Evaluating dialectical structures

Gregor Betz
Freie Universität Berlin
Institute of Philosophy
g.betz@philosophie.fu-berlin.de

February 26, 2007

Abstract

This paper develops concepts and procedures for the evaluation of complex debates. They provide means for answering such questions as whether a thesis has to be considered as proven or disproven in a debate or who carries a burden of proof. While being based on classical logic, this framework represents an (argument-based) approach to non-monotonic, or defeasible reasoning. Debates are analysed as dialectical structures, i.e. argumentation systems with an attack- as well as a support-relationship. The recursive status assignment over the arguments is conditionalised on proponents in a debate. The problem of multiple status assignments arising on circular structures is solved by showing that uniqueness can be guaranteed qua reconstruction of a debate. The notion of burden of proof as well as other discursive aims rational proponents pursue in a debate is defined within the framework.

1 The problem

Starting with Aristotle, logicians and philosophers have developed means for evaluating single arguments. In a nutshell: Once a reasoning is reconstructed as an argument—what is anything but trivial as Tetens (2004) reminds us—two question are to be addressed: (1) Are the inferences valid? (2) Are the premisses true? If both are answered with yes, the argument is sound and its conclusion is true.[1]

[1]This general approach which considers the analysis of natural language argumentation as an application of formal logic has been critizised in fields of argumentation theory following Toulmin (1958) and in so-called informal logic (e.g. Beardsley, 1950; Nolt, 1984; Govier, 1985; Thomas, 1986). There remain, however, two reasons for reconstructing arguments as premiss-conclusion structures and applying formal logic to evaluate them. The first is the universal applicability of this method: every reasoning can be reconstructed as a premiss-conclusion structure. The second consists in the fact that the deductive rules of inference can be considered a minimalist set of rules of sound reasoning against which further rules such as rules of inductive reasoning can be explicated by adding them as

The purpose of this kind of analysis is obviously a normative one. We want to know: Is some thesis t true? Do these reasonings represent good grounds for believing that t? What else would I have to believe in order for these reasonings to represent compelling grounds for t? Applied logic has turned out to be a suitable framework for adressing these questions—at least as long as individual reasons are to be evaluated separately.

Yet real debates, whether about war, social security, climate change, or such peculiar topics as epiphenomenalism never consist of isolated arguments. The central theses of these debates are typically supported or attacked by multiple arguments which in turn are themselves backed or challenged by further arguments. Given such a variety of conflicting, supporting and attacking arguments, can the central thesis still considered be well justified, or true? Supposed I wanted to claim t, what else would I have to maintain in order not to contradict myself? Is my position coherent in the light of all these arguments at all?

This paper adresses the problem how such questions can be answered systematically. Building on applied logic, it aims at providing methods for the evaluation of complex argumentation which consist of many arguments. While the next section places my approach in the wider context of argumentation theory, section 3 briefly presents the general framework of the theory of dialectical structures as developed so far: dialectical structures will be described as a specific type of argumentation system with two types of relations—support and attack—holding between the individual arguments. As a next step, a status assignment ("validity function") on these specific argumentation systems will be defined while considering in particular whether and in which way the support relationship enters the recursive definition of an argument's status (section 4). In order to apply the evaluation procedure to many-proponent debates, the validity function will be conditionalised relative to a proponent's position (section 5). The last two sections (6 and 7) address the problem of multiple status assignments which arises if a dialectical structure is circular. It will be argued that uniqueness and existence of a validity function can be guaranteed qua reconstruction of a debate.

2 Approaches to debate evaluation

The following section briefly reviews some approaches to the evaluation of complex argumentation, pinpointing similarities and differences to the framework I will suggest below.

As a first point to note, the technique of visualising complex argumen-

implicit premisses to an argument reconstruction. By saying so, one is in no way commited to claiming that classical logic is the most appropiate logic to evaluate natural language arguments.

tations and mapping the dialectical relations between arguments has been developed as a major tool for investigating the structure of argumentation in informal logic (see Snoeck Henkemans, 2000). Still, the similarities to the theory of dialectical structures are very limited as the latter builds on formal logic and analyses arguments as premiss-conclusion structures, not as single propositions.

Secondly, some approaches to debate evaluation can be described as game-theoric (or dialectical) in the following sense: They analyse controversial debates as games played by different proponents. A specific approach, then, consists in a set of procedural rules rational proponents are supposed to follow. Moreover, the rules determine under which conditions a proponent has won the game. Rescher (1977) is an important example of such a primarily game-theoretic or rule-based approach.[2] Accordingly, debate evaluation consists in verifying whether the proponents have played by the rules and, finally, determining the winner. While I do not deny that a good theory of argumentation should say something about rules of rational discursive behavior[3], I maintain that such dialogue rules are, at least, not sufficient for debate evaluation. A first problem the game-theoric approaches face is that we do not always know the exact order in which arguments that make up a controversial debate have been put forward. Even worse, the arguments one might consider in a complex argumentation could originally have occured in very different contexts, their authors not knowing each other. If I compare for instance the reasonings of philosophers who lived at very different times at very different places, yet who adressed the very same problem, say the problem of free will, it is nonsense to ask whether these philosophers played by the rules when putting forward their arguments: they didn't play with each other at all. Still, whether the central thesis their arguments aim at is true, and whether it can be held true consistently in the light of these arguments represents an entirely meaningful question which deserves an answer. To strenghten the critique, one might generally wonder whether the order in which certain arguments are put forward has or should have any effect on the evaluation of the final complex argumentation. A single argument is deductively valid or sound independent of how it has been constructed. Similarly, a thesis is well justified in a complex argumentation irrespective of

[2]Others include, for example, Hamblin (1971), Mackenzie (1979), Hintikka (1981), Walton and Krabbe (1995), and van Eemeren and Grootendorst (2004), see also Prakken (2000). Furthermore, the work by Lorenzen and Lorenz (1978) is sometimes mentioned in this context. Their dialogue games, however, were not meant to represent rules rational argents are supposed to follow in a discussion and which thence would enable us to evaluate a debate, rather, they should serve as a constructivist semantic for different systems of formal logic.

[3]Procedural rules provided in addition to essentially non-procedural evaluation procedures and argumentation systems clearly shed an interesting light on the latter (e.g. Baker and Ginsberg, 1989; Vreeswijk, 1993; Simari et al., 1994; Prakken and Sartor, 1996; Vreeswijk, 1997; Vreeswijk and Prakken, 2000; Dung et al., 2006).

the latter's development and construction. What matters for the evaluation in both cases seem to be but the inferential and logical relations between the propositions put forward. A method for evaluating complex debates which builds on applied logic is well-advised not to take certain rules governing rational discussions as fundamental.

Thirdly, formal argumentation frameworks have been developed in a fruitful collaboration of philosophy and artificial intelligence.[4] Some of these frameworks which describe the abstract structure of complex argumentation exhibit interesting similarities to the theory of dialectical strutures as I will point out in due course. Yet the framework developed by Dung (1995) deserves particular attention. Since Dung's application of his formal theory to logic programming, interpreting arguments as deductively valid premiss-conclusion structures, is almost identical with the concept of a dialectical structure (see below), I shall comment on Dung's work in some more detail.

Dung suggests to evaluate complex argumentations consisting of arguments which attack each other by introducing the concept of an admissible set of arguments, i.e. a set of arguments one can coherently and rationally claim. The key notion Dung's theory operates with is that of an argument's acceptability vis-à-vis a set of arguments: a is acceptable with regard to S if every argument that attacks a is itself attacked by an argument from S. A conflict-free[5] set S is admissible iff every argument $a \in S$ is acceptable with respect to S. Dung then defines specific admissible sets (with properties such as maximality with regard to set inclusion) which are supposed to identify those arguments a rational agent accepts given the complex argumentation, and in this sense represent alternative "semantics" of the argumentation system. It is, however, the very fundament of Dung's argumentation theory which may incite criticism. Dung writes: "[It] is reasonable to assume that a rational agent accepts an argument only if it is acceptable." (Dung, 1995, p. 326) But doesn't this put too heavy a burden on what is a rationally acceptable argument anyway? Assume an argument I put forward is attacked by a silly reasoning of one of my opponents. It is obvious to everybody that one premiss in the attacking argument is false. Is my position irrational as long as I haven't positively shown by an additional argument that the apparently false premiss is false? Indisputably, proponents have to be able to identify the false premiss in an argument that attacks one of their claims—otherwise their position were not coherent anymore—this, however, is not to say that they do have to put forward arguments against the premiss in question. On an earlier occassion, Dung formulates more carefully that "a statement is believable if it *can* be argued successfully against attacking arguments"[6] (Dung, 1995, p. 323). That is clearly much more plausible,

[4] See Chesñevar et al. (2000) for a recent review.

[5] No two arguments in that set attack each other.

