Noname manuscript No.
(will be inserted by the editor)

A Structured Argumentation Framework for Modeling Debates in the Formal Sciences

Marcos Cramer · Jérémie Dauphin

Received: date / Accepted: date

This paper is a significant extension of a workshop paper presented at the 2017 International Workshop on Theory and Applications of Formal Argument (Dauphin and Cramer, 2017).

Jérémie Dauphin has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 690974 for the project "MIREL: MIning and REasoning with Legal texts".

M. Cramer
TU Dresden, International Center for Computational Logic, Nöthnitzer Straße 46, 01187 Dresden, Germany, Tel.: +49-35146338426, E-mail: marcos.cramer@uni-dresden.de

J. Dauphin
University of Luxembourg, 2, avenue de l'Université, 4365 Esch-sur-Alzette, Luxembourg, Tel.: +352-4666445364, E-mail: jeremie.dauphin@uni.lu

Abstract Scientific research in the formal sciences comes in multiple degrees of formality: fully formal work; rigorous proofs that practitioners know to be formalizable in principle; and informal work like rough proof sketches and considerations about the advantages and disadvantages of various formal systems. This informal work includes informal and semi-formal debates between formal scientists, e.g. about the acceptability of foundational principles and proposed axiomatizations. In this paper, we propose to use the methodology of structured argumentation theory to produce a formal model of such informal and semi-formal debates in the formal sciences. For this purpose, we propose ASPIC-END, an adaptation of the structured argumentation framework ASPIC+ which can incorporate natural deduction style arguments and explanations. We illustrate the applicability of the framework to debates in the formal sciences by presenting a simple model of some arguments about proposed solutions to the Liar paradox, and by discussing a more extensive – but still preliminary – model of parts of the debate that mathematicians had about the Axiom of Choice in the early 20th century.

Keywords Argumentation theory · Formal sciences · Natural deduction · Hypothetical reasoning · Axiom of Choice

1 Introduction

Scientific research in the formal sciences (mathematics, logic, theoretical computer science, axiomatic metaphysics, formal linguistics, game theory etc) comes in multiple degrees of formality: fully formal work, which is often performed with the help of computer systems for interactive theorem proving, as it quickly becomes too tedious for humans to explicate all their reasoning in a formal system; fully rigorous proofs that practitioners precisely know how to formalize; practically rigorous work that practitioner know to be formalizable in principle; and informal work like rough proof sketches and considerations about the advantages and disadvantages of various formal systems. Historically, there has been a move from more informal approaches to more formal ones, e.g. in the mathematics of the first half of the 20th century, when the foundational crisis led to an increased attention to axiomatization and to rigorous proofs. This move has generally been accompanied by debates among formal scientists, e.g. about the acceptability of foundational principles and proposed axiomatizations. Despite being informed by formal considerations, these debates have generally been articulated in an informal or semi-formal way.

In this paper, we propose to use the methodology of *structured argumentation theory* (see Besnard et al, 2014) to produce formal models of such informal and semi-formal debates in the formal sciences. Structured argumentation theory allows for a fine-grained model of argumentation and argumentative reasoning based on a formal language and evaluated according to the principles developed in Dung-style *abstract argumentation theory* (see Dung, 1995; Baroni et al, 2011).

One of the dominant formal frameworks for structured argumentation is the *ASPIC+ framework* (see Modgil and Prakken, 2014). In ASPIC+, arguments are built from axioms and premises as well as from strict and defeasible rules, in a similar manner as proofs are built from axioms and rules in a Hilbert-style proof system. The distinction between strict and defeasible rules amounts to the difference between deductively valid modes of inference (e.g. conjunction introduction), and defeasible principles that generally hold but allow for exception (e.g. that dogs generally have four legs). Three kinds of attacks between arguments, *undermines*, *undercuts* and *rebuttals*, are defined between arguments, and finally an *argumentation semantics* from abstract argumentation theory (see Baroni et al, 2011) is applied to determine which sets of arguments can be rationally accepted.

Arguments in the formal sciences often involve hypothetical reasoning, which involves reasoning based on an assumption or hypothesis that is locally assumed to be true for the sake of the argument, but to which there is no commitment on the global level. Such hypothetical reasoning is captured well by natural deduction proof systems, whereas the Hilbert-style definition of arguments in ASPIC+ cannot account for such hypothetical reasoning.

