (2006) first noted the connections between judgment aggregation and belief merging, and applied the methods developed in the latter field to address the paradoxes arising in the former. More recently, Hartmann and Sprenger (2012) have studied the problem of truth tracking from the perspective of judgment aggregation. The relations between belief merging, judgment aggregation and verisimilitude are worth studying, starting from the crucial role that distance-based methods of analysis play in all these three areas.

Debate dynamics, opinion dynamics, and truth approximation Finally, both truth approximation and truth tracking can be studied, as recently shown by Douven and Kelp (2011) and Betz (2013), within different simulation-based accounts of opinion dynamics and rational consensus formation. In this field, one studies how the beliefs of a community of agents evolve in time, assuming that, at each step, an agent's beliefs are determined both by his own beliefs and by the beliefs of his peers at the previous step. Two well-known models of this kind are the Lehrer–Wagner model and the Hegselmann–Krause model, with their modifications and extensions (Riegler and Douven 2009). Using computer simulations, one can study the evolution over time of the beliefs of the community, investigating under what conditions such beliefs converge on a overall consensus. A more fine-grained approach, taking into account the "micro-dynamics" of the debate going on within the community itself, is provided by the theory of dialectical structures, developed by Betz (2013) in the field of formal argumentation theory. All these approaches can be construed as contributions to the research program on "veristic" social epistemology (Goldman 1999), i.e., the systematic study of practices of rational belief change and consensus formation with respect to their tendency to disseminate true beliefs in communities of truth-seeking agents. Douven and Kelp (2011) and Betz (2013) analyze the truth-conduciveness of these practices with regard to, respectively, the Hegselmann–Krause model and the theory of debate dynamics.

Remarkably, as emphasized by Betz (2013, p. 40, note 7), this latter theory presents some very interesting connections with the BF-approach of Sect. 4. In the theory of debate dynamics, the beliefs of each agent are essentially represented by a constituent of the underlying language, i.e. as a "complete" c-theory; thus, both the verisimilitude of a theory and its distance from other theories can be expressed in terms of the (normalized) Hamming distance between constituents, i.e., as a special case of the contrast measure of verisimilitude defined in (7). Agents then minimally change their beliefs or theories in order to make them compatible with a shared "dialectical structure", i.e., the set of interconnected arguments (deductively valid inferences from some premises to a conclusion) proposed so far by the participants to the debate. In this richer framework, one can analyze the role of argumentation both in reaching a consensus among the agents and in tracking down the truth; moreover, such an analysis is more fine-grained than, but in principle perfectly compatible with, the one proposed in the present paper. Finally, the BF-approach can be used to extends the theory of debate dynamics by allowing agents to maintain also "incomplete" beliefs about the world, i.e., proper c-theories instead of constituents (cf. Betz 2013, p. 10). Further work is needed, however, to clarify the details of the connection between these two approaches (for a preliminary exploration see Cevolani 2014).

Acknowledgments Earlier versions of this paper were presented at the workshop on "Realism, Antirealism, and the Aims of Science" in Trieste (30 June 2012) and at the 5th Copenhagen Lund Workshop on Social Epistemology in Lund (6–7 December 2012). I'd like to thank the participants in those conferences, and in particular Gregor Betz, Jesús Zamora Bonilla, Vincenzo Crupi, Vincent Hendricks, Erik Olsson, Carlo Proietti, and Frank Zenker, for their useful comments. I'm also grateful to Roberto Festa, Theo Kuipers, Gabriella Pigozzi, Luca Tambolo and to two anonymous referees for providing valuable feedback on the first draft of the paper. Support is acknowledged from the *Deutsche Forschungsgemeinschaft* (priority program "New Frameworks of Rationality", SPP 1516, grant CR 409/1-1) and from the Italian Ministry of Scientific Research (FIRB project "Structures and Dynamics of Knowledge and Cognition", Turin unit, D11J12000470001).