[6] My emphasis.

but to say that it can be argued against an attack does not imply that it is argued against an attack. Tracing the disagreement with Dung's evaluation even back further, I would deny that argumentation evaluation should be based on the principle: "*The one who has the last word laughs best*" (Dung, 1995, p. 322). Consider complex historical philosophical argumentations: It was physically impossible for Hume to defend his position against Kant's attacks. So Kant laughs best and Hume was wrong, just because the latter died too early? And if I critcize Carnap and nobody argues against my criticism, I am right and Carnap wrong? Dung demonstrates impressive applications of his framework to n-persons games and the stable marriage problem, yet his theory is not a good theory of natural and, as a special case, philosophical argumentation—at least not as long as it is interpreted as a theory of real reasoning as opposed to ideal reasoning in which all possible arguments are explicit.[7] Still, Dung's notion of admissible sets and the fixed-point semantics in general can be directly related to parts of the theory of dialectical structures as I will show in section 5. The theory of dialectical structures not only departs from Dung's theory because sets of arguments that represent a rational position given the complex and controversial argumentation need not be admissible, moreover, they need not be conflict-free, either: Conditionalising status assignments to proponent positions, I will propose an interpretation such that not all sentences in the arguments we attribute to a proponent are actually claimed by the proponent. The reason for doing so will be given below.

Let me make a last preliminary remark: I argued above that evaluation procedures for complex argumentation should be based on those theories—namely formal and applied logic—that provide successful methods for single argument evaluation. However, the theory of dialectical structures and in particular the evaluation procedures for debates are, as we will see, underdetermined by formal and applied logic. New normative principles will have to be introduced into the theory. What is their status? On the one hand, they are reconstructions of those rules and principles of rational argumentation that are implicit in our socio-cognitive, communicative practices. Yet, the task is not simply to *discover* these implicit principles since discursive practice is by far not sufficiently precise and coherent to allow for such induction. Thus, on the other hand, these principles also represent a *specification* of our communicative practices. The enterprise of developing methods for the evaluation of complex argumentation inevitably contains *constructive elements.*

[7] Compare footnote 8.

3 The framework and a reformulation of the problem

The core of the theory of dialectical structures remains what I have already introduced in Betz (2005); I will only suggest minor modifications to that framework later in this paper. Here comes a concise summary. The subject matter of a theory of dialectical structures are debates. These consist of arguments which can be reconstructed as premiss-conclusion structures. Moreover, I will assume that arguments are reconstructed as deductively valid. The set of reconstructed arguments is labeled T.[8] An argument $a_1 \in T$ supports (attacks) an argument $a_2 \in T$ if and only if the conclusion of a_1, briefly: $C(a_1)$, is equivalent to (contradicts) a premiss of a_2. The support- and attack-relation, U and A respectively, that are thus defined on T make up the dialectical structure of the debate $\tau = \langle T, A, U \rangle$.[9]

A dialectical structure is a description of an argument's context, containing information about the fundamental functions an argument may fulfill, namely supporting or attacking other arguments.[10] Dialectical structures are models of—or: realise the more general structure of—what Cayrol and Lagasquie-Schiex (2005), extending Dung (1995), have called a "bipolar argumentation framework".[11] Dialectical structures, however, are not identical with bipolar argumentation frameworks: The latter are far more abstract structures, giving rise to many different interpretations of the notion of argument.[12]

[8]Note that, unlike in approaches by Lin and Shoham (1989) or the interpretation by Prakken and Vreeswijk (2001, p. 256) of Dung (1995), T is not supposed to contain arguments which can be *constructed* given the propositions put forward in a debate (or, more generally, some INPUT) but only those arguments that have been explicitly stated (though not necessarily fully). This emphasis on real reasoning as opposed to ideal reasoning seems to be more in line with the approaches of Pollock (1987, 1995), Vreeswijk (1997), or Verheij (1996).

[9]Accordingly, if two arguments conflict, i.e. possess contrary conclusions, they do not necessarily attack each other as defined here. The "assumption attack" as well as "undercutting" an argument (see Pollock, 1970; Prakken and Vreeswijk, 2001) can both be represented in this framework as an attack on an argument's premiss. Moreover, indirect attacks, i.e. attacks on an argument's subconclusion $c-$, can be made explicit be reconstructing the attacked argument as two arguments, a_1 and a_2, such that $c-$ is the conclusion of a_1 and a premiss of a_2, a_1 supporting a_2 and a_2 being the argument attacked.

[10]Birnbaum (1982) aims at representing these main functions, too, whereas he, however, analyses arguments as simple propositions, not as premiss-conclusion structures.

[11]Cayrol and Lagasquie-Schiex (2005) extend Dung's notion of acceptability and admissibility to these frameworks. Yet like Dung, they seem to put much too strong a constraint on rationally acceptable sets of arguments, too, as they require, for example, that if a attacks c and b supports c, I cannot rationally accept a and b in the same time. Arguments a and b, however, might address different premisses of c and it might thence be perfectly consistent to accept these arguments.

[12]Applying his theory to n-person games and the stable marriage problem, Dung (1995) interpretes arguments as payoff vectors and possible marriages, respectively. In the theory of dialectical structures, in contrast, arguments consist in premisses and conclusions whose

A two-coloured, directed graph is an appropriate mathematical model of a dialectical structure, and a helpful visualisation, too. Accordingly, arguments are the graph's nodes; a green (red) arrow between two arguments signifies that one supports (attacks) the other. I will use curly and straight arrows to express these relations conveniently, i.e. $A(a,b) \iff a \rightsquigarrow b$ and $U(a,b) \iff a \rightarrow b$.

Our problem of determining the truth status of a thesis in a complex debate is equivalent to determining the status of an argument in that debate provided that we define:

Definition 1 (τ-truth) *A conclusion of an argument is true according to a debate with dialectical structure $\tau = \langle T, A, U\rangle$ (in short: τ-true) iff the argument in the debate is (dialectically) valid.*[13]

What do we gain if we reduce the problem of determining the truth status of a thesis to that of determining the validity of arguments? This: The validity of an argument in a complex debate obviously depends but on the debate's dialectical structure. All other features of a debate, e.g. the order by which proponents have put forward the arguments, are hence irrelevant to that question. The notion of an argument's validity in a dialectical structure becomes thus the central concept of the theory of dialectical structures. The first task lying ahead is to find an appropriate explication of this very concept.

4 The recursive validity-function and its implications

This section sets up a validity-function ϑ that attributes to each argument in a dialectical structure the value 1 for being valid and 0 for being invalid. It does so by adopting an 'affirmative single-proponent-perspective', i.e. we imagine a single person who adheres to all arguments of a debate[14] and wonders: "What should I think about argument a and its conclusion?"

Whether an argument is valid depends on whether it is attacked or supported by other arguments and on whether these are valid or not. This

semantical relations induce the support- and attack-relationship between the arguments. Accordingly, it is the application of Dung's theory to logic programming which comes closest to the theory of dialectical structures as mentioned earlier. This corresponds to the less abstract approach in Bondarenko et al. (1997).

[13]Dialectical validity is a property of an argument as part of a dialectical structure, it must not be confused with deductive validity which is an internal structural property of an argument. Unless stated otherwise, "valid" and "validity" refer in the following to dialectical validity. The notion of dialectical validity plays a similar role as the concepts of an argument being "justified", "undefeated", "in force", "preferred" in the literature on deafeasible argumentation (see Prakken and Vreeswijk, 2001, p. 233).

[14]Whereas the notion of "adhering to an argument" will be made explicit in section 5.

observation motivates a recursive definition of ϑ.[15] I see, *prima facie*, three promising candidates.

- An argument $a \in T$ is valid$_1$ iff all its supporting arguments and none of its attacking arguments are valid.
- An argument $a \in T$ is valid$_2$ iff one of its supporting arguments (in case it has any at all) and none of its attacking arguments are valid.
- An argument $a \in T$ is valid$_3$ iff none of its attacking arguments are valid.

Which of these definitions is adequate? Alternatives 1 and 2 seem to be at odds with argumentative practice based on applied logic. Consider $\tau = \langle T, A, U \rangle$:

$$\begin{tikzcd} a_1 \arrow[r, squiggly] & a_2 \arrow[d] \\ & a_3 \end{tikzcd}$$

Under alternatives 1 and 2 (represented in the following table by the corresponding validity functions ϑ_1 and ϑ_2), a_3 wouldn't be valid:

	ϑ_1	ϑ_2	ϑ_3
a_1	1	1	1
a_2	0	0	0
a_3	0	0	1

However, it is fallacious to follow that because a supporting argument is unacceptable, i.e. successfully attacked, its supported argument is so, too. a_3 can be a convincing argument—because its premisses can be true no matter how bad the arguments that attempt to justify them. This said, only the minimal definition of validity, alternative 3, seems to me adequate.[16]

[15]This is what Prakken and Vreeswijk (2001, p. 233) have termed a "procedural" as opposed to "declarative" form of defining a status assignment. Moreover, it also the main idea underlying the evaluation of argument structures as dialectical trees in the MTDR framework (see Simari et al., 1994; Simari and Loui, 1992): Controversial argumentation is analysed as a tree of attacking arguments, a node in that tree is defined as undefeated iff all its attacking nodes ar defeated. The theory of dialectical structures departs from that approach with regard to the following aspects: (i) Dialectical structures are made up of attack- *and* support-relations. (ii) Dialectical structures are not necessarily trees, i.e. non-circular. (iii) Proponent's positions are explicitly captured in the theory of dialectical structures and many-proponent (>2) debates can be analysed.

