Debate Dynamics: How Controversy Improves Our Beliefs

SYNTHESE LIBRARY

STUDIES IN EPISTEMOLOGY,

LOGIC, METHODOLOGY, AND PHILOSOPHY OF SCIENCE

VOLUME 357

For further volumes:
http://www.springer.com/series/6607

Gregor Betz

Debate Dynamics: How Controversy Improves Our Beliefs

Springer

Gregor Betz
Institute of Philosophy
Karlsruhe Institute of Technology
Kaiserstrasse 12
D-76131 Karlsruhe
Germany

ISBN 978-94-007-4598-8 ISBN 978-94-007-4599-5 (eBook)
DOI 10.1007/978-94-007-4599-5
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2012946202

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Acknowledgements

This book would not exist without the generous support I received from the Stiftung Alfried Krupp Kolleg Greifswald. I'm very grateful for the year I was allowed to spend as a fellow in Greifswald, and I'd like to thank Prof. Bärbel Friedrich, Dr. Christian Suhm, and their team personally for the close and cordial collaboration. Moreover, I'm indebted to colleagues and friends who critically assessed parts of the reasoning this book unfolds, specifically to Sebastian Cacean, Moritz Cordes, David Löwenstein, Friedrich Reinmuth, Holm Tetens, and Christian Voigt as well as the participants of two colloquia at the University of Greifswald and at the Freie Universität Berlin. Last but not least, I'd like to thank the anonymous reviewer of the Synthese Library Series for the astute comments and very helpful suggestions.

Contents

Chapter 1
General Introduction

1.1 The Aims of Argumentation

The idea that controversial argumentation improves our beliefs is as old as argumentation theory—the systematic reflection on argumentation—itself and probably even older. But what precisely does this alleged improvement we aim at when engaging in a controversy mean? In which sense does the game of giving and taking reasons, presumably, better our beliefs?

We may discern, by and large, two distinct, fundamental rationales one can pursue in an argumentation. The first one is to overcome dissent and to reach a consensus. The second one consists in rectifying our errors and tracking down the truth.

These two aims can be pursued, and achieved, quite independently of each other. The proponents in a debate may very well reach a consensus position without having acquired correct beliefs, in which case the consensus is a spurious one. Similarly, a proponent might acquire true beliefs, achieving the second rationale, while continuing to disagree with her opponents. But obviously, if all proponents have found the truth, they eo ipso agree.

Proponents who engage in an argumentation don't necessarily strive both for consensus *and* for truth. Some debates, such as, for example, the moral controversy about preimplantation genetic diagnosis or the debate about legalizing voluntary euthanasia, might primarily aim at reaching broad societal agreement, and the proponents might simply not be interested in the additional question whether a consensus position, should it once emerge, is also true (if they judge that question meaningful at all). In other debates, however, finding the truth constitutes the primary, or even the sole rationale. Think of scientific controversies. Peer agreement is not what intrinsically motivated scientists are ultimately striving for. In the end, they aspire to correct answers—say, to the question whether the Earth was completely covered by ice sheets once or what caused ancient civilizations to collapse—and not merely to a consensus. In these debates, agreement is at most of indirect interest, namely insofar as persisting dissent indicates that not all proponents have found the truth yet. Finally, the relative importance of the two

G. Betz, *Debate Dynamics: How Controversy Improves Our Beliefs*,
Synthese Library 357, DOI 10.1007/978-94-007-4599-5_1,

rationales may not only vary from debate to debate, but even within a debate, from proponent to proponent, or from one (historic) phase to another.

In spite of being principally distinct aims, consensus and truth are intricately related. We have already noted that a consensus among proponents represents a necessary condition for all proponents having found the truth.[1] This simple observation—that dissent indicates falsity—raises the more interesting question whether consensus, vice versa, indicates truth as well. Further questions pertain to potential trade-offs between the two rationales. Does, for example, a controversy which effectively generates agreement among proponents enable them to track down truth? Or, does a truth-conducive argumentation tend to obstruct rapprochement of the proponents? We will return to these issues in due course.

We engage in argumentation in order to reach agreement and to find the truth. Putting forward and listening to arguments is no self-sufficient activity, no *l'art pour l'art*. The rationality of a controversial argumentation thus resides in its effectiveness as to realizing its aims. It is an instrumental rationality. We can distinguish, accordingly, an instrumental *consensual* value of argumentation on the one hand, and an instrumental *veritistic* value of argumentation on the other hand. This book amounts to an investigation of both. It studies the consensual and veritistic value of different argumentative practices—practices which diverge in regard to the way proponents put forward arguments, and modify their convictions in the light of newly introduced reasons. One reading of the ensuing inquiry, hence, consists in conceiving it as a contribution to the reliabilistic program of social epistemology (see Goldman 1999).

1.2 An Example of a Controversial Argumentation

Before I set forth the methods and assumptions of the following investigation, we shall consider an example of a controversial debate. This miniature case study illustrates the kind of argumentation to which the formal investigation, unfolded in this book, applies. Moreover, it helps to frame and grasp the rather abstract concepts which are to be introduced henceforth.

The scientific controversy concerning the origin of the so-called Nördlinger Ries—an uncommon, circular depression of the landscape, which circumscribes the town of Nördlingen in Southern Germany and which represents in fact, as we know today, the remnants of an impact crater, testifying to the impact of a meteorite roughly 15 million years ago—will serve as an example.[2] The origin of the Ries, and of its unusual rocks, has been unclear for a long time, and only relatively lately, namely in 1960, did Shoemaker and Chao (1961) succeed in demonstrating that the Ries represents an impact crater, effectively closing a controversy whose beginnings date back to the end of the eighteenth century.

[1] As Descartes has already remarked in his *Rules* (Descartes 1984, p. 11).

[2] The following account is based on von Engelhardt (1982) and Kölbl-Ebert (2003).

In search of building materials, the Bavarian engineer C. v. Caspers found, in the 1780s, that specific rocks from the Ries area (referred to as suevite today) are suited for mortar production. Likening these rocks to the trass in the Rhenish area, which possesses similar properties, and whose volcanic origin had only recently been established, Caspers argued:

Hypothesis 1 (Volcanic origin) *Suevite is a volcanic product.*

The volcanic hypothesis has been largely agreed upon in subsequent decades. Geologists who studied the Ries such as Flurl (1805)[3], Schübler (1825), Cotta (1834), and Voith (1835) assented to this theory and, in part, provided additional arguments by drawing further analogies or pointing out that the scattered suevite occurrences may be understood as lava bombs.

Mapping the geology of the Ries area, Schnitzlein and Frikhinger (1848) found that the rocks' sequence in the basin doesn't accord with their normal geological position (i.e., older rocks were situated on top of younger ones). On the basis of this observed stratigraphic disturbance, they argued in favor of an extended and modified volcanic hypothesis.

Hypothesis 2 (Volcanic origin) *Volcanic forces lifted old rocks from deeper depth to the surface, caused multiple eruptions which gave rise to suevite occurrences, and, finally, triggered the subsidence of the basin.*

At the same time, Schafhäutl (1849) proposed a completely different theory, rejecting, in consequence, the volcanic hypothesis.

Hypothesis 3 (Viscous underground) *Vast underground resources of a viscous silicate gel contracted (due to water loss) and sparked off the subsidence; the gel ascended, henceforth, along the resulting fractures at the margins of the depression, and eventually solidified, forming granite-like rocks (including suevite).*

Schafhäutl's main argument was based on a chemical analysis of suevite, which revealed a close resemblance to granite and disclosed, moreover, significant differences between suevite and the Rhenish trass, thereby undermining the argument, originally introduced by Caspers, in favor of the volcanic theory. Given that suevite and granite occurrences are, in addition, locally correlated, Schafhäutl inferred that both stem from one and the same origin, which eventually triggered his inventive hypothesis.

Some 20 years later, Deffner (1870) suggested yet another rival hypothesis.

Hypothesis 4 (Ries glacier) *The Ries basin once hosted a glacier; concentric ice flow in all directions powered a corresponding transport of material.*

Deffner supported his theory on the basis of new evidence pertaining to large amounts of debris, consisting in rocks from different geological periods, outside

[3]For bibliographic references to this as well as to the other original contributions to the Ries controversy mentioned below, compare von Engelhardt (1982).

the Ries basin. Moreover, he had discovered, together with his colleague O. Fraas, signs of lateral transport such as polished surfaces and striations pointing towards the basin's center. A glacier, Deffner thought, was the only mechanism which could account for the vast displacements of rocks, and the characteristic traces which had, at other places, already been attributed to glacial activities.

Fraas, however, well aware of the evidence for massive horizontal displacements, didn't concur in his colleague's glacier hypothesis, questioning, in particular, the ability of a glacier to cause mass transport on the required scale. In addition, he provided a detailed description of the shapes of the suevite bombs, which witness, he argued, to the air resistance in the course of their flight through the atmosphere, and which therefore support the volcanic theory.

Gümbel (1870,91,94), too, accepted the volcanic hypothesis yet argued for a specific modification.

Hypothesis 5 (Volcanic origin) *The suevite occurrences result from a single volcanic ejection.*

He had discovered that all suevite bombs, scattered over the Ries area, display a similar microscopic structure and composition, which suggests that they spring from a single event.

Being apparently still puzzled by the traces of lateral transport such as the polished surfaces and striations, which, allegedly, only pertain to geologically young rocks, Koken (1901,02) argued in favor of a revival of Deffner's glacier hypothesis, which he now understood, however, as a complementary rather than a rival hypothesis to the volcanic theory. By distinguishing different phases of the Ries' evolution, with volcanic activities preceding its later glaciation, Koken tried to reconcile some of the different theories previously proposed.

Branco (1901,03) and Fraas (1901,03,19), however, strictly opposed Koken's modified glacier hypothesis and attempted to refute it on different grounds. They insisted, first, that substantial amounts of debris have been discovered at places far removed from the basin: This documents a mass transport glacier theory cannot account for. Second, the characteristic striations in younger rocks, which supposedly testify to a relatively recent glaciation, are in fact also discernible in much older rocks.