7 Appendix: Proofs

We first introduce some notation and terminology. Given a c-theory T, a *completion* of T is any constituent c_i of $\mathscr{L}_n$ which agrees with all the claims of T. One can check that $c_i \in \mathscr{R}_T$ iff c_i is a completion of T. The *reversal* of a constituent is another constituent, which is the conjunction of all the negations of the conjuncts of the former constituent. By extension, the reversal of a sentence X is the sentence which contains, in its range, the reversal of each constituent appearing in $\mathscr{R}_X$ (Oddie 1986, p. 50). One can check that, given a c-theory T, the reversal of T is just the conjunction of the negations of the claims of T; this was called the "specular" of T by Cevolani et al. (2011, p. 186).

Proof of result (8) Let be $E = [T_1, \ldots, T_k]$ a profile of k c-theories. Recall from (3) that $E^\circ = \bigvee\{c_i : \Delta_{min}(E, c_i) \text{ is minimal}\}$, where $\Delta(E, c_i) = \sum_{T_j \in E} \Delta_{min}(T_j, c_i)$. Note that, if T_j is a c-theory, the closest constituent in $\mathscr{R}_{T_j}$ to c_i is the completion of T_j which agree with c_i on all the basic sentences which are indeterminate in T_j; thus, $\Delta_{min}(T_j, c_i)$ is just the number of mismatches between T_j and c_i. In turn, $\Delta(E, c_i)$ is the sum of the number of mismatches between c_i and each T_j in E. Moreover, one can check that this sum can be equivalently expressed as the the sum, for each basic sentence b appearing in c_i, of the number of theories in E rejecting b. This means that c_i is the closer to E, the smaller the number of theories in E rejecting each of its claims.

Let be c_i and c_h two constituents differing only because b is accepted in c_i and rejected in c_h. Then c_i is closer than c_h to E iff: 1) either some T_j in E accepts b and no T_j in E rejects b; 2) or there are in E both theories accepting b and theories rejecting b, but the former are more than the latter. This means that all the constituents closest to E will agree on each basic sentence b such that the number of theories in E accepting b is greater than the number of those rejecting it. The conjunction of such b is exactly the c-theory E°.

Proof of result (9) Let be T_1 and T_2 two c-theories. As Fig. 2 shows, they can be expressed, respectively, as $T_1 \equiv O_{T_1T_2} \wedge C_{T_1T_2} \wedge X_{T_1T_2}$ and $T_1 \equiv O_{T_2T_1} \wedge C_{T_2T_1} \wedge X_{T_2T_1}$. Note also that, on the one hand, the claims of $O_{T_1T_2} = O_{T_2T_1}$, of $X_{T_1T_2}$, and of $X_{T_2T_1}$ are rejected neither by T_1 nor by T_2. On the other hand, each claim of $C_{T_1T_2}$ and of $C_{T_2T_1}$ is rejected by one theory and accepted by the other. From result (8), it follows that $T_1 \circ T_2 = O_{T_1T_2} \wedge X_{T_1T_2} \wedge X_{T_2T_1}$.

Proof of result (10) Let be T_1 and T_2 two c-theories. Note first that, as one can easily check, measure Vs_ϕ as defined in (7) is *additive* in the sense that, given a c-theory T, $Vs_\phi(T) = \sum_b Vs_\phi(b)$ where b is a claim of T. Thus, e.g., $Vs_\phi(T_1) = Vs_\phi(O_{T_1T_2}) + Vs_\phi(C_{T_1T_2}) + Vs_\phi(X_{T_1T_2})$. From result (9), it follows that $T_1 \circ T_2 = O_{T_1T_2} \wedge X_{T_1T_2} \wedge X_{T_2T_1}$. Then $Vs_\phi(T_1 \circ T_2) > Vs_\phi(T_1)$ iff $Vs_\phi(O_{T_1T_2}) + Vs_\phi(X_{T_1T_2}) + Vs_\phi(X_{T_2T_1}) > Vs_\phi(O_{T_1T_2}) + Vs_\phi(C_{T_1T_2}) + Vs_\phi(X_{T_1T_2})$ iff $Vs_\phi(X_{T_2T_1}) > Vs_\phi(C_{T_1T_2})$. Similarly, $Vs_\phi(T_1 \circ T_2) > Vs_\phi(T_2)$ iff $Vs_\phi(X_{T_1T_2}) > Vs_\phi(C_{T_2T_1})$.