[16]In the end, this is the simple recursive definition discussed in Prakken and Vreeswijk (2001, p. 235) as well as the recursive rule Simari et al. (1994) have chosen to evaluate dialectical trees, too (see also footnote 15). It should be noted, however, that this definition is not recursive in the sense that an argument's validity depends on the status of its substructures.

Definition 2 (Validity-function) *Let $\tau = \langle T, A, U \rangle$ be a dialectical structure. A function $v : T \to \{0, 1\}$ is called a validity-function on τ iff for all $a \in T$: $(v(a) = 0 \leftrightarrow \exists b \in T : b \rightsquigarrow a \wedge v(b) = 1))$.*

If the validity-function exists on τ and is unique, it is labeled "ϑ" and an argument $a \in T$ is called "τ-valid" iff $\vartheta(a) = 1$, "τ-invalid" otherwise.

It follows by induction over the maximum length of directed red paths on an argument $a \in T$ that ϑ is unique on dialectical structures without directed red circles—a precondition we will therefore assume until section 6.

This definition somehow implements the juridical principle of ,,in dubio pro reo", following the basic logical fact that a conclusion of an invalid argument can nevertheless be true. Does it declare arguments as valid too generously? As we evaluate debates from an affirmative single-proponent-perspective here, this doesn't seem to be the case. For if one person adheres to all arguments and some $a \in T$ is not successfully attacked by any argument, she is entitled to consider it valid, and its conclusion as true.

Next, we introduce three further important notions of the theory of dialectical structures. The first definition gives a name to all premisses whose truth value is not determined by the inferential relations in a dialectical structure.

Definition 3 (Free premiss) *Let $\tau = \langle T, A, U \rangle$ be given. A premiss p of an argument in τ is called "bound in τ" iff*

$$\exists a \in T : \Big[\vartheta(a) = 1 \wedge \Big((p \Leftrightarrow C(a)) \vee (p \Leftrightarrow \neg C(a))\Big)\Big].$$

If and only if a premiss is not bound in τ, it is "free in τ". The set of all free premisses of τ is called Π_τ.

The second definition seizes the idea that a sentence c can be deduced from sentences $p_1 \dots p_n$ by using but the inferential relations encoded in the dialectical structure τ.

Definition 4 (τ-deducibility) *Let $\tau = \langle T, A, U \rangle$ be given. A statement c is deducible in τ from $S = \{p_1 \dots p_n\}$, briefly "$S \vdash_\tau c$", iff there is an argument $a \in T$ with $c \Leftrightarrow C(a)$ and there is a green subgraph $\tau' \subseteq \tau$ such that (i) a is the only sink of τ', and (ii) $\Pi_{\tau'} \subseteq S$.*

For the sake of generality, S may include any sentences whatsoever.

The third definition transposes the idea of consistency to dialectical structures.

Definition 5 (Equilibrium) *A dialectical structure $\tau = \langle T, A, U \rangle$ is said to be in equilibrium iff not*

$$(p \in \Pi_\tau \vee \Pi_\tau \vdash_\tau p) \wedge (\neg p \in \Pi_\tau \vee \Pi_\tau \vdash_\tau \neg p)$$

for some sentence p.

It might be worth noting that a dialectical structure τ can be in equilibrium although Π_τ is inconsistent. For that a contradiction can be derived from the free premisses doesn't entail that the required inferential relations are represented in τ and that the contradiction is thus τ-deducible. But if Π_τ is inconsistent, τ can be enlarged by further arguments to τ^* such that τ^* is not in equilibrium and $\Pi_\tau = \Pi_{\tau^*}$.

With these conceptual tools at hand, we can now proof our first theorem.

Proposition 1 (Validity and deducibility) *Let $\tau = \langle T, A, U\rangle$ be a dialectical structure in equilibrium without directed circles, and let c be a sentence. The following two statements are equivalent.*

1. *There is a τ-valid $a \in T$ with $C(a) \Leftrightarrow c$.*
2. $\Pi_\tau \vdash_\tau c$.

Proof: (1) $\Rightarrow$ (2): Induction over the maximum length n of directed paths in τ pointing on a (possible because τ is non-circular).

Basis (n=0): The argument a is neither attacked nor supported by any other arguments, thus all its premisses are free and $\Pi_\tau \vdash_\tau C(a)$.

Induction hypothesis: For all arguments a with paths in τ pointing on a no longer than n, it is true that if a is τ-valid then $\Pi_\tau \vdash_\tau C(a)$.

Inductive step: Let $a \in T$ with paths on a no longer than $n+1$ be τ-valid. Thus, any $b \in T$ with $b \rightsquigarrow a$ is τ-invalid. So, all bound premisses of a are actually conclusions of τ-valid arguments, say $a_1 \dots a_m$. But no path on any $a_i, 1 \le i \le m$ is longer than n such that the induction hypothesis applies. All bound premisses of a are therefore τ-deducible from Π_τ, and so is $C(a)$.

(2) $\Rightarrow$ (1): First, we define the depth of a non-circular dialectical structure $d(\tau)$ as the length of the longest directed path in τ. The proof goes by induction over the maximum depths n of subgraphs $\tau' \subseteq \tau$ pointing on c, i.e. τ' contains an argument as sole sink whose conclusion is equivalent to c. (This induction is possible because τ is non-circular.)

Basis (n=0): If c is τ-deducible from Π_τ, there are $a_i, i = 1 \dots m$ with $C(a_i) \Leftrightarrow c$. None of these is supported by an argument. As $\Pi_\tau \vdash_\tau c$, all the premisses of at least on $a_j, 1 \le j \le m$ have to be free. This a_j is τ-valid.

Induction hypothesis: For any sentence c such that subgraphs which contain an argument $a \in T$ with $C(a) \Leftrightarrow c$ as sole sink are of maximum depths n, $\Pi_\tau \vdash_\tau c$ implies that there is a τ-valid $a \in T$ with $C(a) \Leftrightarrow c$.

Inductive step: Let c be a sentence such that subgraphs which contain an argument $a \in T$ with $C(a) \Leftrightarrow c$ as sole sink are of maximum depths $n+1$. Assume further $\Pi_\tau \vdash_\tau c$. Thus, there is an $a \in T$ with $c \Leftrightarrow C(a)$ and there is a green subgraph $\tau' \subseteq \tau$ such that (i) a is the only sink of τ' and (ii) $\Pi_{\tau'} \subseteq \Pi_\tau$. Now this a is valid! Otherwise, there would be a τ-valid $a' \in T$ with $a' \rightsquigarrow a$. Set $p := \neg C(a')$. Then, $p \notin \Pi_\tau$ and hence $p \notin \Pi'_\tau$.

Consequently, there is at least one τ-valid $a^* \in T' \subset T$ such that $C(a^*) \Leftrightarrow p$. As subgraphs that contain a^* or a' as sole think are of maximum depths n, the induction hypothesis applies and $\Pi_\tau \vdash_\tau p$ as well as $\Pi_\tau \vdash_\tau \neg p$. Contrary to our assumption, τ is not in equilibrium. $\square$

This is an important result. It represents a justification *a posteriori* for our initial strategic decision to focus on the validity of arguments instead of the truth value of sentences when evaluating debates, though these sentences are what one is ultimately interested in. To see whether a sentence c is "inferred in a debate", i.e. τ-deducible from its free premisses, it is sufficient to spot for τ-valid arguments warranting c. As the latter is determined by the debate's dialectical structure, it is not necessary to zoom into single arguments. Yet, we should bare in mind that the concept of being τ-deducible from the debate's free premisses, in contrast to the pre-theoretic notion of being inferred in a debate, is built upon the definition of τ-validity. In other words: the definition of τ-validity not only enters statement (1) but also statement (2) of proposition 1.

The second theorem is rather of practical relevance. It identifies quickly verifiable necessary and sufficient conditions for being a τ-valid argument. Let us define a long red path on $a \in T$ as a directed red path f with a as sink such that there is no other directed red path f' with a as sink which includes f. The theorem states:

Proposition 2 (Long red paths and validity) *Let $\tau = \langle T, A, U \rangle$ be a dialectical structure in equilibrium without directed red circles. Then, for every $a \in T$ which is attacked at all, the following two statements hold:*

1. *If every long red path f on a is of odd length, then $\vartheta(a) = 0$.*
2. *If every long red path f on a is of even length, then $\vartheta(a) = 1$.*

Proof: Induction over the maximum length of long red paths on a (is possible because directed red circles are excluded).

Basis (n=2): Statement (1). If every long red path f on a is of odd length, then a is attacked by arguments which are not attacked themselves and are thus τ-valid. Consequently, $\vartheta(a) = 0$. Statement (2). If every long red path f on a is of even length, then it is of length 2. Hence, every $b \in T$ with $b \rightsquigarrow a$ is itself attacked by some unattacked argument. So every such b is τ-invalid and a is τ-valid.

Induction hypothesis: For all $a \in T$ such that all long paths f on a are no longer than $n \geq 2$, statements (1) and (2) are true.

Inductive step: Assume $a \in T$ with long red paths f on a of maximum length $n + 1$.