ASPIC+ does not allow strict rules to be attacked, which means that debates about which rules of inference are correct, cannot be modeled in ASPIC+. But sometimes formal scientists debate about which rules of inference are deductively valid. ASPIC-END replaces the strict rules of ASPIC+ by *intuitively strict rules*, which formalize the *prima facie* laws of logic which we pre-theoretically consider to be valid without exceptions, but which can nevertheless be given up after more careful examination. Unlike the strict rules of ASPIC+, an intuitively strict rule can be attacked by another argument, but unlike for a defeasible rule, the conclusion of an intuitive strict rule cannot be rejected if both the antecedent of the rule and the rule itself is accepted.

Scientific discourse is characterized not only by the exchange of arguments in favor and against various scientific hypotheses, but also by the attempt to provide scientific *explanations*. In the context of abstract argumentation, Šešelja and Straßer (2013) have therefore proposed to incorporate the notion of *explanation* into argumentation theory, in order to model scientific debate more faithfully. So far, this incorporation of explanation into argumentation theory has not been extended to the case of structured argumentation. The two contributions of the current paper in this direction are a general framework for incorporating explanation into structured argumentation and a particular proposal for how to define explanations in instantiations of that framework in the domain of paradoxes arising in the formal sciences.

We propose an adaptation of the ASPIC+ framework called *ASPIC-END* that allows for incorporating hypothetical reasoning and explanations (see Section 3). We illustrate the applicability of the framework to debates in the formal sciences through two instantiations of the framework: First, we present in detail a model of a very simple set of arguments about proposed solutions to the Liar Paradox (see Section 4). The presentation of this model only serves to illustrate the functioning of ASPIC-END on a simple example and does not

purport to be a model of philosophically noteworthy arguments on this topic. In Section 5 we sketch and discuss a more extensive model that formalizes parts of the debate that mathematicians had about the Axiom of Choice in the early 20th century (see Moore, 1982). Given that the model still leaves out many contributions to that debate and additionally simplifies some of the contributions that it does take into account, we consider it to only be a preliminary model that we plan to extend in the future. However, we hope that this more extensive model gives some insight into the strengths and drawbacks of the modeling capacities of ASPIC-END, as well as inspiration for further research into this direction.

In order to ensure that the ASPIC-END framework behaves as one would rationally expect, as was previously done for ASPIC+ (see Modgil and Prakken, 2013), we have proved multiple rationality postulates about ASPIC-END in a technical online appendix (Cramer and Dauphin, 2018).

We see two primary motivations for applying the methodology of structured argumentation theory to debates in the formal sciences: First, it is a suitable testbed for structured argumentation theory: Applying structured argumentation theory to real-life debates is often very challenging, because of many layers of uncertainty and imprecision in the interpretation of most types of debates, caused by ambiguities and vagueness of natural language, by a lack of a formal understanding of the domain of discourse of the debate, as well as by the limited rationality of the humans involved in the debate. In the case of debates in the formal sciences, all of these problems are alleviated to some degree: Formal scientists tend to avoid ambiguities and minimize vagueness in their scientific usage of natural language, especially so in the more formal parts of their work, but also in the more informal parts. We have a much better formal understanding of the domains of discourse of the formal sciences than of practically any other domains of discourse. And the debates that scientists have on scientific topics of their field generally show a higher degree of rationality than debates that non-scientists have. For these reasons, it can be hoped that structured argumentation theory can be more easily, and thus hopefully more fruitfully, applied to debates in the formal sciences than to many other kinds of debates to which it has been applied so far. This could also more clearly than existing application bring to light the drawbacks of current approaches in structured argumentation theory, which could become an impetus for further developments in the field.

The second motivation for applying structured argumentation theory to debates in the formal sciences is that in the long run, once the methodology and the models it produces become more mature, such models could contribute to a better understanding of what is at stake in debates in the formal sciences, and hence to a better understanding about the foundations of formal sciences. In this respect, we see the proposed methodology as complimentary to and combinable with the work within the emerging field of *computation metaphysics*, in which methods from automated and interactive theorem proving are used to fully formalize axiomatic theories of metaphysics. The term *computation metaphysics* was first coined by Fitelson and Zalta (2007), who formalized parts

of Abstract Object Theory (see Zalta, 2012) with PROVER9. More recently, significant contributions to this field of research were made by Benzmüller and Woltzenlogel Paleo (2016), who with the help of an automated higher-order theorem prover discovered a so far undetected inconsistency in Gödel's ontological argument, and by Benzmüller et al (2017), who used higher-order theorem provers to expose some mistakes and novel insights in a long-standing controversy between Háyek and Anderson concerning a variant of Gödel's ontological argument. This work shows that full formalization of work in a formal field of research can yield real benefits to advance the research in such a field. But so far, this methodology has been limited to the study of the object level of formal axiomatic theories, whereas the meta-level debates that formal scientists have about such theories could not be captured within the formalizations. One way in which the methodology proposed in this paper could complement the existing methodology of automated theorem proving is that is could allow such meta-level debates to also be captured within a formal model, so that the discovery of mistakes and new insights with the help of automated theorem proving could be extended to this level.