Proof of result (11) Let be T_1 and T_2 two c-theories and suppose that there is no weak disagreement between T_1 and T_2, i.e., that their extra parts are "empty". From result (9) it follows that $T_1 \circ T_2 = O_{T_1T_2}$. Moreover, from result (10) it follows that $Vs_\phi(T_1 \circ T_2) > Vs_\phi(T_1)$ iff $Vs_\phi(C_{T_1T_2}) < 0$ and $Vs_\phi(T_1 \circ T_2) > Vs_\phi(T_2)$ iff $Vs_\phi(C_{T_2T_1}) < 0$. Note that, by definition, the conflicting part of T_1 is the reversal of the conflicting part of T_2, and vice versa. One can check that, if $\phi \geq 1$, then $Vs_\phi(C_{T_1T_2}) > 0$ iff $Vs_\phi(C_{T_2T_1}) < 0$. In fact, $Vs_\phi(C_{T_1T_2}) > 0$ iff, by definition (7), $cont_t(C_{T_1T_2}, c_\star) > \phi cont_f(C_{T_1T_2}, c_\star)$, i.e., if T makes more matches than mistakes. But since, by definition, $cont_t(C_{T_1T_2}, c_\star) = cont_f(C_{T_2T_1}, c_\star)$ and $cont_f(C_{T_1T_2}, c_\star) = cont_t(C_{T_2T_1}, c_\star)$, this means that $C_{T_2T_1}$ makes more mistakes than matches and hence, given $\phi \geq 1$, that $Vs_\phi(C_{T_2T_1}) < 0$. It then follows that $Vs_\phi(T_1 \circ T_2) > Vs_\phi(T_1)$ iff $Vs_\phi(T_1 \circ T_2) < Vs_\phi(T_2)$, and vice versa.

References

Alchourrón, C., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, *50*, 510–530.

Betz, G. (2013). *Debate dynamics: How controversy improves our beliefs*. Dordrecht: Springer.

Cevolani, G. (2013). Truth approximation via abductive belief change. *Logic Journal of the IGPL*, *21*(6), 999–1016.

Cevolani, G. (2014). Social epistemology, debate dynamics, and truth approximation. Manuscript.

Cevolani, G., Crupi, V., & Festa, R. (2011). Verisimilitude and belief change for conjunctive theories. *Erkenntnis*, *75*(2), 183–202.

Cevolani, G., Festa, R., & Kuipers, T. A. F. (2013). Verisimilitude and belief change for nomic conjunctive theories. *Synthese*, *190*(16), 3307–3324.

Cevolani, G., & Tambolo, L. (2013). Progress as approximation to the truth: A defence of the verisimilitudinarian approach. *Erkenntnis*, *78*(4), 921–935.

Douven, I., & Kelp, C. (2011). Truth approximation, social epistemology, and opinion dynamics. *Erkenntnis*, *75*, 271–283.

Everaere, P., Konieczny, S., Marquis, P. (2010). The epistemic view of belief merging: Can we track the truth? In: Proceedings of the 2010 conference on ECAI 2010: 19th European Conference on Artificial Intelligence, IOS Press, Amsterdam, pp 621–626, http://dl.acm.org/citation.cfm?id=1860967.1861089.

Festa, R. (1987). Theory of similarity, similarity of theories, and verisimilitude. In T. A. F. Kuipers (Ed.), *What is closer-to-the-truth?* (pp. 145–176). Amsterdam: Rodopi.

Festa, R. (2007). Verisimilitude, cross classification, and prediction logic. Approaching the statistical truth by falsified qualitative theories. *Mind and Society*, *6*, 37–62.

Frances, B. (2010). Disagreement. In D. Pritchard & S. Bernecker (Eds.), *Routledge companion to epistemology*. London: Routledge.

Gärdenfors, P. (1988). *Knowledge in flux: Modeling the dynamics of epistemic states*. Cambridge, MA: MIT Press.

Goldman, A. I. (1999). *Knowledge in a social world*. Oxford: Oxford University Press.