Statement (1). Let's assume that every long red path f on a is of odd length. Consider all arguments $a_i, 1 \leq i \leq m$ that attack a, $a_i \rightsquigarrow a$. It follows for all $i \in \{1 \dots m\}$: every long red path on a_i is of even length

$l_i \leq n$ (otherwise there would be a long red path on a with even length). The induction hypothesis applies (namely statement 2) and we infer that all a_i are τ-valid. Thus, a is not τ-valid.

Statement (2). This follows analogously. Let's assume that every long red path f on a is of even length. Consider again all arguments $a_i, 1 \leq i \leq m$ that attack a, $a_i \rightsquigarrow a$. It follows for all $i \in \{1 \ldots m\}$: every long red path on a_i is of odd length $l_i \leq n$ (otherwise there would be a long red path on a with odd length). The induction hypothesis applies (namely statement 1) and we infer that no a_i is τ-valid. Thus, a is τ-valid. □

It is time for a preliminary resumé. Our theory has so far provided means that enable a person who adheres to an entire debate, i.e. accepts all its free premisses, to determine whether she should think that a certain thesis is true or not. Now, we rarely approve all arguments and all free premisses in a debate. Opponents characteristically accept only some arguments in course of a controversial discussion. We will in the next section extend our theory so that it can be applied to such multi-proponent debates.

5 Application to multi-proponent debates

The general idea is the following: We break up the dialectical structure into (possibly overlapping) substructures such that each proponent o_i, $i = 1 \ldots k$, accepts the arguments in her part only. The concepts and propositions developed above can then be applied to these subdebates in order to determine what a proponent should think of a certain thesis, for example.

Accordingly, our next step consists in specifying how to pick the proponent-specific substructures τ_i, $i = 1 \ldots k$. In Betz (2005), I introduced a stance-attribution that assigns a set of proponents to each argument and interpreted it in the following way: a proponent accepts all premisses and the conclusion of an argument she is assigned to via the stance-attribution. Here, I want to put forward a modified interpretation of stance-attribution and subsequently explain why I consider it as more appropriate.

Definition 6 (Stance-attribution) *Let $\tau = \langle T, A, U \rangle$, and $O = \{o_1, \ldots, o_k\}$ be a set of proponents. A function $S : O \rightarrow \mathcal{P}(T)$ is called a stance-attribution on τ. $\tau_i = \langle S(o_i), A|_{S(o_i)}, U|_{S(o_i)} \rangle$ is the subdebate accepted by o_i. A proponent o_i claims that*

- *All $p \in \Pi_{\tau_i}$ are true.*
- *All $C(a)$ (with $a \in S(o_i)$ is τ_i-valid) are true.*

According to this new definition of stance-attribution, a proponent will adhere to more arguments than under the old one. However, inconsistencies among premisses of arguments a proponent o_i approves don't imply anymore that o_i has contradictory beliefs since she only subscribes to the τ_i-free premisses and conclusions of τ_i-valid arguments.

The general reason why I favour this modified stance-attribution over the original one is that it is more informative. Before I can explain that, we have to introduce a general criterion for stance-attributions. Regardless of this, the need for a criterion stems from the fact that we wouldn't accept arbitrary stance-attributions on τ. A proponent o_i should, for instance, approve an argument whose premisses are equivalent to conclusions of already accepted τ_i-valid arguments. Proponents are thus not entirely free to choose which arguments to adhere to. The following is therefore a necessary criterion for 'permissible', 'acceptable', or 'rational' stance-attributions.

Definition 7 (Closed subdebates) *Let $\tau = \langle T, A, U \rangle$ be a dialectical structure and $S : O \to \mathcal{P}(T)$ a stance-attribution on τ. A subdebate τ_i induced by S is called "closed" iff there is no $a \in (T \setminus T_i)$ such that $\Pi_{\tau_i} = \Pi_{\tau'}$, $\tau' = \langle S(o_i) \cup \{a\}, A|_{S(o_i)\cup\{a\}}, U|_{S(o_i)\cup\{a\}} \rangle$.*

In other words, if we can't add an argument to a subdebate without increasing the set of free premisses, then this subdebate is closed.

A stance-attribution is permissible only if all induced subdebates are closed. For if this criterion is not satisfied, a proponent doesn't accept an argument she should accept given her other commitments.

Now, the modified interpretation of stance attribution is in the following sense more informative: There are cases where the stance-attribution according to the old definition doesn't indicate that a proponent is obliged to accept a certain argument, whereas the one according to the modified definition does. Consider as an example the following dialectical structure τ:

$$a_1: \boxed{\begin{array}{c} p_1 \\ p_2 \\ \hline \neg p_3 \end{array}} \rightsquigarrow a_2: \boxed{\begin{array}{c} p_3 \\ p_4 \\ \hline \neg p_5 \end{array}} \rightsquigarrow a_3: \boxed{\begin{array}{c} p_5 \\ p_6 \\ \hline \neg p_7 \end{array}} \rightsquigarrow a_4: \boxed{\begin{array}{c} p_7 \\ p_8 \\ \hline \neg p_9 \end{array}}$$

$$a_5: \boxed{\begin{array}{c} p_4 \\ p_8 \\ \hline c \end{array}}$$

and suppose that the proponent o beliefs in the following sentences: p_1, p_2, $\neg p_3, p_4$, $p_5, p_6, \neg p_7, p_8$. According to the old interpretation of stance-attribution, proponent o approves all arguments whose premisses and conlusion are included in her set of beliefs. Thence $S_{old}(o) = \{a_1, a_3\}$, and the corresponding subdebate τ_o is closed since adding arguments to it will inevitably increase the set of free premisses. Let us now consider which arguments o adheres to according to the modified interpretation of stance-attribution in definition 6. Here, $S_{new}(o) = \{a_1, a_2, a_3, a_4\}$ since o claims the free premisses and the conclusions of τ_o-valid arguments only. In contrast to the previous case, the corresponding subdebate τ_o is not closed! But adding a_5 to $S_{new}(o)$ will close it. Now why does this show that the new interpretation is more appropriate than the old one? We introduced the property of being a closed subdebate in order to express whether a proponent meets her discursive obligations by approving all arguments she has to accept in the light of the arguments already approved. In the specific example above, o is actually—given her set of beliefs—obliged to accept a_5, and in particular its conclusion c. As the stance-attribution in line with the old interpretation was closed, it didn't contain this information, whereas the new one—not being closed—does.

All this said and done, we start realising that the evaluation of debates comprises several tasks. Let p be a controversial thesis in a multi-proponent debate of structure $\tau = \langle T, A, U \rangle$ and with proponents $O = \{o_1, \ldots, o_k\}$. Its evaluation consists in addressing the following questions:

- For every τ_i $(i = 1 \ldots k)$: Is τ_i closed? If not, proponent o_i doesn't meet her discursive obligations.
- For every τ_i $(i = 1 \ldots k)$: Is τ_i in equilibrium? If not, proponent o_i makes inconsistent claims.
- For every τ_i $(i = 1 \ldots k)$: Is p true in τ_i, i.e. is there a τ_i-valid argument a such that $C(a) \Leftrightarrow p$? If so, p is true for o_i.

This paves the way for analysing burdens of proof. I suggest to distinguish two types of burden of proof.

Definition 8 (Burden of proof) *Let $\tau = \langle T, A, U \rangle$ be given with opponents O and stance-attribution S. A proponent $o_i \in O$ carries a*

(i) first-order burden of proof regarding the thesis p iff p is not τ_i-deducible from Π_i.

(ii) second-order burden of proof regarding the thesis p iff there is $j \neq i$ such that p is not τ_j-deducible from Π_j.

If a proponent meets her first-order burden of proof with respect to p at least for her, i.e. on the background of the arguments she accepts, p is true. It is an entirely different question as to whether her opponents think so, too. Accordingly, she carries a second-order burden of proof with respect to p iff at least one opponent isn't convinced that p. The distinction of the two types of burden of proof therefore corresponds to the distinction between justifying a thesis for oneself on the one hand and justifying a thesis in order to convince one's opponents on the other hand.

(Let us step back for a minute and note that we have here an example for how concepts and procedures of everyday argumentation are embedded in the theory of dialectical structures. But not only are these common concepts such as the burden of proof reformulated within that theory, they are moreover and more importantly specified and stated in a more precise way such that the theory of dialectical structures refines the rules of rational argumentation.)

The above criteria subdebates have to satisfy not only enable us to evaluate multi-proponent debates but also suggest a typology of rational discursive strategies. Broadly, we can distinguish two types of discursive moves: An argumentative strategy by proponent o_i is (i) defensive iff it aims at τ_i, e.g. is supposed to ensure that τ_i meets an evaluation criterion such as being in equilibrium; (ii) offensive iff it aims at some $\tau_j, j \neq i$, for instance with the objective of justifying a thesis in τ_j or pushing τ_j out of equilibrium.

A specific subtype of offensive strategies is what is widely known as "internal critique" (e.g. Schleichert, 1998). An argumentative strategy by proponent o_i is an internal critique iff she not only argues with a view to some τ_j, $j \neq i$, but also introduces arguments in the debate that she doesn't approve herself, i.e. that are not part of τ_i, but are supposed to be introduced into τ_j, $j \neq i$.