As an alternative to transport by glaciation, Branco proposed the following hypothesis, which still remained, by and large, within the cluster of volcanic theories.

Hypothesis 6 (Ries mountain) *A magma pocket of 25 km diameter created, through expansion, a colossal mountain with steep slopes. As a result of gas release (or other reasons), the mountain eventually disappeared and gave way to today's basin.*

Branco argued that, on account of the Ries mountain's steep slopes, gravitational slides shredded and transported material far beyond today's basin. Yet, realizing that gravitational forces don't yield enough energy to bring about the observed dislocations, Branco modified his original hypothesis.

Hypothesis 7 (Ries mountain) *The sudden elevation of the Ries mountain was accompanied by a (water vapor) explosion.*

This modified hypothesis, Branco argued, provides a better explanation of the chaotic debris. Moreover, he pointed out analogies to other sudden volcanic explosions without lava ejection.

Still, even Branco's modified hypothesis didn't satisfy Kranz (1911,14–52), who questioned the ability of a conventional volcanic eruption to release sufficient amounts of energy. As a result, Kranz proposed the so-called blasting theory.

Hypothesis 8 (Blasting) *Ground water entered a magma chamber situated in shallow depth, triggering a massive explosion.*

Kranz' theory gave a unified account of the diverse evidence. It explained, for example, the different items of polished and striated surfaces as resulting from the impacts of rock fragments which were catapulted by the explosion. According to Kranz' blasting theory, though, the Ries was conceived as a geologically unique phenomenon. Nonetheless, blasting theory became, gradually, the consensus view of geologists, and remained so until 1960.

It was the strangeness of the Ries which led Werner (1904) to surmise a radically different theory about the Ries' origination.

Hypothesis 9 (Impact) *The Ries represents the remnants of an impact crater.*

Werner likened the Ries to lunar craters, but failed to give arguments in favor of his hypothesis. Two further authors, Kaljuwee (1933) and Stutzer (1936), had advanced the impact theory before 1960. While Kaljuwee argued that the Earth had been subject to massive meteoritic bombardment in the past, whose traces cannot have been entirely erased from Earth's surface, Stutzer compared the Ries with the Barringer Crater in Arizona, noting significant morphological similarities. The impact hypothesis was nonetheless dismissed by the scientific community. This changed, however, radically and sustainably in 1960 once Shoemaker and Chao detected the rare mineral coesite, which was first discovered in the 1950s and which crystallizes only at very high pressures, in samples of rocks from the Ries. Both had previously found coesite in the Barringer Crater, as well. Since the extreme pressures required to form coesite cannot be reached by volcanic activities, this discovery was unanimously regarded as a successful verification of the impact theory, effectively closing the controversy.

The sketch of the Ries debate illustrates the kind of controversial argumentation this book's inquiry is going to analyze. Let us try to describe the lively debate in somewhat more abstract, argumentation-theoretic terms. The debate comprises different proponents who hold specific positions. These are modified in the light of new arguments introduced into the debate. Some of these arguments are intended to support a given position, others are set forth so as to attack opponents. Novel evidence enables the proponents to put forward ever new arguments. Proponents agree with each other to different degrees. Cotta and Voith, for example, holding Hypothesis 1 advanced by Caspers, concur, obviously, by and large with Schnitzlein

and Frikhinger, who maintain only a slightly modified position (Hypothesis 2). In contrast, these proponents disagree more or less fundamentally with advocates of Schafhäutl's viscous underground theory (Hypothesis 3). As the debate evolves, and as proponents alter their positions, the overall agreement changes as well, resulting, as far as we can tell from our brief sketch, in phases with almost unanimous consensus (e.g., 1800–1840, 1920–1940), or outspoken plurality and dissent (e.g., 1900–1920). Besides mutual agreement, the overall truthlikeness, or "verisimilitude," of the proponent positions appears to vary too. So, some hypotheses seem to be closer to the truth (of course, relative to our current knowledge) than others. Gümbel's Hypothesis 5, for example, improves *objectively* upon previous volcanic hypotheses by attributing all suevite occurrences to one and the same source. Likewise, the refined Ries mountain theory (Hypothesis 7) gets much closer to the truth than hypotheses which posit glaciation but is itself outperformed by blasting theory (Hypothesis 8). In sum, there is nothing fundamentally obscure about assessing, in retrospective, the effects of controversial argumentation on the mutual agreement and verisimilitude of proponent positions. (That these concepts call, however, for more precise explications goes without saying.)

The general purpose of this book is to assess the instrumental consensual and veritistic value of controversial argumentation, in other words, to assess its consensus- and truth-conduciveness. One method for doing so would consist in providing, first of all, a *detailed* reconstruction and analysis of our example, specifying, in particular, how arguments are introduced into the debate, how proponents modify their beliefs in response, and how this affects the proponents' overall agreement, as well as the correctness of their convictions. Next, an equally detailed analysis would have to be carried out for dozens if not hundreds of further controversies so as to obtain a sufficiently broad sample of dynamic debate reconstructions. A statistical analysis of this sample could then teach us whether controversial argumentation is, in general, consensus- and truth-conducive and which specific argumentation strategies are particularly effective with regard to generating agreement and discovering the truth.

Obviously, that is a giant task, and is not going to happen, at least not here. It takes already one book to document the reconstruction and analysis of a single debate. The above paragraph hence outlines rather an entire research program than an investigation to be unfolded in a monograph.

Lacking a sufficiently large sample of dynamic debate reconstructions, which would allow us to learn from previous experience with different argumentative practices, we are going to generate our own tailored samples by simulating controversial argumentation. So, instead of studying real debates and their reconstructions, we investigate simulated debates and their automatically generated formal representations. This allows us, in principle, to scrutinize the effects of different argumentative practices in arbitrary many debates and therefore to identify their consensus- and truth-conduciveness accurately. Clearly, the simulation of controversial argumentation has to rely on an adequate model which incorporates the relevant aspects of debate dynamics. This model will be described, informally, in the following section. It extends the approach developed in Betz (2010) by a dynamic component and is carefully set forth in Chap. 2.

1.3 Modeling Controversial Debate

A fixed state of some debate is essentially characterized by (a) the arguments which have been uncovered and introduced so far and (b) the positions maintained by the debate's proponents. We assume that the arguments are—or are reconstructed and thence represented as—deductively valid inferences from some premisses to a conclusion. Arguments may support or attack each other, giving rise to a complex argumentation which we will refer to as a dialectical structure and which may be visualized as an argument map.

Given a dialectical structure containing arguments which mutually support and attack each other, we can identify a position, actually or potentially held by some proponent, with a truth-value assignment to the sentences which figure in the debate. We refer to a truth-value assignment to all the sentences which occur in the debate as a "complete position"; a partial position, in contrast, maps truth values to some of the sentences only. While assuming, throughout the following study, that proponents hold complete positions, we do distinguish, in some cases, so-called core and auxiliary beliefs, capturing the Lakatosian idea that proponents don't regard all sentences which figure in a debate as equally important.

The arguments advanced in a debate entail certain constraints a position ought to satisfy so that a proponent may reasonably adopt it. In addition to assigning equivalent sentences identical truth values, and contradictory sentences complementary ones, a position must, on account of deductive validity, consider a conclusion of some argument true, if its premisses are deemed correct. We shall call a complete position "dialectically coherent" if and only if it satisfies these constraints.

Since positions are identified with truth-value assignments, their mutual agreement can simply be assessed by counting the number of sentences to which two positions assign the same truth value. This gives us a simple metric on the set of all positions, and allows us to picture the set of coherent positions as a space in which the proponents (provided they hold dialectically coherent positions) are located.

When investigating the veritistic value of controversial argumentation, we stipulate that some position (truth-value assignment) is correct and represents the true truth-value assignment, that is, the truth. The truth is dialectically coherent (for no deductive valid argument has true premisses and a false conclusion) and is itself located in the space of coherent positions. Assessing a position's agreement with the truth yields a convenient way for gauging its truthlikeness (verisimilitude).

The background knowledge shared by the debate's proponents (*endoxa*) represents an additional characteristic of a given state of debate. We model background knowledge as a partial position which is accepted by all proponents; more precisely, the complete proponent positions necessarily agree with the background knowledge. Moreover, we assume that the background knowledge is (1) constant and (2) correct, that is, agrees with the truth. In other words, we don't consider the case where proponents systematically err with respect to background assumptions.

So far, we have merely given a static account of a debate, which focuses on some fixed state of a controversial argumentation, taking a single snapshot.

Clearly, this framework could be applied to reconstruct consecutive states of a real debate, which would provide a dynamic picture of how the individual states evolved into one another. Yet, a simulation of debate dynamics (in contrast to its mere reconstruction) requires, in addition, that we model the way a given state of a debate triggers a further one. In particular, we have to describe how the two most important constituents of some state of debate, that is, the dialectical structure (comprising the arguments advanced so far) and the proponents' positions, evolve. Accordingly, we must detail an *argument construction mechanism* and a so-called *update mechanism*. The individual simulation studies documented in this book's chapters vary, primarily, in regard to the specific argumentation and update mechanism they presume. We shall roughly summarize and categorize these various assumptions in the following.

The most simple *argument construction mechanism* posits that new arguments be devised randomly, that is, that the premisses and conclusion of a new argument be drawn randomly from the set of all sentences which pertain to the debate. Consequently, arguments are not purposefully contrived by proponents and relate only coincidentally to the positions proponents maintain. According to random argument construction, arguments—previously unseen inferential relations—are rather discovered than designed.