Hansson, S. O. (2011). Logic of belief revision. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy*.

Hartmann, S., & Sprenger, J. (2012). Judgment aggregation and the problem of tracking the truth. *Synthese*, *187*(1), 209–221.

Konieczny, S., & Pino Pérez, R. (2002). Merging information under constraints: A logical framework. *Journal of Logic and Computation*, *12*, 773–808.

Konieczny, S., & Pino Pérez, R. (2011). Logic based merging. *Journal of Philosophical Logic*, *40*, 239–270.

Kuipers, T., & Schurz, G. (2011). Introduction and overview. *Erkenntnis*, *75*, 151–163.

Kuipers, T. A. F. (1987). A structuralist approach to truthlikeness. In T. A. F. Kuipers (Ed.), *What is closer-to-the-truth?* (pp. 79–99). Amsterdam: Rodopi.

Kuipers, T. A. F. (2000). *From instrumentalism to constructive realism*. Dordrecht: Kluwer.

Kuipers, T. A. F. (2011). Basic and refined nomic truth approximation by evidence-guided belief revision in agm-terms. *Erkenntnis*, *75*, 223–236.

Linstone, H. A., & Turoff, M. (1975). *The Delphi method*. Reading, MA: Addison-Wesley.
List, C. (2012). The theory of judgment aggregation: An introductory review. *Synthese*, *187*(1), 179–207.
Mill, J. S. (1848). The principles of political economy. In J. M. Robson (Ed.), Collected works, vol. III. Toronto & London: University of Toronto Press/Routledge and Kegan Paul, 1965. http://oll.libertyfund.org.
Mill, J. S. (1859). On liberty. In *On liberty and the subjection of women*, 1879. New York: Henry Holt & Co. http://oll.libertyfund.org.
Miller, D. (1974). Popper's qualitative theory of verisimilitude. *The British Journal for the Philosophy of Science*, *25*(2), 166–177.
Niiniluoto, I. (1987). *Truthlikeness*. Dordrecht: Reidel.
Niiniluoto, I. (1998). Verisimilitude: The third period. *The British Journal for the Philosophy of Science*, *49*(1), 1–29.
Niiniluoto, I. (1999a). Belief revision and truthlikeness. In B. Hansson, S. Halldén, N. E. Sahlin, W. Rabinowicz (Eds.), *Internet Festschrift for Peter Gärdenfors*, Department of Philosophy, Lund University, Lund. http://www.lucs.lu.se/spinning/.
Niiniluoto, I. (1999b). *Critical scientific realism*. Oxford: Oxford University Press.
Niiniluoto, I. (2011). Revising beliefs towards the truth. *Erkenntnis*, *75*(2), 165–181.
Oddie, G. (1986). *Likeness to truth*. Dordrecht: Reidel.
Oddie, G. (2008). Truthlikeness. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.
Oddie, G. (2013). The content, consequence and likeness approaches to verisimilitude: Compatibility, trivialization, and underdetermination. *Synthese*, *190*(9), 1647–1687.
Pigozzi, G. (2006). Belief merging and the discursive dilemma: An argument-based account to paradoxes of judgment aggregation. *Synthese*, *152*, 285–298.
Popper, K. R. (1963). *Conjectures and Refutations: The growth of scientific knowledge* (3rd ed.). London: Routledge and Kegan Paul.
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. (2011). Verisimilitude and belief revision. With a focus on the relevant element account. *Erkenntnis*, *75*(2), 203–221.
Schurz, G., & Weingartner, P. (1987). Verisimilitude defined by relevant consequence-elements. In T. Kuipers (Ed.), *What is closer-to-the-truth?* (pp. 47–77). Amsterdam: Rodopi.
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.
Tichý, P. (1974). On Popper's definitions of verisimilitude. *The British Journal for the Philosophy of Science*, *25*(2), 155–160.
Zamora Bonilla, J. (2007). Optimal judgment aggregation. *Philosophy of Science*, *74*(5), 813–824.
Zamora Bonilla, J. (2012). The economics of scientific knowledge. In U. Mäki (Ed.), *Philosophy of economics* (pp. 823–862). Amsterdam: Elsevier.