Moreover, it should be noted that the theory developed in this paper, while being based on a monotonic logic which serves for the internal evaluation of arguments, represents a framework for non-monotonic reasoning. The introduction of a new argument might alter the status of other arguments, formerly τ-valid arguments might become invalid and propositions which have been τ-true might not be so anymore once the new argument is fully taken into account.

Before considering the more general case of circular dialectical structures, I shall briefly show how Dung's notion of admissible sets of arguments can be embedded within this theory. Dung defined a complete extensions as an admissible set S of arguments such that all arguments that are acceptable with respect to S belong to S (Dung, 1995). Dung showed that every argumentation framework that is "well-founded" (which signifies that there is no infinite sequence of pairwise attacking arguments) has exactly one complete extension. The following theorem embeds these results within the theory of dialectical structures.

Proposition 3 (τ-valid arguments as complete extension) *Let $\tau_i = \langle T_i, A_i, U_i \rangle$ be a subdebate induced by a stance attribution on the acyclic dialectical structure $\tau = \langle T, A, U \rangle$. Then $\mathrm{AF} = \langle T_i, A_i \rangle$ is a well-founded argumentation framework in the sense of Dung (1995) and the set Θ of all τ_i-valid arguments is the complete extension of* AF*.*

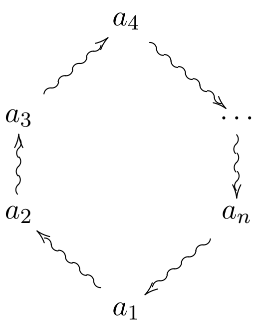
Proof: As τ is finite and non-circular, AF is well-founded. We have to show that the set $\Theta := \{a \in T_i : \vartheta_i(a) = 1\}$ (i) is conflict-free, (ii) is admissible, and (iii) contains all arguments that are acceptable with respect to Θ. Ad (i): By the recursive definition of the validity function, a valid argument cannot attack another valid argument. Ad (ii): Consider an arbitrary τ_i-valid argument a. If a is not attacked in τ_i, a is acceptable with regard to Θ. If a is attacked by some $b \in \tau_i$, then b is τ_i-invalid and that implies that there is some other τ_i-valid argument c ($c \in \Theta$) which attacks b. So a is acceptable with respect to Θ. Since a was chosen arbitrarily, Θ is admissible. Ad (iii): Let $a \in T_i$ be an argument which is acceptable with regard to Θ, that is every argument which attacks a is itself attacked by a τ_i-valid argument. But then a is itself τ_i-valid and belongs to Θ. $\square$

6 Circular structures

We now relax our assumption that the dialectical structure is non-circular. The first questions we have to answer are: Does a validity-function as defined recursively in definition 2 exist, and is it unique? We will reaffirm that this is not necessarily so. In the literature, two types of solution have been proposed to address this problem, namely (i) to modify the recursive definition of an arguments's status (validity) and (ii) to embrace the existence of multiple status assigments (validity functions) as a positive feature (see Prakken and Vreeswijk, 2001, p. 236). I will depart from these solution insofar as I will keep the simple and very plausible recursive definition of validity and show that qua reconstruction of a debate as a dialectical structure one can ensure the existence and uniqueness of a validity function. This section primarily investigates necessary and sufficient conditions for the existence and uniqueness of validity functions whereas the last section elaborates the eventual solution.

As sketched in section 4, the validity-function ϑ exists and is unique on structures $\tau = \langle T, A, U \rangle$ which don't contain red directed circles as subgraphs. Thus, let us now consider dialectical structures τ that do contain red directed circles $C = \{a_1 \dots a_n\} \subseteq T$ with $a_1 \rightsquigarrow a_2 \rightsquigarrow \dots \rightsquigarrow a_n \rightsquigarrow a_1$ while distinguishing red circles of even and of odd length. Note also that green edges are irrelevant for determining the validity-function of a dialectical structure —; this is the reason why we consider but the red ones in the following.

First, we shall consider the question of uniqueness and have a closer look at *even* red circles. Let n be an even number and $\tau =$



a dialectical structure. Then, there are two functions $v_1, v_2 : T \to \{0,1\}$ that satisfy the recursive definition of validity, namely

$$v_1(a_i) = i \bmod 2$$

and

$$v_2(a_i) = (i+1) \bmod 2.$$

Thence, in this case, validity is underdetermined by the dialectical structure. The following theorem, assuming that there are functions on τ that satisfy the recursive validity definition, states a sufficient and necessary condition for ϑ being unique.

Proposition 4 (Uniqueness of validity-functions) *Let $\tau = \langle T, A, U\rangle$ be a dialectical structure, $C_1, \dots, C_m \subseteq T$ enumerate all the red directed circles in τ, and $v_1, \dots, v_l$ be the functions that satisfy the recursive definition of validity on τ. Then, the following two statements are equivalent:*

1. *ϑ is unique on τ.*
2. *For every C_j $(1 \leq j \leq m)$ there is an argument $a_j \in C_j$ such that for every v_i $(1 \leq i \leq l)$: $v_i(a_j) = 0$.*

Proof: (1) $\Rightarrow$ (2). Clear.

(2) $\Rightarrow$ (1). The proof consists in two steps. First, we transform τ to τ' such that τ' doesn't contain any red directed circles. Then, we show that every validity-function v_i on τ is also a validity-function on τ'. And since ϑ is unique on structures without red directed circles, it follows that it is unique on τ, too.

Step (i). Let τ be given as specified above. We set $T^* := \{a \in T | \exists j : (a \in C_j \wedge \forall i v_i(a) = 0)\}$. By assumption (2), at least one argument of each circle is in T^*. We transform τ by deleting all outgoing red arrows from all arguments in T^*, formally: $\tau' = \langle T, A \setminus \{(a,b) | a \in T^* \wedge b \in T\}, U\rangle$. As every red directed circle in τ "loses" at least one red arrow, τ' doesn't contain red circles.

Step (ii). It remains to be shown that every validity-function v_i on τ is also a validity-function on τ'. More precisely, we have to proof that $\forall a \in T : (v_i(a) = 0 \leftrightarrow \exists b \in T : b \rightsquigarrow_{\tau'} a \wedge v_i(b) = 1))$ for all $i = 1 \ldots l$. Consider arbitrary $a \in T$ and v_i. "$\Rightarrow$": Assume $v_i(a) = 0$. Consequently, there is a b that attacks a in τ with $v_i(b) = 1$. But then, $b \notin T^*$ and therefore $b \rightsquigarrow_{\tau'} a$. "$\Leftarrow$": If there is a $b \in T$ such that $b \rightsquigarrow_{\tau'} a \wedge v_i(b) = 1$, then $b \rightsquigarrow_{\tau} a$ and as v_i is a validity-function on τ, $v_i(a) = 0$. $\square$

Let's consider two examples. $\tau_1 =$

$$\begin{tikzcd}
a_2 \arrow[r, rightsquigarrow] & a_3 \arrow[d, rightsquigarrow] & a_5 \arrow[l, rightsquigarrow] \\
a_1 \arrow[u, rightsquigarrow] & a_4 \arrow[l, rightsquigarrow] &
\end{tikzcd}$$

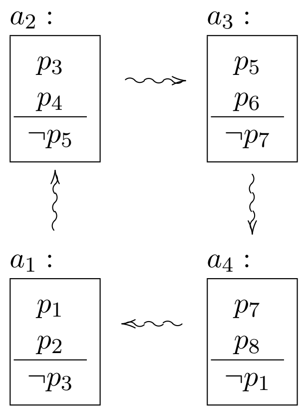
Here, $v_i(a_5) = 1$ for every validity-function v_i and so $v_i(a_3) = 0$; the theorem applies. $\tau_2 =$

$$\begin{tikzcd}
& a_3 \arrow[dr, rightsquigarrow] & \\
a_2 \arrow[ur, rightsquigarrow] & & a_4 \arrow[dl, rightsquigarrow] \\
& a_1 \arrow[ul, rightsquigarrow] \arrow[uu, rightsquigarrow] &
\end{tikzcd}$$

Here, too, a_3 is successfully attacked under every v_i. But why? Well, either $v_i(a_1) = 0$ or $v_i(a_1) = 1$. In the first case, $v_i(a_2) = 1$ and thus $v_i(a_3) = 0$; in the second case, a_3 is successfully attacked by a_1 and so $v_i(a_3) = 0$. This example shows that even in a dialectical structure where all arguments belong to some circle, ϑ can be given uniquely. Besides, it yields a useful generalisation: Whenever there are an even and an odd red directed path on some argument a (i) which originate from the same argument and (ii) whose arguments are not attacked by arguments outside the respective path, a is invalid according to every validity-function on τ.

This said, what does it signify that a validity-function is underdetermined on some subdebate τ_i asserted by proponent o? It signifies that we cannot infer from the stance-attribution which sentences o actually claims.

To see this, consider the following example, $\tau_i =$



There are two validity-functions v_1, v_2 which satisfy the recursive definition of validity. The following table shows which sentences o asserts according to our interpretation of stance-attributions.

val. funct.	valid arguments	sentences claimed by o
v_1	a_1, a_3	$p_1, p_2, \neg p_3, p_4, p_5, p_6, \neg p_7, p_8$
v_2	a_2, a_4	$\neg p_1, p_2, p_3, p_4, \neg p_5, p_6, p_7, p_8$

So, in case of underdetermined validity-functions, o is free to choose between the two sets of sentences and, correspondingly, the two validity-functions.