2 Related work & motivation for ASPIC-END

The work of Dung (1995) introduced the theory of *abstract argumentation*, in which one models arguments by abstracting away from their internal structure to focus on the relation of conflict between them. This gives rise to the notion of an *argumentation framework*, which formally is just a directed graph, whose informal interpretation is that the vertices stand for arguments and the edges stand for the attack relation between arguments, i.e. the relation between a counterargument and the argument that it counters. Given an argumentation framework, the goal is to select a set of arguments deemed acceptable on the sole basis of the attack relation between the arguments. There are various approaches for making such selections, based on different criteria such as conflict-freeness (i.e. never simultaneously accepting two arguments where one attacks the other), defense (accepting an attacked argument only if you also accept counterarguments to all its attackers), and maximality (which among other things ensures that an unattacked argument will always be accepted). A selection of arguments that are deemed simultaneously acceptable according to some criteria is called an *extension*. Sometimes, especially when there are cycles in the argumentation framework, there might be multiple extensions that satisfy the given criteria. For this reason, the formal definition of an *abstract argumentation semantics* is that it is a function that maps any given argumentation framework to a set of sets of arguments (vertices) of that argumentation framework.

In *structured argumentation*, one models also the internal structure of arguments through a formal language in which arguments and counterarguments can be constructed (Besnard et al, 2014). One important family of frameworks for structured argumentation is the family of ASPIC-like frameworks, which

is based on the work of John Pollock (e.g. 1987; 1995) and consists among others of the original ASPIC framework (Prakken, 2010), the ASPIC+ framework (Modgil and Prakken, 2014), and the ASPIC- framework (Caminada et al, 2014). We briefly sketch ASPIC+, as it is the basis for our framework ASPIC-END.

In ASPIC+, one starts with a knowledge base and a set of rules[1] which allow one to make inferences from given knowledge. There are two kinds of rules: *Strict rules* logically entail their conclusion, whereas *defeasible rules* only create a presumption in favour of their conclusion. Arguments are built either by introducing an element of the knowledge base into the framework, or by making an inference based on a rule and the conclusions of previous arguments. Attacks between arguments are constructed either by attacking a fallible premise of an argument (*undermining*), by attacking the conclusion of a defeasible inference made within an argument (*rebuttal*), or by questioning the applicability of such a rule (*undercutting*). Preferences between arguments can be derived from preferences between rules. An abstract argumentation framework can thus be built and acceptable arguments can be selected using any abstract argumentation semantics.

Caminada and Amgoud (2007) have introduced the notion of *rationality postulates* for structured argumentation frameworks. These are conditions that structured argumentation frameworks would rationally be expected to satisfy, such as closure under strict rules of the output and consistency of the conclusions given consistency of the strict rules. Caminada and Amgoud (2007) showed that the original ASPIC system did not satisfy these postulates, but proposed minor changes that made it satisfy them. These changes have been incorporated into ASPIC+ (Modgil and Prakken, 2013).

ASPIC-END features three main differences from ASPIC+. The first is that it allows for arguments to introduce an assumption on which to reason hypothetically, just like in natural deduction. In natural deduction, hypothetical derivations are employed in the inference schemes called $\neg$-Introduction (or *proof by contradiction*), $\supset$-Introduction (we use $\supset$ for the material implication), and $\vee$-Elimination (or *reasoning by cases*). Allowing for the usage of defeasible rules within hypothetical reasoning leads to specific problems that have been studied for the inference scheme of reasoning by cases in a recent paper by Beirlaen et al (2017). In the current paper we avoid these problems by not allowing defeasible rules within hypothetical reasoning. However, a conclusion made on the basis of an inference scheme involving hypothetical reasoning

[1] In this paper, we use the word *rule* in the way in which it is usually used in the structured argumentation literature. There is one important difference between this usage of *rule* and the way the word is usually used in the logical literature outside of structured argumentation theory: A *rule*, as the word is used in structured argumentation theory, is what would normally be called an instance of a rule. For this reason, it makes sense to speak of a *rule scheme* (as we will frequently do in Section 5), which is what would normally be just called a rule.

may still be incorporated into an argument that uses defeasible rules, so that there is some integration of defeasible and hypothetical reasoning.[2]

The second difference is that ASPIC-END allows for arguments about the correct rules of logical reasoning. In ASPIC+, such arguments cannot be modeled, as the rules of logical reasoning represented by strict rules, and arguments involving only strict rules can never be attacked. Argumentation about the correct rules of logical reasoning is quite common in debates in the formal sciences. For example, our *prima facie* intuitions suggest that it is a law of logic that a sentence that is not true must be false. However, the Kripke-Feferman solution to the Liar paradox (Reinhardt, 1986; Feferman, 1991) suggests that some sentences, such as the Liar sentence, are neither true nor false, since giving them either one of the two truth values leads to a contradiction. This solution is not putting forward an argument against the falsehood of the sentence by rebutting it, nor is it undermining any of the argument's premises. It is undercutting the argument by attacking the inference made from the negation of truth to falsehood.