More sophisticated argument construction mechanisms assume that arguments are introduced by a specific proponent who follows a certain argumentation rule, taking the positions, held by the debate's participants, into consideration. An important aspect which distinguishes such argumentation rules is the relative importance attached to the position of the proponent who advances the new argument versus her opponents' positions. We may thus distinguish argumentation rules which prescribe to introduce an argument that (a) backs the proponent's position (the conclusion is maintained as true by the proponent) or (b) criticizes an opponent's position (the conclusion is denied by an opponent). Likewise, we can discriminate between rules which demand that the premisses of a new argument be accepted (a) by the proponent who advances the argument or (b) by one of her opponents. These distinctions give rise to four basic types of argumentation strategies:

Fortify: An argument satisfies the *fortify* rule iff the proponent who puts forward the argument considers its premisses and its conclusion true.

Attack: An argument satisfies the *attack* rule iff the proponent who puts forward the argument maintains its premisses while one of her opponents denies its conclusion.

Convert: An argument satisfies the *convert* rule iff the proponent who puts forward the argument maintains its conclusion while one of her opponents accepts its premisses.

Undercut: An argument satisfies the *undercut* rule iff an opponent of the proponent who puts forward the argument denies its conclusion while conceding its premisses.

Whereas arguments that satisfy *fortify* and *attack* take off from the proponent's convictions, *convert* and *undercut* urge proponents to base new arguments on

premisses accepted by their opponents. The latter strategies are opponent-sensitive, whereas the former ones may be characterized as self-centered. Moreover, by prescribing to criticize an opponent's position, *attack* and *undercut* are more aggressive than, respectively, *fortify* and *convert*.

Besides random argumentation, the four argumentation rules, or variants thereof, constitute the primary argument construction mechanisms studied in this book. This enables us to examine how opponent-sensitive and self-centered as well as more or less aggressive argumentation strategies affect the consensual and veritistic dynamics of debates.

A final type of argumentation strategy we consider can be employed by proponents who hold a core position plus further auxiliary beliefs. Such a core position possesses a specific degree of justification, or robustness, relative to a given state of debate. A core position's degree of justification can be precisely defined in the framework of the theory of dialectical structures (cf. Sect. 2.2), and proponents may hence introduce new arguments so as to maximize the degree of justification of their own core position. We will refer to this rule, which represents a self-centered strategy, as *maximize robustness*.

The argument construction mechanism specifies how a dialectical structure grows from one step in a debate to another. Likewise, the *update mechanism* describes how proponent positions evolve and, in particular, respond to arguments that have been newly discovered and introduced into the debate. In a nutshell, we assume that proponents hold and retain dialectically coherent positions, and try to minimize the number of belief revisions which are necessary to do so. To understand the dynamics of proponent positions, it is important to note that dialectical coherency hinges sensitively on the dialectical structure against which positions are assessed, and hence on the arguments discovered so far. More precisely, a position which is dialectically coherent given a state of debate might become dialectically incoherent as new arguments are introduced and proponents have to take account of further inferential relations. If a newly introduced argument renders the position maintained by some proponent dialectically incoherent, the proponent modifies her truth-value assignments so as to reestablish dialectical coherency with respect to the enlarged dialectical structure. We assume, in addition, that proponents are conservative in the sense of revising their convictions only reluctantly. Specifically, proponents minimize the individual revisions of truth-value assignments so as to regain a dialectically coherent position. Or, in other words, proponents update their position to the closest coherent one.

The closest coherent update mechanism underlies most of the simulations presented in this inquiry. Occasionally, however, we presume a slightly more sophisticated revision policy. As previously remarked, we distinguish, in some simulations, core and auxiliary convictions. Proponents are assumed to stick particularly vehemently to their core beliefs while being much more willing to modify their auxiliary convictions. This suggests the following modification of the simple closest coherent update mechanism: If the complete position held by a proponent is rendered dialectically incoherent, the proponent determines, in a first step, all coherent positions that agree maximally with her core convictions. In a

second step, she chooses from those positions the one that displays the greatest overall agreement with her previous position. This lexicographic update mechanism will be employed whenever we distinguish core and auxiliary sentences.

This sketch of how we model debate dynamics clearly exposes some simplifications and therefore suggests obvious extensions. To begin with, there is no reason to assume that proponents maintain but complete positions. To withhold judgement in regard to some sentence may very well be a reasonable doxastic state. Moreover, that is what happens in real debates all the time. Accordingly, a first interesting extension of this investigation could posit that proponents hold but partial positions. This would trigger a corresponding modification of the debate dynamics, in particular, of the update mechanism, which must allow for retracting truth-value assignments to some sentences altogether as well as for extending one's partial position. Secondly, future research might loosen the assumption that (explicit) background knowledge is constant and correct. The externally fixed background knowledge might itself grow in the course of a debate; it is, moreover, not immune to revisions and might therefore vary considerably. A particularly interesting extension consists in studying the effects of background knowledge correction, that is, the revision of false yet previously universally shared beliefs. These brief remarks demonstrate that the investigation carried out in this book is, by no means, to be read as a final word. Rather, it paves the way for possibly even more interesting inquiries into the dynamics of controversial argumentation within the framework of the theory of dialectical structures.

1.4 Results Pertaining to Consensus-Conduciveness

In the following, we report and summarize the main results regarding the consensual value of controversial argumentation, which are derived, and discussed, in Part I of this book:

C1 (General Results) Controversial argumentation is, all things considered, consensus-conducive. Although the concrete agreement evolution in an individual debate seems to depend, mainly, on random factors, we may nonetheless discern substantial statistical differences between different argumentative practices.

C1.1 (Long Run) A controversial argumentation compels proponent positions to converge, eventually. This is, however, hardly surprising inasmuch as, in the long run, only one single position remains dialectically coherent. Different argumentative practices vary substantially with respect to the pace of this convergence.

C1.2 (Alienation) Controversial argumentation may very well, in particular during the initial phase of a debate, lead to the alienation of proponent positions, and undo coincidental agreement. Instead of generating agreement, controversy sparks dissent. This effect, too, depends strongly on the argumentative strategies employed by the proponents. It is, in line with (C1.1), inevitably reversed in the long run.

C1.3 (Global Agreement Versus Partial Consensus) There exists a trade-off between (a) increasing the overall mean agreement between *all* proponents in a debate and (b) prompting at least some proponents to agree fully. Debate evolutions which foster partial consensus (i.e., full agreement between some proponents) tend to slow down the global rapprochement of proponent positions.

C1.4 (The Space of Coherent Positions) The characteristic consensus dynamics of argumentative practices, such as, for example, the result (C1.3), can be explained in terms of how the corresponding argumentation shapes the space of coherent positions. In particular, the degree of fragmentation of the space of coherent positions—whether the remaining coherent positions, that is, are all closely related to each other, or form, on the contrary, distant and isolated opinion clusters—turns out to be of pivotal importance for the belief dynamics. The concept of the space of coherent positions is, in fact, the primary theoretical tool for understanding the consensus-conduciveness of argumentative practices.

C2 (Background Knowledge) The introduction of background knowledge into a debate fosters, very much as one would expect, the mutual agreement between proponents.

C2.1 (Multiplier Effect) The introduction of constant background knowledge accelerates the rapprochement of proponent positions. This is because, as the debate unfolds, ever more sentences are derived from the constant body of background beliefs. These sentences become, consequently, part of the effective background knowledge themselves and may, in turn, serve as a basis for the derivation of further statements. This multiplier effect drives the discernible speed up of mutual rapprochement.

C2.2 (Favorable Fragmentation) With a sufficiently broad body of background knowledge, the fragmentation of the space of coherent positions, which tends to obstruct mean agreement increase without background knowledge (C1.4), favors both the generation of partial consensus and the global increase of mean agreement, thus resolving the trade-off reported above (C1.3).

C3 (Argumentation Strategies) The consensus-conduciveness of specific argumentative practices varies widely. The most noteworthy differences pertain to self-centered argumentation rules on the one side and opponent-sensitive ones on the other side.

C3.1 (Self-Centered Argumentation) Self-centered argumentation strategies, that is, argumentation rules (such as *fortify* and *attack*) which stipulate that a proponent advances but arguments with premisses she accepts as true, are totally ineffective in generating agreement. Strategies which are in addition aggressive, recommending direct attacks against opponent positions (e.g., the *attack* rule), consistently destroy agreement in all phases of a debate and drive proponent positions systematically apart.[4]

[4]Note that this stands in no contradiction to result (C1.1) because at some point in a debate, there are typically no more arguments that satisfy the *attack* rule, and proponents have to resort to other strategies if the debate shall continue.

C3.2 (Opponent-Sensitive Argumentation) Opponent-sensitive argumentative practices, however, are highly consensus-conducive. So, using, as premisses of the arguments one introduces to back up one's position, statements which an opponent considers true represents the most effective way for generating agreement. This result underlines the importance of explicitly addressing opponents by taking their positions as starting points for new arguments.

C3.3 (Aggressiveness and Disagreement) Aggressive opponent-sensitive strategies, that is, extremely critical strategies such as the *undercut* rule, are, in general, less consensus-conducive than their non-aggressive counterparts (*convert*). Too much criticism and too many direct attacks seem to inhibit rapprochement. The less aggressive *convert* rule, moreover, allows for an apparently highly beneficial strategy: Before directly refuting an opponent position, potential back doors (adjacent fall-back positions) which are available to the opponent and which are farther removed from the proponent than the opponent's current position are closed (rendered incoherent). When the opponent position is, afterwards, directly refuted, the opponent is compelled to relocate towards the proponent. Our simulations identify this complex mechanism and demonstrate its consensual value.

C3.4 (Friends and Fundamentalists) The effectiveness of an argumentation strategy in generating consensus depends on whether the initial agreement with one's opponent is very high ("friend") or very low ("fundamentalist"). Thus, a sharply critical, aggressive opponent-sensitive rule is advisable when arguing with a fundamentalist. Frequent falsifications due to "internal critique" represent in fact the most appropriate means for overcoming extreme dissent (cf. Schleichert 1998, pp. 93–111). Massive criticism impedes, however, finding consensus when arguing with a friend. Minor disagreement, instead of being effectively resolved, is typically deepened by aggressive argumentation. Here, the less critical *convert* strategy is much more consensus-conducive than the *undercut* strategy.