We shall now turn our attention to the existence of validity-functions. Let τ be a red circle of *odd* length, i.e. a vicious circles (see Betz, 2005). In contrast to the case of even circles, ϑ is not underdetermined on τ; for even worse, it doesn't exist! It follows easily (by indirect proof), that for every function $g : T \rightarrow \{0, 1\}$ there are $a_i \rightsquigarrow a_{i+1}$ (index summed modulo n) with $g(a_i) = g(a_{i+1})$, but this contradicts the recursive definition of validity. So, the vicious circles turn out to be vicious in this context, too. Theorem 5 below states that the absence of vicious circles, i.e. red directed circles of odd length, is a sufficient condition for a validity-function to exist. To proof it, we need

Definition 9 (Circuit) *Let $\tau = \langle T, A, U\rangle$ be a dialectical structure. $D \subseteq T$ is called a circuit iff for all $a, b \in D$ there is a circle C such that $a \in C$ and $b \in C$. D is an "even circuit" iff all circles in D are of even length.*

and

Lemma 1 (Validity function on even circuits) *Let $\tau = \langle T, A, U\rangle$ be a dialectical structure such that all its red circles are part of one even red circuit. Then, there is a validity-function on τ.*

Proof: By induction over the number of red edges in τ.

Basis (n=3): Dialectical structures satisfying the above conditions with 3 red arrows are either non-circular, and then, the validity-function exists, or of one of the following forms: $[a_1 \rightsquigarrow a_2 \rightsquigarrow a_1, a_1 \rightsquigarrow a_3]$, $[a_1 \rightsquigarrow a_2 \rightsquigarrow a_1, a_3 \rightsquigarrow a_1]$. In both cases, a validity-function exists.

Induction hypothesis: Let $\tau = \langle T, A, U \rangle$ be a dialectical structure such that all its red circles are part of one even circuit and the number of red edges is not larger than n. Then, there is a validity-function on τ.

Inductive step: Let $\tau = \langle T, A, U \rangle$ be a dialectical structure such that all its circles are part of one even red circuit and the number of red edges is $n+1$. τ's arguments fall into three categories: $T_1 = \{a \in T | \exists C : C \,\text{is circle} \wedge a \in C\}$, $T_2 = \{a \in T \setminus T_1 | \exists f, b : (f \,\text{is path from}\, a \,\text{to}\, b) \wedge b \in T_1\}$, $T_3 = \{a \in T \setminus T_1 | \nexists f \exists b : (f \,\text{is path from}\, a \,\text{to}\, b) \wedge b \in T_1\}$. Either $T_2 = \emptyset$ (i) or not (ii). Case (i): Let a^* be an arbitrary argument in T_1. We define $v : T \rightarrow \{0,1\}$ on T_1 as $v(a) = \mathit{distance}(a, a^*) \bmod 2$. As all circles in τ are even, v satisfies the recursive definition of validity so far. Finally, v can be constructed on T_3 iteratively. Case (ii): $v : T \rightarrow \{0,1\}$ is constructed on T_2 iteratively. Consider all arguments $a_1 \ldots a_m \in T_1$ such that $\exists b \in T_2 : b \rightsquigarrow a \wedge v(b) = 1$. Case (ii.1): There are such arguments: Transform τ to τ' be deleting all outgoing red arrows from these arguments a_i. By induction hypothesis, there is a validity-function v' on τ'. As $\forall i : v'(a_i) = 0$, v' is a validity-function on τ. Case (ii.2): There are no such arguments. Then proceed as in case (i). □

These preliminaries will now enable us to proof the following theorem.

Proposition 5 (Validity function on non-vicious structures) *Let $\tau = \langle T, A, U \rangle$ be a dialectical structure. If τ doesn't contain vicious circles (red circles of odd length), there is a validity-function on τ.*

Proof: If τ is non-circular, there is nothing to proof. If it is circular, consider the subsets $D_i \subseteq T$ such that $a \in D_i \wedge b \in D_i \iff \exists C : (C \,\text{is circle} \wedge a, b \in C)$. These D_i are disjunct even red circuits which are possibly interconnected by red paths. Yet, these interconnection are in any case non-circular. But then, Lemma 1 tells us that they can be evaluated recursively, allowing for an iterative construction of v on τ. □

τ_2 from above represents an example of a dialectical structure that comprises a vicious circle but on which the validity-function nevertheless exists, and is unique.

In general, a dialectical structure can possess none, one, or many validity functions. The following algorithm represents a mechanical procedure for evaluating dialectical structures:

```
<tau> = <T,A,U>;
set v(a)=NaN for all a in T;
spot for pairs of even-odd paths on a and set v(a)=0; //optional
```

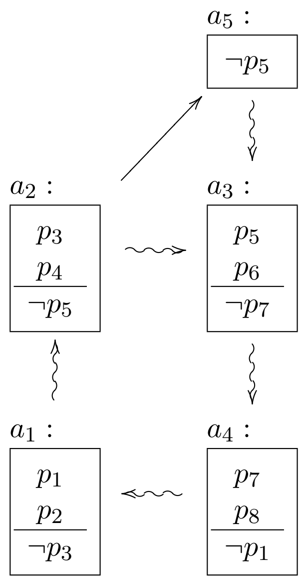
```
04 constructValidityFunction(v);
05
06 method constructValidityFunction(v) {
07   if (v(a)!=NaN for all a) do: {
08     if (v is validity-function) do: output v;
09     return;
10   }
11   T*  = set of all args a with v(a)=NaN and
12         a is attacked by an arg b with v(b)=1;
13   T** = set of all args a with v(a)=NaN and
14         for all args b that attack a v(b)=0;
15   if (T* and T** empty) do: {
16     take some a in T with v(a) = NaN;
17     v(a) = 0;
18     constructValidityFunction(v);
19     v(a) = 1;
20     constructValidityFunction(v);
21   } else {
22     for all a in T* do: set v(a) = 0;
23     for all a in T** do: set v(a) = 1;
24     constructValidityFunction(v);
25   }
26 }
```

This algorithm generates all validity-functions on a dialectical structure τ because it constructs a validity function v iteratively only to the extend that v is determined by the dialectical structure (lines 21-25) and if, at some point, further values are underdetermined, the algorithm makes an exhaustive distinction of cases: it sets an arbitrary argument's value tentatively to 0 respectively 1 and tries to construct validity functions for each of these cases (lines 15-21).

7 Debate reconstruction and theses introduction

This final section attempts to tackle the problems that emerged in the previous section by showing that one can ensure *qua reconstruction* of a natural language argumentation as a dialectical structure that the validity function exists and is unique. In order to do so, a specific type of argument, namely theses, will have to be introduced.

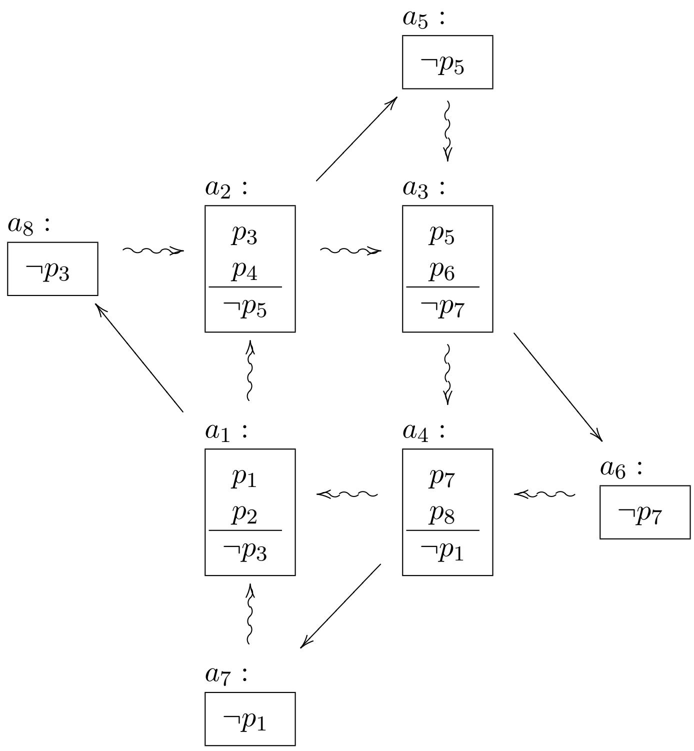
Reconsider the illustrative subdebate on page 19 with its two validity functions. The situation (i.e. the assignment of some proponent to that subdebate) can be interpreted as follows: τ_i is ambiguous and does not precisely determine which sentences the proponent o asserts. This ambiguity, however, can be overcome by introducing theses (i.e. arguments that contain their conclusion as sole premiss and that do not represent further inferential constraints) into the dialectical structure which o additionally asserts.[17] Thus, if o's sentence evaluation actually corresponds to validity function v_2, we might make this explicit in the dialectical structure by introducing a thesis which maintains $\neg p_5$. τ_o':



The validity function (v_1) exists uniquely on τ_o'. Generally, by requiring that at least one conclusion in the red circle is asserted as thesis, the ambiguity of the dialectical structure is avoided. It is, besides, not inconsistent to assert

[17]Only the conclusion of theses will be displayed in the graphs below.