C4 (Consensus Bias) Different argumentative practices do not only vary with respect to their consensus-conduciveness. The argumentation strategies employed by the proponents affect, in addition, the distance between the proponents' initial positions and the debate's final consensus.

C4.1 (Resilient Argumentation) The final consensus reached in a debate tends to be closer to the initial positions held by proponents who employ an opponent-sensitive argumentation strategy (i.e., the *convert* or *undercut* rule) than to the initial positions maintained by proponents who argue in a self-centered way (implementing the *fortify* or *attack* rule). Thus, a proponent who follows an opponent-sensitive argumentation rule does not only foster consensus but also benefits from the ensuing fact that the final consensus is, on average, biased towards her initial position.

C4.2 (Robust Core Positions) Proponent core positions with a high degree of justification at an early phase of the debate tend to be closer to the final consensus than core positions which exhibit a low degree of justification. Degree of justification (at an early state of the debate) correlates with a position's agreement with the final consensus. This is because the higher the degree of justification, the more

robust the corresponding core position and the more flexibly can a proponent adapt her complete position to critical arguments without modifying her core beliefs. This result provides a first justification for why adopting a position with a high degree of justification is rationally desirable at all (see also T4.3 below).

C4.3 (Sensitivity of Indicator) Proponents who introduce arguments so as to maximize the robustness of their core position don't reach a consensus any faster than proponents who apply opponent-sensitive strategies. Yet, in debates where proponents maximize their robustness through argumentation, the accuracy of the degree of justification as an indicator of a position's agreement with the final consensus increases dramatically. In contrast, the correlation between robustness and agreement with the final consensus almost vanishes entirely if proponents pursue very aggressive and critical strategies. Robustness seems to be a highly sensitive indicator.

1.5 Results Pertaining to Truth-Conduciveness

This section summarizes our findings about the veritistic value of controversial argumentation. The following results are spelled out in much more detail in Part II of this book:

T1 (General Results) In toto, controversial argumentation enables proponents to track down the truth. Individual veritistic dynamics vary substantially from debate to debate and are mainly determined by random factors. Still, different argumentative practices give rise to specific mean verisimilitude evolutions, and can thence be characterized statistically.

T1.1 (Long Run) Proponent positions converge, in the long run, against the truth. As explained above (C1.1), this is not surprising. Argumentative practices differ, however, significantly with respect to the speed and timing of the verisimilitude increase.

T1.2 (Epistemic Deterioration) Controversial argumentation may trigger a temporary loss of, instead of a gain in, verisimilitude. Still, verisimilitude evaporates to a much lesser degree than mutual agreement in the course of a debate. That is because it is comparatively difficult to render proponent positions which are close to the truth dialectically incoherent.

T1.3 (Engine of Progress) Criticism, as Mill argued no less eloquently than prominently, is indeed the main driver of epistemic progress.[5] The pace at which

[5] In *On Liberty*, Mill defends freedom of speech on grounds of the epistemic virtues of controversial discussion."Complete liberty of contradicting and disproving our opinion, is the very condition which justifies us in assuming its truth for purposes of action; and no other terms can a being with human faculties have any rational assurance of being right." (Mill 2009, p. 60) "The steady habit of correcting and completing his own opinion," Mill details, "is the only stable foundation for a just reliance on it: for, being cognisant of all that can, at least obviously, be said against him, and

proponents approach the truth is largely determined by the frequency at which their positions are rendered incoherent (successfully criticized). Rendering a proponent position incoherent requires, however, that one pinpoints an internal inconsistency pertaining to a subset of the proponent's beliefs, not all of which must, as deductive logic has it, be true. The fact that not all sentences figuring in an alleged inconsistency may be true, whereas, of course, they may all very well be false, amounts to a small but nonetheless influential asymmetry, which assures that, on average, internal critique tends to target more false than correct beliefs, and thus prompts a proponent to modify her position for the better.

T1.4 (Consensual and Veritistic Value) The relationship between consensus- and truth-conduciveness is intricate. A highly truth-conducive practice is necessarily consensus-conducive, at least to a certain degree, for it impels proponents to assent, gradually, to one and the same position, the truth. Yet, consensus-conduciveness alone does not guarantee truth conduciveness, and can, in fact, prevent proponents from approaching the truth. Argumentative practices which are highly effective in promoting agreement tend to generate spurious consensus.

T1.5 (Space of Coherent Positions) As in the case of consensus-conduciveness, the degree of fragmentation of the space of coherent positions exerts a markable influence on a debate's veritistic dynamics and represents thus a pivotal explanatory variable. As a rule (with several notable exceptions, though), debates with a highly fragmented space of coherent positions display lower verisimilitude increase.

T2 (Background Knowledge) Background knowledge affects an argumentation's truth-conduciveness in similar ways as its consensus-conduciveness.

T2.1 (Multiplier Effect) Constant background knowledge does not simply increase the mean verisimilitude of proponents by a fixed amount but accelerates their approaching the truth, since ever more sentences can be derived from the constant body of background beliefs during a debate.

T2.2 (Favorable Fragmentation) With sufficiently many correct background beliefs, the fragmentation of the space of coherent positions turns out to be favorable rather than detrimental to an argumentation's truth-conduciveness. This effect, however, is less pronounced than with consensus-conduciveness (C2.2).

T3 (Argumentation Strategies) The veritistic value of an argumentative practice does not correspond, one-to-one, with its consensual value. A proponent's ability to track down the truth is determined by her own argumentation strategy as much as by her opponents'. We find that argumentative practices differ significantly in terms of their characteristic truth-conduciveness.

---

having taken up his position against all gainsayers—knowing that he has thought for objections and difficulties, instead of avoiding them, and has shut out no light which can be thrown upon the subject from any quarter—he has a right to think his judgement better than that of any other person, or any multitude, who have not gone through a similar process." (Mill 2009, p. 64) Criticism, though, has no intrinsic but merely instrumental epistemic value. "Such negative criticism would indeed be poor enough as an ultimate result; but as a means of attaining any positive knowledge or conviction worthy the name, it cannot be valued too highly [. . .]" (Mill 2009, p. 128).

T3.1 (Veritistic Value of Critique) As the advancement towards the truth is primarily driven by criticism (T1.3), proponents whose positions are frequently rendered incoherent exhibit a comparatively rapid verisimilitude increase. In consequence, it is the argumentation strategy employed by one's opponent, and this opponent's ability to advance critical arguments, which controls the pace at which one acquires more and more true beliefs. Proponents whose opponents argue in an aggressive and opponent-sensitive way (*undercut* rule) display, accordingly, the strongest verisimilitude rise. Opponents, in contrast, who don't address a proponent's position at all, arguing in a self-centered way, don't allow the proponent to improve her position. These findings, too, corroborate Mill's methodology of controversial debate, in particular his emphasis on being criticized by able opponents: "So essential is this discipline to a real understanding of moral and human subjects, that if opponents of all important truths do not exist, it is indispensable to imagine them, and supply them with the strongest arguments which the most skilful devil's advocate can conjure up" (Mill 2009, p. 108).

T3.2 (Veritistic Value of Plurality) Outstanding consensus-conduciveness and the inability to question (and give up) a reached consensus contributes to an argumentative practice's consensual value but tends to curtail its veritistic one. This is strikingly revealed by our simulations, where proponents who implement the *convert* rule fare poorly in terms of verisimilitude. Now, high initial disagreement and the employment of agreement-reducing strategies, side by side with consensus-conducive ones, can help to avoid the emergence and persistence of a spurious consensus and enable proponents to continue questioning their beliefs. Plurality, we find, is an instrumental epistemic virtue, and argumentative practices which explicitly cultivate it (in an, otherwise, extremely consensus-conducive climate) foster a debate's overall truth-conduciveness. Once more, Mill had it right: "[The] 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" (Mill 2009, pp. 62–64).

T3.3 (Consensus First) Aggressive and opponent-sensitive argumentation (i.e., the *undercut* strategy) represents the most truth-conducive practice in dualistic (i.e., two-proponent) debates. This is, however, not the case if multiple proponents engage in a controversy. Instead of fervently criticizing the various proponent positions simultaneously, it is more efficient to generate a consensus, possibly a spurious one, in a first step, and to criticize the consensus position (by way of self-critique) in a second step. This more conciliatory strategy, it turns out, is, in sum, more truth conducive than an immediate criticism of the diverse proponent positions. A specific version of the *convert* rule has, consequently, a role to play in truth-seeking controversies, as well.

T4 (Veritistic Indicators) Because—on average, and irrespective of the argumentative practice employed—proponent positions approach the truth only in a relatively advanced phase of a debate and since, in addition, real debates (for lack of new arguments) often don't attain these advanced phases, it becomes a decisive

question whether there are reliable methods for gauging the verisimilitude of proponent positions in an early stage of a debate. We may identify, accordingly, three veritistic indicators, whose characteristic properties are summarized below: consensus, stability, and degree of justification. Remarkably, these indicators suggest a novel, "dialectic" foundation of the two major methodologies which have been developed in philosophy of science, that is, falsificationism and verificationism.

T4.1 (Consensus) Consensus, for being possibly spurious, may obviously be a misleading indicator of truth. Still, a consensus which is reached not simply by two, but by at least five or six (independently arguing) proponents is typically a very good indicator of truth. In general, the greater the size of a consensus (in terms of proponents), the higher its expected verisimilitude. If the proponents who reach the consensus display substantial initial disagreement, the reliability tends to improve even further. The accuracy of consensus as an indicator of truth depends, moreover, on the specific argumentation strategy employed by the proponents. The more consensus-conducive the argumentative practice, the less reliable the indicator. In a highly critical controversy (proponents follow the *undercut* rule), however, even a two-proponent consensus represents a highly accurate indicator of truth, especially at an early stage of the debate. Thus, given the appropriate argumentative practice, consensus allows one to infer verisimilitude in a reliable way. Our inquiry hence confirms a very old, in fact ancient, methodological idea, which runs, for instance, already through Plato's dialogues.[6]

T4.2 (Stability) The stability of a proponent position can be measured in different ways—as agreement of the position with the proponent's initial position, or as relative frequency at which the proponent had to modify her previously held positions. No matter how one gauges stability, however, it yields a telling indicator of a position's verisimilitude at an early stage of a debate. The accuracy of stability as an indicator of truth depends on the argumentation strategies pursued by the debate's proponents. Specifically, the more critical the argumentation, the more accurate the indicator. With proponents who implement the *undercut* strategy, stability becomes in fact an extremely reliable indicator of truth. This allows us to make sense and to justify core tenets of a refined falsificationist methodology. In a modified account of his earlier views, Popper (1963), reaffirming the importance of submitting theories to severe tests (criticism), introduces the idea that the iterative process of continuous testing gradually increases the verisimilitude of our theories, namely of those which pass the successive batteries of tests. Here is

[6]Consider, for example, how Socrates, having attested to Callicles' goodwill, frankness, and (pivotally) critical competence, addresses the latter: "Well then, the inference in the present case clearly is, that if you agree with me in an argument about any point, that point will have been sufficiently tested by us, and will not require to be submitted to any further test. For you could not have agreed with me, either from lack of knowledge or from superfluity of modesty, nor yet from a desire to deceive me, for you are my friend, as you tell me yourself. And therefore when you and I are agreed, the result will be the attainment of perfect truth" (*Georgias*, 487).

a dialectic reformulation suggested by our inquiry's results: Theories (positions held by proponents) which remain stable in the face of critique are, on average, closer to the truth. Their ability to pass controversies unspoiled testifies to their verisimilitude. And criticism is crucial, as Popper rightly sees, because stability constitutes a revealing indicator of truth only on the condition that the debate is highly controversial and proponents argue in an aggressive and opponent-sensitive way.

T4.3 (Degree of Justification) The verisimilitude of a proponent's core position is, at an early stage of a debate, correlated with its degree of justification. Degrees of justification thus signal proximity to the truth. Holding a partial position with a high degree of justification is veritistically valuable. The correlation between degree of justification and verisimilitude is particularly strong if arguments are discovered randomly or introduced by proponents with a view to maximizing their positions' robustness. This relatively simple result seems to provide a justification of inductive modes of reasoning—understood as meta-reasoning on dialectical structures with probabilities interpreted as degrees of justification. As I've tried to show elsewhere, such a dialectic framework, in turn, licenses inference to the best explanation and confirmatory inferences in line with the hypothetico-deductive account of confirmation, besides Bayesian inferences and reasoning with precise probabilities (see Betz 2011c,d). With a view to apparent theoretical parallels to Wittgenstein's[7] and Carnap's[8] definition of logical probability, one may conceive these results as a *dialectic foundation of probability*.[9]

T4.4 (Methodological Trade-Off) The last two results seem to suggest that falsificationism and verificationism don't represent mutually exclusive methodologies but stand for alternative, yet equally viable ways to estimate the verisimilitude of hypotheses (proponent core positions) at an early stage of a controversy. Nothing seems to prevent one from using both stability and degree of justification as veritistic indicators. As both indicators are fairly accurate in random debates, this is principally possible. Yet, if one attempts to sharpen the accuracy of stability by stipulating that proponents argue in a highly critical way, the reliability of degree of justification as an indicator of truth is completely lost. If not a definite opposition, there seems to remain a certain trade-off between the two indicators, and, accordingly, between falsificationist and verificationist methodologies, because the accuracy of the indicators, and the reliability of the corresponding inferences, hinges sensitively on the argumentation strategies pursued by the proponents. In addition, this fact obviously complicates the application of these methods to real controversies.

---

[7] cf. *Tractatus*, 5.15.

[8] Carnap (1950).

[9] See also Betz (2010, 61). Let me clarify, however, that we have, of course, not solved Hume's problem. Far from giving a justification for induction in general, we rely, throughout this inquiry, on inductive methods, for example, in the form of basic statistical reasoning (applied to the debate ensembles).

1.6 Objections and Caveats

Before we put the approach pursued in this inquiry in perspective, relating it to other theories of rational belief dynamics in the subsequent section, we shall consider some general limitations of our approach and objections which may be raised against it. This will allow us to delineate the scope of our investigation and to announce important clarifications and caveats.

First, the proponents in our model are (unrealistically) rational. They revise their convictions but in the face of arguments, and their belief dynamics are merely determined by the inferential relations encoded in the dialectical structure. (If, however, these logical constraints underdetermine the belief revision, because there are several closest coherent positions, the proponents are indifferent and make a random choice.) Yet, we all know that our beliefs are shaped not only by arguments but by many other factors as well. Pride or arrogance might prevent proponents from taking novel arguments fully into account. Different kinds of attachment such as fondness or loyalty might cause us to lean towards the positions held by a dear proponent. Hostility or contempt, in contrast, might prevent us from conceding a point. The way a claim is framed and the way it is rhetorically presented might affect our inclination to agree or disagree. The frequency at which a statement is uttered, too, seems to influence our tendency to believe it. This illustrative enumeration raises the question whether it is admissible to ignore these factors when modeling controversies. I posit it is—as long as one reads our model as a normative one. Thus, we are not trying to give an empirically adequate account of real debates and opinion dynamics but trying to assess certain fundamental properties of debates of an ideal type, namely, their consensus- and truth-conduciveness. Rather than allowing for, say, accurate predictions of real belief formation, these studies enable us to conclude how we *should* try to argue if we are interested in achieving consensus or in finding the truth by rational argumentation.

Second, our model has it that the proponents, when pursuing an argumentation strategy, invariably succeed in designing and introducing tailored arguments. But is this even possible? Can we construct arguments at will? Or, more precisely, can one always find, given some sought-after conclusions and a set of potential premisses, a deductive argument which inferentially links some of the premisses with one of the wanted conclusions? At first glance, the answer is, plainly, no. Consider, for instance, the case where the potential premisses on the one hand and the desired conclusions on the other hand are logically independent. Then no deductive argument whatsoever links premisses and conclusion in an appropriate way. But that seems to imply that real proponents simply cannot mimic the ideal types of controversies we are studying, even if they tried hard, because the argumentation rules, so successfully employed by our model proponents, face real limitations, even logical ones. This amounts to a severe challenge, and there are two ways to address it. The first rebuttal of the challenge stresses that, while the model assumption might be unrealistic indeed, it is not that far off as the objection makes believe. Our actual capacity to design arguments is not negligible. So, while the

no-failure assumption, that is, the supposition that proponents never fail in introducing an argument which satisfies certain intended specifications, clearly overstates our abilities, the opposite no-success assumption, which presumes that proponents never find arguments that satisfy an argumentation rule, is equally mistaken. If a person, engaged in a debate, attempts to support her own position, and tries to advance corresponding arguments, she usually succeeds in doing so, at least from time to time. Likewise, if one attempts to criticize an opponent, searching for arguments that attack the opponent's claims, this is far from being a hopeless endeavor and leads, albeit not always and invariably, to the introduction of suitable arguments. In sum, I concede that the no-failure assumption, built into our model, is unrealistic and, strictly speaking, unattainable. Yet, our argumentative practice teaches us that we can successfully design arguments to some degree. With a view to identifying and contrasting the effects of different argumentation strategies (which, admittedly, can only be imperfectly applied in real debates), the no-failure assumption seems to me, nevertheless, an appropriate first-order approximation, since it allows for the most distinct assessment of the (purely applied) argumentation rules. Let us now turn to the previously announced, second rebuttal of the objection against the no-failure assumption. It requires that we step back for a moment and reflect on different interpretations of our simulated dialectical structures. Recall that real debates, no matter whether they unfold in an oral or a written exchange, don't consist in deductively valid arguments and don't explicitly realize a dialectical structure. It takes a substantial amount of analysis and reconstruction to transform a raw argumentation into well-formed arguments and a uniform dialectical structure. The arguments in such a reconstructed dialectical structure are deductively valid by virtue of the reconstruction, and relative to a given logic, the reconstruction logic, which the analysis rests upon. The first and obvious interpretation of our simulated dialectical structures is to understand them as debate reconstructions, containing arguments that are valid inferences relative to some reconstruction logic, with all premisses made explicit. It is within this first interpretation where the doubts about the no-failure assumption (Can one always establish, given the reconstruction logic, sought-after inferential relations between some sentences?) arise. Yet, an interpreter, reconstructing a debate, has not merely some leeway in choosing the reconstruction logic but may also decide on the reconstruction's "degree of explicitness." Thus, she may judge that, for example, mathematical principles, which are not warranted by the reconstruction logic itself but which are nevertheless universally agreed upon, might be omitted and don't have to be explicitly stated as premisses. Likewise, in a reconstruction of a debate about a company's future investment strategy, physical knowledge might equally be taken for granted and might hence not be presented explicitly in the reconstructed arguments. In Sect. 1.3 above, we introduced a direct representation of background beliefs, namely as fixed truth-value assignments. Here, we spot a second, indirect representation of background knowledge: Background beliefs might simply be omitted in the reconstructed debates, so that only premisses which don't belong to the body of background beliefs are made explicit. Now, this yields a second interpretation of simulated dialectical structures: The modeled arguments stand for deductively valid

inferences (relative to the reconstruction logic) which derive a conclusion from the explicitly stated premisses *and* the (implicit) global background beliefs. With this interpretation, designing arguments, which take off from given premisses and back a specified conclusion, becomes a completely different task: It doesn't merely consist in scrutinizing and crunching the inferential relations between the sought-after premisses and the wanted conclusion but comprises, primarily, the search for appropriate background beliefs (not made explicit in the argument) together with which the explicit premisses imply the conclusion. Clearly, this is much less difficult a task than finding a logical relation between a couple of statements held by a proponent, especially if the proponents have, implicitly, a lot of sufficiently diverse beliefs in common. In that case, a proponent can typically find an argument which inferentially links some premisses—on the basis of further implicit and shared background assumptions—with a wanted conclusion and which dovetails with a given argumentation rule. The idea that proponents can construct arguments which fit a given argumentation rule becomes even more plausible, if we consider that, in some controversies, proponents can *generate* shared background beliefs, for example, by collecting specific observational data or carrying out experiments. So, the second rebuttal stresses that the no-falsity assumption ceases to be unrealistic, and arguments can be successfully designed indeed, if the (implicit) body of background knowledge is sufficiently broad.