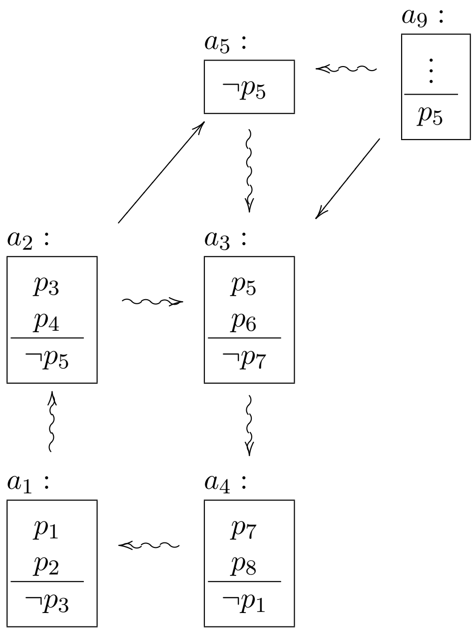
the conclusions of all arguments as theses:



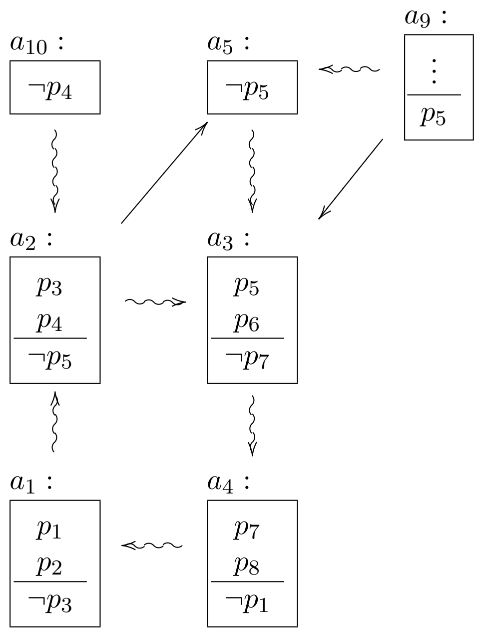
In this extreme case, none of the arguments $a_1, \ldots, a_4$ is τ-valid.

The underlying reason why introducing theses fixes a dialectical structure's ambiguity is given by theorem 4 on the uniqueness of validity functions above. Introducing theses which assert conclusions of arguments in the red circle and which, consequently, attack arguments in the red circle, fixes the validity-value of these attacked arguments, in other words: renders condition (2) of theorem 4 true.

Still, there is a problem lurking around the corner. What if the newly introduced thesis is itself attacked by a τ-valid argument? In that case, the thesis doesn't fix the validity-value of any argument for not being τ-valid itself. The situation is this, τ''_o



Clearly, a_5 is τ''_o-invalid and thence its attacking a_3 doesn't make a_3 τ''_o-invalid. Yet, as a_9 is τ''_o-valid, a_2 shouldn't be so since otherwise the debate were not in equilibrium! If, however, we simply set, in line with this reasoning, $\vartheta(a_2) = 0$, the thus set validity function doesn't necessarily satisfy the recursive definition of validity: a_2 is τ-invalid although it might not be attacked by a τ-valid argument. This problem can be overcome by introducing a further thesis which asserts the negation of one of a_2's premisses—and the above reasoning warrants that there has to be such a thesis the proponent approves of. In the following example, that is thesis a_{10}.



It therefore seems to be safe to assume that when *reconstructing* a proponent's subdebate, it is always possible to add for every red circle a thesis which the proponent asserts (i.e. is τ_i-valid) and which attacks one of the circle's arguments. Solving the first aspect of the red-circle problem, the following theorem summarises this section's reasoning.

Proposition 6 (Theses-introduction and uniqueness) *Let $\tau = \langle T, A, U \rangle$ be a dialectical structure without vicious circles and in equilibrium such that there is at least one function $v : T \rightarrow \{0, 1\}$ satisfying the recursive definition of τ-validity. If for every red circle $C = \{a_1, \ldots, a_n\}$ in τ there is a thesis $a_C \in T$ that asserts a conclusion of one of C's arguments, then the validity function is unique, or it is possible to add further theses, transforming τ to τ', such that τ' is in equilibrium and ϑ is unique on τ'.*

Proof: The theorem follows with proposition 4 and a case distinction for each thesis a_C according to whether a_C is τ-valid or not. If it is τ-valid, then the argument in C it attacks, say a_k, is not. If it is not τ-valid, then the argument a_l in C which attacks a_k is τ-invalid by virtue of τ being in equilibrium. The latter must be made explicit by introducing a thesis which attacks a_l without being successfully attacked itself. $\square$

We shall now see how the strategy of fixing validity-values in a red circle by introducing theses can also be applied in order to tackle the second aspect of the red-circle problem, i.e. the non-existence of a validity function on some dialectical structures. In order to do so, we have to improve theorems 4 and

6 above which assume that there is at least one function which satisfies the recursive definition of τ-validity—and hence don't cover the question whether a validity function exists at all.

Proposition 7 (Theses-introduction end existence) *Let* $\tau = \langle T, A, U \rangle$ *be a dialectical structure in equilibrium. If for every red circle* $C = \{a_1, \ldots, a_n\}$ *in* τ *there is a thesis* $a_C \in T$ *which*

(i) attacks one of C*'s arguments,*

(ii) is neither part of a red circle itself nor connected to a red circle via a red directed path from that circle to a_C*, and*

(iii) is assigned the validity value 1 according to a partial evaluation of τ*,* $\vartheta_{\text{partial}}$*, which excludes all the arguments in red circles,*

then the validity function ϑ *on* τ *exists, and is unique.*

Proof: The proof applies basically the same ideas as the proof of proposition 4 above. We specify, first, how to transform the dialectical structure's characteristic graph G_τ to G'_τ. Secondly, we show that ϑ exists and is unique on G'_τ. And, thirdly, we prove that every validity function on G'_τ is a validity function on G_τ.

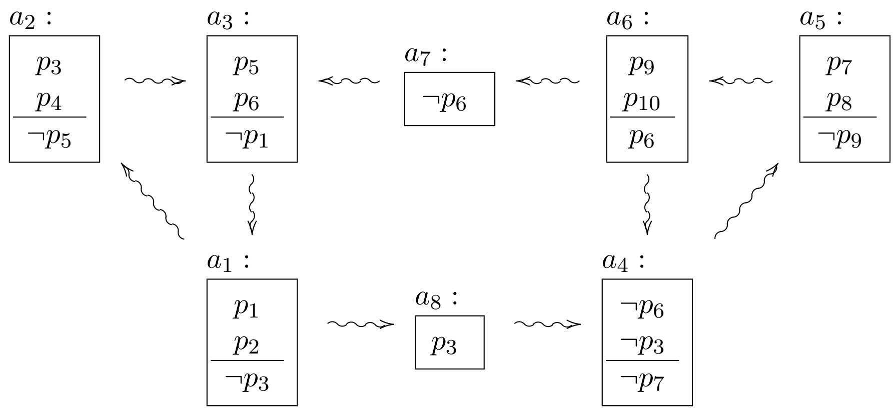
Step (1): Since every thesis a_C is neither part of a red circle itself nor connected to a red circle via a red directed path from that circle to a_C, a partial evaluation of τ—excluding all the arguments in red circles—will determine the validity value of a_C. Introducing such a thesis a_C for every red circle $C = \{a_1, \ldots, a_n\}$ identifies an argument $b_C \in C$ whose validity value should be zero given a_C's validity value according to the partial evaluation. We transform the graph G_τ to G'_τ by deleting the ingoing red arrows to every b_C—except those which depart from the corresponding a_C.

Step (2): By assumption that a_C is neither part of a red circle itself nor connected to a red circle via a red directed path from that circle to a_C, G'_τ doesn't contain any red circles. Consequently, there is a unique validity function ϑ on G'_τ. Moreover, ϑ agrees with $\vartheta_{\text{partial}}$.

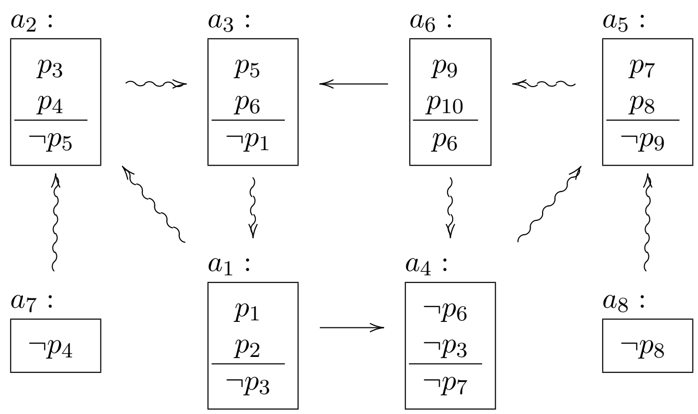
Step (3): To see that ϑ is a validity function on τ, we have to show that it satisfies the recursive definition of validity on G_τ, and, in order to do so, we have to consider but the differences of G_τ and G'_τ only, more precisely: the arguments whose ingoing red arrows were deleted in the transformation. These b_C where chosen above such that $\vartheta(b_C) = 0$. Yet this is in accordance with the recursive definition of validity on G_τ, since the τ-valid argument that attack b_C in G'_τ, namely the corresponding a_Cs, do attack it in G_τ, too.