Third, the simulations, in particular those which are supposed to study the veritistic dynamics of debates, stipulate that some given position is correct. This position, the truth, is chosen randomly (at the very beginning). But, then, don't the simulations merely demonstrate, for example, how quickly proponents approach an arbitrarily selected position and not how rapidly they find the truth? In the end, the truth is not simply a randomly chosen position, is it? Admittedly, the specific setup of the simulations is prone to triggering confusions and worries of this kind. So let me try to clarify, with the help of an analogy, why choosing the true position randomly when initializing the simulation is not only unproblematic but even necessary in order to obtain meaningful results at all. Picture an engineer who has designed a machine which is supposed to test freshly fabricated footballs. Specifically, the machine scans a football's skin by moving a highly accurate and well-tested sensor over its surface. The sensor's path is determined by a complicated algorithm. Yet, before the procedure is employed at a large scale, the manufacturer urges to assess its reliability, that is, its ability to track down fissures in a football's hull. With the sensor itself being well tested, the crucial ingredient is the algorithm that prescribes the sensor's path. Thus, the engineer sets up a simulation which represents (1) a damaged football and (2) the sensor moving over the football's surface according to the corresponding algorithm. This simulation determines, for a given initial position of the sensor and the location of the fissure, whether the sensor, controlled by the algorithm, moves over the fracture, or not. Based on sufficiently many simulations, with varying initial conditions, the engineer is in a position to assess whether the algorithm reliably prompts the sensor to move over the fissure. Or, to put it—with a view to our analogy—differently, by way of simulating the procedure, the engineer assesses its instrumental value with respect to tracking

down the rupture. Now, it is not merely perfectly fine, but even crucially required that the simulations assume that the fracture be located on a randomly chosen spot on the football's surface. For the engineer wants to assess the procedure's reliability in regard to finding any fracture, no matter where it is located. To assume, in contrast, that the football is damaged at a very specific spot, say 5 cm south of the sensor's initial location, is obviously not very helpful, since it doesn't evaluate the procedure's ability to detect damages at different places. In close analogy, it is crucial for our simulations that we don't make any (arbitrary!) assumptions about the particular location of the truth within the space of coherent positions. So, by presuming that the truth is an arbitrary (randomly chosen) position, we avoid, in fact, fatal arbitrariness and ensure that our simulations assess the veritistic value of controversial argumentation, that is, its instrumental value with respect to tracking down the truth.

Fourth, our inquiry is, critically, language-relative. Thus, we assess the consensus- and, in particular, the truth-conduciveness of controversial argumentation under the assumption that proponents speak a common and unvarying language. More seriously, we frame the truth by identifying it with a fixed position, within a given conceptual scheme. But then, one may object, we don't assess the ability of argumentative practices to track objective, in the sense of language-independent truth.[10] I suggest that the appropriate response to this alleged *reductio* consists in embracing its conclusion. Indeed, we suppose that, throughout a debate, proponents speak one and the same language. And we study which argumentative strategies allow them to achieve their epistemic aims, given the linguistic and conceptual means they have. Furthermore, the concept of truth we posit is, in fact, not a metaphysical one. This book's investigation doesn't attempt to demonstrate how to reach, by way of controversial argumentation, a mind- and language-independent, eternal and infallible truth. Hence, it doesn't address the fundamental skeptical challenge, which vexes traditional epistemology, either. In contrast, our investigation into the veritistic value of argumentation is based on an internal rather than metaphysical realism. It builds, accordingly, on a language-relative (yet nonetheless objective) notion of truth, as introduced by Carnap (1956), later defended by Putnam (1981), and, I take it, proficiently wrapped up by Kitcher (2001). As a consequence, we assess the ability of proponents, who engage in a debate, to track down the truth given the language they speak. Such an ability, however, does not imply that the verisimilitude of proponent positions thus attained is invariant to translations into other languages. So, let us assume, for the sake of illustration, that a controversial debate has led proponents reliably towards the truth. Assume, in addition, that the proponents' positions have to be translated, subsequently, into a new language because the proponents have decided to modify their conceptual scheme substantially. Now, the translated proponent positions, in spite of having emerged from an argumentation (though in a different language), might be completely wrong. In other words, the

[10] A similar objection is advanced, more generally, against explications of verisimilitude (cf. Oddie 2008).

presumed ability of controversial debate to track down the truth does not guarantee that proponent positions retain a high verisimilitude once the underlying conceptual scheme is changed. This is important to notice since it draws a relevant limitation to our inquiry's scope: Thus, we disregard, and exclude from our investigation, far-reaching conceptual change, which plays, for instance, an important role, as Kuhn (1962) famously argues, in some ("revolutionary") scientific controversies.

Fifth, the evaluation of controversial argumentation, and hence our results about its consensual and veritistic value depends not only on the common language spoken by the proponents but on the reconstruction of the natural language argumentation as well. Since a precise reconstruction of a debate is seriously underdetermined by the speech acts which the proponents actually advance and since, more specifically, an interpreter, when reconstructing an argumentation, may choose, more or less at liberty, how to individuate single premisses, thereby determining the number of premisses per argument as well as the number of sentences in the dialectical structure, the assessment of a debate within the framework of the theory of dialectical structures appears to be, by and large, arbitrary. All the crucial evaluative variables seem to hinge on the interpreter's subjective choice: the degree of agreement between proponents, the verisimilitude of a position, a debate's inferential density, a partial position's degree of justification, the stability of a proponent position, etc. And this seems to imply that not the argumentative practices but rather the way a debate is interpreted determines whether the proponents have reached consensus or attained the truth. This objection doesn't, in the first place, criticize our simulation studies but questions the applicability of our findings to real debates. Now, the underdetermination of a detailed debate reconstruction is clearly a hermeneutical challenge, yet I doubt it undermines our investigation. Let's assume, for the sake of the argument, that the dialectic evaluation of a debate depends in fact sensitively on arbitrary choices of the interpreter. This alone does, however, not interfere with a meaningful assessment of how evaluative variables (such as agreement, degree of justification, etc.) have evolved during one and the same debate or are correlated with each other—provided the arbitrary hermeneutic decisions are not varied in the reconstructions of the debate's consecutive states. If, for example, the interpreter reconstructs a given reason as an argument containing three premisses in the initial phase of the debate, she must reconstruct it in the same way in later phases. What is, admittedly, obstructed by hermeneutic underdetermination is a sound interdebate comparison of evaluation results, such as, for instance, juxtaposing the verisimilitude of partial positions in two different debates. Still, an assessment of the consensual and veritistic value of argumentative practices *relative to* the corresponding debate's starting point is all we are aiming at in this inquiry. So we examine, for example, which argumentation strategies tend to improve a proponent's veritistic situation in the course of a debate. And this sort of inquiry is not threatened by hermeneutic underdetermination. An analogy may clarify the rebuttal even further. Consider climatologists who try to assess the impact of a volcanic eruption in the nineteenth century on worldwide surface temperatures. The scientists base their investigation on temperature records from dozens of meteorological stations on different continents. Unfortunately, though, the stations used instruments which

were not calibrated against each other, and the measurements of the different stations cannot be directly compared with each other, in consequence. This does, however, not interfere with a meaningful assessment of each station's temperature record—provided the stations didn't modify their instruments (within the relevant period). So for each individual station, the temperature effect of the volcanic eruption might very well be gauged. Moreover, by normalizing the data record with respect to the temperature before the eruption, the *relative* global mean effect of the eruption (e.g., $-0.5\,\mathrm{K}$) can be estimated. By analogy, we can assess the instrumental consensual and verististic value of different argumentative practices, even if evaluative variables were, in absolute terms, hardly comparable across different debates.

Sixth, by assuming that all arguments contained in a dialectical structure are deductively valid, we seem to pay no attention to inductive reasoning, which undeniably plays a crucial role in real debates and which certainly contributes to the consensual and veritistic value of controversial argumentation. This objection raises, of course, not only a challenge for our inquiry but for any approach in argumentation theory which subscribes to the principle of deductivism, that is, the view that all arguments, including the allegedly inductive ones, can and, ultimately, should be reconstructed as deductively valid inferences. I'm certainly not able, in the remainder of this section, to defend this view in depth. Such a defense, I suppose, requires to go through all types of so-called inductive arguments (e.g., reasoning by analogy, inference to the best explanation, statistical inference, enumerative induction), which allegedly resist a charitable (!) deductive interpretation, and to suggest, for each such type of argument, how to reconstruct it, deductively, in an appropriate way. In the following, however, I'd merely like to highlight (a) two different general reconstruction strategies which typically prove useful when reconstructing inductive reasoning, and to outline (b) a further possibility for embedding inductive modes of reasoning in the framework of the theory of dialectical structures. Note, ad (a), that every inductive argument relies on a specific inductive inference rule. An inductive inference rule has characteristically a different status than a logical one: It doesn't hold necessarily on account of certain logical constants, it may only be applicable as long as certain conditions are shown to prevail, and it might cease to be a sound inference rule as soon as new evidence emerges. Now, one straightforward strategy for reconstructing an inductive argument as deductively valid consists in making the inference rule, plus its additional applicability conditions and restrictions (e.g., ceteris paribus or total evidence clauses), explicit by stating it as a premiss of the argument. Consider, as an example, the following reasonably good inductive argument:

(P1) Tara is Indian.
(P2) Most Indians (>80%) are Hindi.
(C) Thus, Tara is Hindu.

A charitable deductive reconstruction makes the underlying inference rule explicit and adds the following additional premisses:

(P3) If (1) most F are G, (2) a is F, and (3) a being F represents our total evidence relevant to the question whether a is G or not, then a is G.

(P4) Tara being Indian represents our total evidence relevant to the question whether she is Hindu or not.

By adding (P3) and (P4), we obtain a deductively valid, monotonic argument. In particular, a defeat of the original inductive argument, such as learning that Tara lives actually in the region of Punjab, ca. 90% of whose inhabitants are Muslims, does not miraculously undermine the inference anymore but can now be explicitly related to premiss (P4), which becomes false as the new evidence surfaces. Furthermore, we may reconstruct the modified inductive inference, which makes use of the novel evidence, as follows:

(P1) Tara is an Indian living in Punjab.
(P2) Most Indians living in Punjab are Muslims.
(P3) If (1) most F are G, (2) a is F, and (3) a being F represents our total evidence relevant to the question whether a is G or not, then a is G.
(P4) Tara being an Indian living in Punjab represents our total evidence relevant to the question whether she is Muslim or not.
(C) Thus, Tara is Muslim.

Besides making the underlying inductive inference rule explicit, qualifying the conclusion of an argument represents a further valuable maneuver when reconstructing inductive arguments. More specifically, it might be necessary to insert probabilistic or epistemic operators so as to obtain plausible premisses and a charitable reconstruction. Likewise, our illustrative reconstruction might be further improved along the following lines:

...
(P3') If (1) most F are G, (2) a is F, and (3) a being F represents our total evidence relevant to the question whether a is G or not, then it is *likely/very likely/reasonable to accept/permissible to assume in further arguments...* that a is G.
...
(C') Thus, it is *likely/very likely/reasonable to accept/permissible to assume in further arguments...* that Tara is Muslim.

Demonstrating that inductive arguments can in fact be reconstructed as deductive inferences in a charitable way represents an effective rebuttal of the objection to deductivism. Yet, ad (b), a further and in some sense even more interesting rejoinder embeds inductive reasoning within the theory of dialectical structures not merely by interpreting these arguments as deductive inferences but by showing that inductive arguments can be understood as meta-inferences on dialectical structures. Thus, I have tried to explain, in separate articles, how (1) inductive reasoning in line with the hypothetico-deductive account of confirmation (Betz 2011c) and (2) inferences to the best explanation (Betz 2011d) may be understood as meta-syllogisms on a given dialectical structure. Moreover, by establishing that a partial position's degree of justification represents a significant indicator of verisimilitude (see T4.3 above), this inquiry strongly supports those results and contributes to a dialectic foundation of inductive inferences based on degrees of justification. All this dismantles the fear of our inquiry not fully and adequately accounting for inductive modes of reasoning.

1.7 Putting the Approach in Perspective

The endeavor to model the dynamics of belief change, and to understand the rationality thereof, is far from being novel. The investigation carried out in this book thence relates to a couple of alternative approaches in epistemology, philosophy of science, logic, and artificial intelligence, which attempt to explain the dynamics of rational belief formation and revision.

A major dimension along which these approaches can be ordered is the degree of logical competence which agents are assumed to possess according to the corresponding approach. One may, for example, assume that agents are logically omniscient, being aware not only of all inferential relations within a given set of sentences but even of all logical implications some sentence carries. This amounts to maximal logical competence and represents an extreme assumption in the spectrum we are considering. At the opposite side of this spectrum lies the presumption of minimal logical competence, or, as we shall call it, "logical ignorance." Agents are (modeled as) logically ignorant if they don't take account of any inferential relations between their convictions when revising their beliefs. This is in particular the case if a model of belief change doesn't represent inferential relations between sentences in the first place.

Models which presume that agents are logically omniscient comprise epistemic logic (Fagin et al. 1995), including dynamic extensions (van Ditmarsch et al. 2007) and theories of belief revision (Hansson 1999), in particular the so-called AGM model (Alchourron et al. 1985; Gärdenfors 1988), as well as, though to a lesser extent, argumentation frameworks as developed in artificial intelligence (Chesñevar et al. 2000; Prakken and Vreeswijk 2001; Bench-Capon and Dunne 2007).

Epistemic logic extends first-oder predicate logic by a knowledge operator K_i, allowing for the logico-semantical analysis of statements about an agent i's knowledge and for the expression of epistemic principles in the corresponding formal language. The syntactic calculus of epistemic logic is complemented by a possible world semantics (as proposed by Hintikka 1962) to the effect that $K_i p$ is interpreted as "p holds in every possible world which is compatible with what agent i knows." But as (1) a logical truth holds in every possible world, and (2) if p holds in some possible world then all its logical implications hold in that very world as well, we have $K_i p^*$ (with p^* being an arbitrary logical truth) and $K_i p \implies K_i q$ (with q being an arbitrary logical consequence of p) for every agent i. In other words, according to epistemic logic, agents are logically omniscient and hold deductively closed knowledge claims (see also Fagin et al. 1995, pp. 333–337; Hendricks 2006, p. 98).

The AGM model, named after its original authors Carlos Alchourrón, Peter Gärdenfors, and David Makinson, represents an agent's beliefs as a set of sentences in some formal language. Belief revision theories in the tradition of the AGM model study the principles of how an agent's overall belief set ought to change given (1) the acquisition of some new belief (*expansion*), the dismissal of some previously held belief (*contraction*), or the replacement of previously held beliefs

by new ones (*revision*). Now, it represents a fundamental assumption of this approach, which seems to be required, as Hansson (2009) notes, in order to carry out an interesting formal treatment in the first place, that an agent's belief set be closed under logical implication. That is, agents are assumed to be logically omniscient. The AGM model has been used, recently, to investigate whether and under which conditions belief revision increases the verisimilitude of an agent's beliefs (see the special issue of *Erkenntnis*, in particular Kuipers and Schurz 2011). The ongoing research effectively brings together AGM theory and the program of (logically) explicating the concept of verisimilitude (cf. Niiniluoto 1998; Oddie 2008). Moreover, rather than simply trying to pin down precisely the notion of truthlikeness, such investigations take on the methodological challenge as formulated, for example, by Zamora Bonilla (1992, 2000), namely, to spell out how (i.e., through which methods) the verisimilitude of a belief set can be increased. Yet, results by Niiniluoto (2011) suggest that belief revision does not necessarily help agents to approach the truth. In general, these specific approaches, while being driven by a similar research interest than this study, remain committed to the assumption of logical omniscience.

Researchers in artificial intelligence (AI), taking Reiter's default logic as a starting point (Reiter 1980), have developed, in recent decades, a variety of approaches to modeling complex argumentation. In AI, controversies are typically analyzed as "argumentation frameworks." Although these theories of argumentation frameworks are not primarily concerned with the rational *dynamics* of belief change but attempt to model, rather, a static knowledge base in terms of its arguments, we shall nevertheless briefly consider them here on account of their resemblance with the theory of dialectical structures (see also Sect. 2.1). Some theories of argumentation frameworks, specifically those in the tradition of Dung (1995), are not explicitly based on formal logic at all, so it is not fully correct to say that these models assume agents to be logically omniscient in a strict sense. Still, I maintain that they (implicitly) assume agents to be logico-argumentatively omniscient, namely inasmuch as the corresponding evaluation procedures suppose that all potentially relevant arguments be taken into consideration when assessing a controversial claim. Let me illustrate this diagnosis with respect to the highly influential approach of Dung (1995). Dung takes it that "[for] a rational agent G, an argument A is acceptable if G can defend A (from within her world) against all attacks on A." (Dung 1995, p. 326) Yet, unless the argumentation framework contains every argument that can possibly be advanced at all (and unless an argument is, consequently, inacceptable if and only if it is simply not possible to defend it against an attack), Dung's explication of the fundamental notion of acceptability appears to be inappropriate, for a rational agent with limited cognitive capacities might, even in the face of undefeated counterarguments against her claim, stick to her position by simply saying that she does not accept the counterargument (denies one of its premisses), without being able, as of today, to back up that refutation with an extra argument. Note that a similar assumption is also built into the model developed by Besnard and Hunter (2008): Assuming that an argument which is not attacked has to be conceded by a rational proponent (p. 108) makes only sense insofar as

the argumentation framework contains all relevant arguments which can possibly be advanced in the debate. In sum, the evaluation procedures established by AI models of complex argumentation seem to suppose that agents are, in a broader sense, logically omniscient as well.

Unlike the approaches considered so far, other models of rational belief dynamics don't represent, at least not explicitly, inferential dependencies between an agent's beliefs at all and hence seem to assume that agents are logically ignorant. The archetypal models of rational consensus formation and opinion dynamics developed by Lehrer and Wagner (1981), on the one hand, and by Hegselmann and Krause (2002), Hegselmann (2004), on the other hand, belong to this type (for a review which focuses on veritistic opinion dynamics see also Douven and Kelb 2011).