Finally, that ϑ is unique on τ is warranted by proposition 4. □

The following dialectical structure exemplifies the case where a thesis attacks an argument in a red circle yet doesn't satisfy conditions (ii) and, thence, (iii):[18]



Both theses a_7 and a_8 are connected to a red circle as forbidden by (ii), moreover, they are part of a big red circle, and their validity values cannot be determined, as a consequence, by a partial evaluation of τ. The crux is that the validity value of a_8 depends on the evaluation of the left red circle which depends on the evaluation of a_7 which depends on the evaluation of the right red circle which, finally, depends on the evaluation of a_8. In the following example, in contrast, the antecends conditions of theorem 7 are met:



The τ-valid arguments are a_3, a_6, a_7, a_8.

As a special case of this general solution the problem of self-attacking arguments is solved, too (compare Prakken and Vreeswijk, 2001, p. 237). If a dialectical structure contains an argument with premisses $\neg c, p_1, \ldots, p_n$ and conclusion c then every rational proponent o_i must negate at least on of that argument's premisses which can thence be introduced as a τ_i-valid thesis into the dialectical structure.

Theses introduction, as a reconstruction method, helped to solve the two aspects of the red-circle problem, i.e. that validity functions on dialectical

[18]Green arrows are omitted.

structures realising red circles might be ambiguous, or not exist at all. Now how can these insights be integrated into the general framework of debate evaluation? One could simple stipulate that a stance attribution has to possess the properties (stated in proposition 7) which ensure the existence and uniqueness of the validity functions on the induced subdebates. These properties are summarised in the following definition.

Definition 10 (Complete stance-attribution) *Let $\tau = \langle T, A, U\rangle$ be a dialectical structure. The stance-attribution $S : \{o_1, \ldots, o_k\} \rightarrow \mathcal{P}\ (T)$ is called "complete" iff there is a thesis a_C for every red circle C in every subdebate τ_i which*

(i) attacks one of C's arguments,

(ii) is neither part of a red circle itself nor connected to a red circle via a red directed path from that circle to a_C, and

(iii) is assigned the validity value 1 according to a partial evaluation of τ_i, $\vartheta_{\text{partial}}$, which excludes all arguments in red circles.

Completeness of a stance-attribution is, unlike fulfilling one's burdens of proof, not a discursive aim proponents pursue consciously in a debate, but rather a requirement which is fulfilled qua rational reconstruction of a debate. Completeness, thus understood, is a guiding principle for those who analyse debates rather than for the proponents themselves. If completeness, as specified above, is required, the rational reconstruction of a debate, as a consequence, guarantees that a validity function exists and is unique on every subdebate induced by the stance-attribution. And if a validity-function exists and is unique on a circular dialectical structure, the definitions of the previous sections and most notably the evaluation procedure sketched in section 5 can be applied. Yet theorem 2 cannot be applied to circular structures. More importantly, however, theorem 1 cannot easily be transposed to the general case either. The most prominent counter-example to that theorem in the general case are circular justifications, i.e. green directed circles. These issues deserve further attention.

References

Andrew B. Baker and Matthew L. Ginsberg. A theorem prover for prioritized circumscription. In N. S. Sridharan, editor, *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 463–467. Morgan Kaufmann, 1989.

M. C. Beardsley. *Practical Logic.* Prentice-Hall, 1950.

Gregor Betz. The vicious circle theorem – a graph-theoretical analysis of dialectical structures. *Argumentation*, 19(1):53–64, March 2005.

Lawrence Birnbaum. Argument molecules: A functional representation of argument structure. In David L. Waltz, editor, *Proceedings of the National Conference on Artificial Intelligence*, pages 63–65. AAAI Press, 1982.

Andrei Bondarenko, Phan Minh Dung, Robert A. Kowalski, and Francesca Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93(1-2):63–101, June 1997.

Claudette Cayrol and Marie-Christine Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In Lluis Godo, editor, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 8th European Conference, ECSQARU 2005*, volume 3571 of *Lecture Notes in Computer Science*, pages 378–389. Springer, 2005. ISBN 3-540-27326-3.

Carlos Iván Chesñevar, Ana Gabriela Maguitman, and Ronald Prescott Loui. Logical models of argument. *ACM Computing Surveys*, 32(4):337–383, 2000. ISSN 0360-0300. doi: http://doi.acm.org/10.1145/371578.371581.

Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

P.M. Dung, R.A. Kowalski, and F. Toni. Dialectic proof procedures for assumption-based, admissible argumentation. *Artificial Intelligence*, 170 (2):114–159, February 2006.

Trudy Govier. *A Practical Study of Argument*. Wadsworth Publishing Company, 1985.

Charles L. Hamblin. Mathematical models of dialogue. *Theoria*, 37:130–155, 1971.

Jaakko Hintikka. The logic of information-seeking dialogues. a model. In Werner Becker and Wilhelm K. Essler, editors, *Konzepte der Dialektik*, pages 212–231. Klostermann, 1981.

Fangzhen Lin and Yoav Shoham. Argument systems: A uniform basis for nonmonotonic reasoning. In Ronald J. Brachman, Hector J. Levesque, and Raymond Reiter, editors, *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning*, pages 245–255. Morgan Kaufmann, 1989.

Paul Lorenzen and Kuno Lorenz. *Dialogische Logik*. Wissenschaftliche Buchgesellschaft, 1978.

J. D. Mackenzie. Question-begging in non-cumulative systems. *Journal of Philosophical Logic*, 8:117–133, 1979.

John E. Nolt. *Informal Logic: Possible Worlds and Imagination*. McGraw-Hill, 1984.

Manuel Ojeda-Aciego, Inman P. de Guzmán, Gerhard Brewka, and Luís Moniz Pereira, editors. *Logics in Artificial Intelligence, European Workshop, JELIA 2000 Malaga, Spain, September 29 - October 2, 2000, Proceedings*, volume 1919 of *Lecture Notes in Computer Science*, 2000. Springer. ISBN 3-540-41131-3.

John L. Pollock. Defeasible reasoning. *Cognitive Science*, 11(4):481–518, 1987.

John L. Pollock. The structure of epistemic justification. *American Philosophical Quarterly*, 4:62–78, 1970.

John L. Pollock. *Cognitive Carpentry. A Blueprint for How to Build A Person*. MIT Press, 1995.

Henry Prakken. On dialogue systems with speech acts, arguments, and counterarguments. In Ojeda-Aciego et al. (2000), pages 224–238. ISBN 3-540-41131-3.

Henry Prakken and Giovanni Sartor. A dialectical model of assessing conflicting arguments in legal reasoning. *Artif. Intell. Law*, 4(3-4):331–368, 1996.

Henry Prakken and Gerard Vreeswijk. Logics for defeasible argumentation. In D.M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 4, pages 219–318. Kluwer, 2 edition, 2001.

Nicholas Rescher. *Dialectics. A Controversy-Oriented Approach to the Theory of Knowledge*. State University of New York Press, 1977.

Hubert Schleichert. *Wie man mit Fundamentalisten diskutiert, ohne den Verstand zu verlieren. Anleitung zum subversiven Denken*. Beck, 1998.

Guillermo Ricardo Simari and Ronald Prescott Loui. A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence*, 53(2-3):125–157, 1992.

Guillermo Ricardo Simari, Carlos Iván Chesñevar, and Alejandro J. García. The role of dialectics in defeasible argumentation. In *Anales de la XIV*

Conferencia Internacional de la Sociedad Chilena para Ciencias de la Computacíon, pages 270–281. Univ. de Concepción, Chile, November 1994.

A. Francisca Snoeck Henkemans. State-of-the-art: The structure of argumentation. *Argumentation*, 14:447–473, 2000.

Holm Tetens. *Philosophisches Argumentieren*. Beck, 2004.

Stephen N. Thomas. *Practical Reasoning in Natural Language*. Prentice-Hall, 3rd edition, 1986.

Stephen E. Toulmin. *The Uses of Argument*. Cambridge University Press, 1958.

Frans H. van Eemeren and Rob Grootendorst. *A Systematic Theory of Argumentation. The pragma-dialectical approach*. Cambridge University Press, 2004.

Bart Verheij. *Rules, Reasons, Arguments. Formal studies of argumentation and defeat*. Dissertation Universiteit Maastricht, 1996.

Gerard Vreeswijk. Defeasible dialectics: A controversy-oriented approach towards defeasible argumentation. *Journal of Logic and Computation*, 3 (3):317–334, 1993.

Gerard Vreeswijk and Henry Prakken. Credulous and sceptical argument games for preferred semantics. In Ojeda-Aciego et al. (2000), pages 239–253. ISBN 3-540-41131-3.

Gerard A. W. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90(1-2):225–279, February 1997.

Douglas N. Walton and Erik C. W. Krabbe. *Commitment in Dialogue. Basic Concepts of Interpersonal Reasoning*. State University of New York Press, 1995.