Both the Lehrer–Wagner and the Hegselmann–Krause models represent an agent's belief as a real number in the interval [0;1] and assume, at least in their most basic variants, that an agent's belief at step $t + 1$ is fully determined by that agent's as well her peers' beliefs at step t. Now, the models diverge in terms of how the belief of agent i and the beliefs of i's peers are aggregated so as to yield the updated belief of i. In the Lehrer–Wagner model, each agent i assigns constant real numbers to her peers and herself, assessing the agents according to their alleged expert status. An agent i's new belief is then defined as the weighted average (based on the weights assigned by i) of all agents' previous beliefs. Lehrer and Wagner (1981) demonstrate that the agents' beliefs necessarily converge (if weights are greater than 0), and postulate that the resulting opinion dynamic represents a rational procedure for consensus generation. In the Hegselmann–Krause model, an agent i doesn't consider the opinions of all other agents when updating her belief but merely those peer beliefs which fall within a certain ε-interval ($\varepsilon > 0$) around i's belief. Agents, according to the intended interpretation, merely take those peers into consideration whose opinions are not too far off. Agent i's new belief is the plain average of all opinions that fall in her ε-interval. Hegselmann and Krause (2002) simulate the ensuing opinion dynamics and show, for example, that the size of the confidence interval (ε) crucially affects whether the agents settle on a consensus position or not. Extending the basic model, Hegselmann and Krause (2006) stipulate that some real number is the correct opinion and assume that a few agents possess the ability to track the truth: The beliefs of truth trackers are both affected by the peers within the corresponding confidence interval and attracted by the truth. Although this extension provides interesting new results, it doesn't amount to an *explicit* inclusion of inferential dependencies in the model. In sum, the Lehrer–Wagner as well as the Hegselmann–Krause model disregard inferential relations which hold between the agents' beliefs altogether, and conceive agents, accordingly, as logically ignorant.

Extending the Hegselmann–Krause model, Riegler and Douven (2009) represent an agent's belief state by a binary evaluation of a propositional basis (rather than a single real number). The modified model, unlike the original one, hence contains a pivotal element of the theory of dialectical structures. The specific propositional basis employed by Riegler and Douven (technically: the canonically ordered set of state descriptions Riegler and Douven 2009, p. 150) allows even to encode

inferential relations between different sentences. Consequently, the extended model comprises a rich representation of opinion sets. As a drawback, however, Riegler and Douven (2009) have to assume that belief states of agents be closed under logical implication. In other words, their modification of the Hegselmann–Krause model relies on the assumption of logical omniscience.

The theories of belief dynamics which presume agents to be logically omniscient on the one side, and the models of opinion dynamics which don't represent inferential dependencies at all on the other side constitute opposite poles of a spectrum of approaches to modeling rational belief dynamics and are both characterized by equally extreme (and, on the face of it, unrealistic) assumptions about the logical competence of agents and its role in rational belief formation. That is precisely what sets the model unfolded in this book, which falls well in between these two extremes, apart from previous approaches. For we assume, on the one hand, that agents are not logically omniscient. Instead, they consider merely the inferential relations discovered so far (i.e., the arguments explicitly introduced into a debate), when inspecting and, if necessary, revising their beliefs. But this is of course, on the other hand, far from presuming that agents don't consider any logical dependencies whatsoever when updating their beliefs. Agents, as modeled by the theory of dialectical structures, are anything but logically ignorant. By acknowledging the actual cognitive limitations of agents who engage in an argumentation, our approach can be understood as a bounded rationality model (e.g., Simon 1982) of belief dynamics.

The previous remarks, however, are not supposed to discard the alternative approaches simply on the grounds that they rely on specific idealizations. Models that assume logical omniscience, for instance, study how ideal agents should form and modify their beliefs (likewise, Levi (1991, p. 8) conceives these models as analyzing an agent's commitments rather than her consciously held beliefs), possibly yielding significant epistemological insights which might, in turn, bear on our everyday epistemic practices. Moreover, these models constitute, obviously, adequate representations of computational multiagent systems and give rise to valuable applications in artificial intelligence and computer science. The Lehrer–Wagner and Hegselmann–Krause models, however, representing agents, ostensibly, as logically ignorant, can be understood as highly aggregated models of opinion formation. By concentrating on the macrodynamics of belief revision, they deliberately disregard detailed (micro) argumentative processes and seek to capture dialectic mechanisms through (1) the general opinion-averaging (assuming that arguments increase peer agreement) and (2) the truth-tracking procedure (assuming that arguments increase verisimilitude). From this perspective, the approach unfolded in this book can be interpreted as a model of the micro- and mesodynamics of rational debate, thus complementing the Lehrer–Wagner and Hegselmann–Krause models, which attempt to represent the corresponding macrodynamics.

Still, notwithstanding other fruitful applications of models with logical omniscience or ignorance, I take it that models of this type provide in any case poor representations of the *detailed* dynamics of rational debates and of the belief change which is triggered by controversial argumentation: Rational proponents who engage

in a debate undeniably adjust their position in the face of new arguments, without, however, being logically omniscient (which would render the entire *process* of argumentation, that is, the introduction of new arguments, superfluous and the corresponding real-world *practice* incomprehensible).

In the remainder of this section, we discuss two further approaches to modeling doxastic dynamics, which resist a straightforward subsumption under the opposite types of alternative theories previously considered. These approaches are, first, Paul Thagard's theory of explanatory coherence and (scientific) controversy and, second, theories of judgement aggregation, which constitute a lively research area, bringing together economists, sociologists, political scientists, scholars of law, computer scientists, and philosophers alike.

In his book *Conceptual Revolutions*, Paul Thagard develops a theory of explanatory coherence with a view to understanding the dynamics of scientific controversies (Thagard 1992, Chap. 4). Thagard considers propositions which state (1) the available observational evidence and (2) the proposed hypotheses at a given state of debate. He represents several relations which may hold between these propositions. Pivotally, his model maps explanatory relations between tuples of hypotheses on the one side and observational statements on the other side. Thagard specifies seven general principles which allow one to translate the explanatory links between propositions into a symmetrical relation that indicates how strongly two individual propositions cohere. A connectionist computer program is then used to determine which hypothesis coheres best with the given observational evidence. By applying this method to consecutive states of a scientific controversy, Thagard seeks to explain its evolution.

Although we share with Thagard the aim to understand the dynamics of rational controversies, Thagard is primarily interested in explaining theory change in science (e.g., Why is it that some hypothesis was well corroborated at state t_1 but justified to a much lesser degree at a later state t_2 of the debate?), whereas the scope of our inquiry surely covers but is not restricted to scientific controversies. Moreover, by evaluating debates in terms of coherence, Thagard's model neither takes account of the agreement between proponent positions nor of their verisimilitude. As a consequence, it cannot assess the consensus- and truth-conduciveness of controversial argumentation, which is this inquiry's main mission. There are, of course, major differences concerning the specific representation of a debate as well. Most importantly, Thagard's account does not, unlike the theory of dialectical structures, represent inferential relations between the statements which figure in a debate (except for contradiction). His approach does thence not qualify as an argumentation-theoretic one in the first place. Yet, in spite of these basic theoretical differences, Thagard's method and the theory of dialectical structures yield seemingly similar results when applied to real debates. Specifically, Thagard's analyses of scientific controversies as explanatory maps evoke, immediately, argument maps that visualize dialectical structures. This superficial resemblance stems from the fact that an explanatory link, relating a couple of hypotheses on the one side with an observational item (the explanandum) on the other side, calls for an interpretation as argument (with the explanandum as conclusion), and

a corresponding reconstruction according to the theory of dialectical structures. Hence, Thagard's concrete applications might actually be neatly transferred into the framework adopted throughout our inquiry.

Theories of judgement aggregation (cf. List and Puppe 2009; List and Polak 2010) study methods for merging various judgements (or sets of judgements, i.e., proponent positions) into a single, collective judgement (or a set of judgements, i.e., a proponent position). At the heart of this research program lies the observation that simple majority voting on individual sentences might aggregate consistent individual positions into an inconsistent collective one. So, consider three agents who assign truth values to the sentences $p \vee q$, $\neg q$, p as follows:

Agent	$p \vee q$	$\neg q$	p
1	True	True	True
2	True	False	False
3	False	True	False

Clearly, each agent holds a logically consistent position. Now, assume it as required to aggregate the mutually distinct positions into a collective judgement about the corresponding three sentences (e.g., because the agents belong to a jury in a U.S. court, or to a scientific advisory body, or to the Cabinet). A straightforward method for doing so is majority voting with respect to the individual statements. This yields, however,

	$p \vee q$	$\neg q$	p
Maj	True	True	False

which is an inconsistent truth-value assignment, hence the "discursive dilemma," as this problem is also referred to. Theories of judgement aggregation seek and study procedures for combining judgements which don't result in inconsistent collective judgements, provided the individual agents hold consistent positions. In analogy to Arrow's impossibility theorem for preference aggregation (Arrow 1963), List and Pettit (2002, 2004) have proven impossibility theorems for judgement aggregation, which demonstrate, generally, that such procedures cannot simultaneously meet a set of given, sought-after criteria.

Now, how do theories of judgement aggregation relate to the model of debate dynamics presumed in this inquiry? To begin with, theories of judgement aggregation neither assume agents to be logically omniscient nor to be logically ignorant. Instead, the discursive dilemma arises, and can be studied, based on the assumption of limited logico-argumentative capacities, which dovetails with the theory of dialectical structures. However, theories of judgement aggregation don't investigate how the beliefs of individual agents change given the introduction of new arguments or the discovery of new evidence. They presume, in contrast, that

the rational debate has come to standstill, without having generated a universal consensus. The question addressed by theories of judgement aggregation reads: What should we do if (a) a consensus position has to be reached—for whatever reasons, if (b) the process of giving and taking reasons has come to an end because no new arguments or facts pertaining to the debate are discovered anymore, and if (c) a residual dissent persists nevertheless? This is of course an interesting and relevant question, yet it concerns a completely different phase of collective belief formation than the one studied in this book. Our investigation assesses the consensus- and truth-conduciveness of controversial argumentation, studying, in particular, whether proponent positions approach each other—and the truth—in the course of a debate, that is, by way of introducing new arguments. The point at which no new arguments are discovered, at which a controversy ends, delimits the scope of our inquiry. But it is precisely at this point where theories of judgement aggregation set in. So, a model of debate dynamics on the one hand and theories of judgement aggregation on the other hand, by virtue of relating to consecutive phases of social belief formation, rather complement, than compete with each other.