To allow such arguments about the correct laws of logic to be modeled in ASPIC-END, we replace strict rules by *intuitively strict rules* whose applicability can be questioned, as in the case of defeasible rules in ASPIC+, but which behave like strict rules when their applicability is accepted. This means that conclusions of intuitively strict rules cannot be rebutted, just as for strict rules in ASPIC+. Intuitively strict rules represent *prima facie laws of logic*, i.e. purportedly logical inference rules which make sense at first but are open to debate.

The third difference is that ASPIC-END has a notion of *explanations* additionally to the notion of arguments. This feature is based on the work of Šešelja and Straßer (2013), who have extended Dung-style abstract argumentation with *explananda* (phenomena that need to be explained) and an *explanatory relation*, which allows arguments to either explain these explananda or deepen another argument's explanation. In Section 3, we will need some definitions from Šešelja and Straßer (2013):

Definition 1. An explanatory argumentation framework (EAF) is a tuple $\langle \mathcal{A}, \mathcal{X}, \rightarrow, \dashrightarrow \rangle$, where $\mathcal{A}$ is a set of arguments, $\mathcal{X}$ is a set of explananda, $\rightarrow$ is an attack relation between arguments and $\dashrightarrow$ is an explanatory relation from arguments to either explananda or arguments.

If $A \dashrightarrow B$, we say that A *explains* B.

Sets of admissible arguments are then selected:

Definition 2. Let $\Delta = \langle \mathcal{A}, \mathcal{X}, \rightarrow, \dashrightarrow \rangle$ be an EAF, $A \in \mathcal{A}$ and $S \subseteq \mathcal{A}$. We say that S is *conflict-free* iff there are no arguments $B, C \in S$ such that $B \rightarrow C$.

[2] The early formalisms of Pollock (1987) and Pollock (1995) also allowed for arguments involving hypothetical reasoning. Most of the work in structured argumentation theory that built on this early work of Pollock ignored this type of arguments. In a recent paper, Beirlaen et al (2018) have critically assessed the way hypothetical arguments function in Pollock's formalisms and have identified three problematic features of the formalism in Pollock (1995). By not allowing defeasible rules within hypothetical reasoning, we avoid these problematic features.

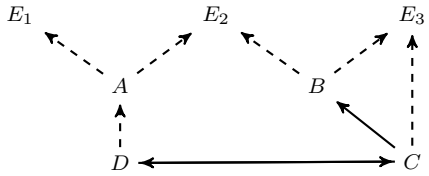

Fig. 1 Example of explanatory power and depth: $\{B\} >_p \{C\}$ and $\{A, B\} >_p \{B\}$, but $\{A\}$ and $\{C\}$ are incomparable with respect to explanatory power. $\{A, D\} >_d \{A\}$, but $\{A\}$ and $\{B\}$ are incomparable with respect to explanatory depth.

We say that S *defends* A iff for every $B \in \mathcal{A}$ such that $B \rightarrow A$, there exists $C \in S$ such that $C \rightarrow B$. We say that S is *admissible* iff S is conflict-free and for all $B \in S$, S defends B.

The most suitable admissible sets are then selected by also taking into account their explanatory power and depth. These are measured by first identifying the explanations present in each set of arguments.

Definition 3. Let $\Delta = \langle \mathcal{A}, \mathcal{X}, \rightarrow, \dashrightarrow \rangle$ be an EAF, $S \subseteq \mathcal{A}$ and $E \in \mathcal{X}$. An *explanation* $X[E]$ for E offered by S is a set $S' \subseteq S$ such that there exists a unique argument $A \in S'$ such that $A \dashrightarrow E$ and for all $A' \in S' \setminus \{A\}$, there exists a path in $\dashrightarrow$ from A' to A.

In order to be able to compare sets of arguments on how many explananda they can explain and in how much detail, the two following measures are required:

Definition 4. Let $\Delta = \langle \mathcal{A}, \mathcal{X}, \rightarrow, \dashrightarrow \rangle$ be an EAF and $S, S' \subseteq \mathcal{A}$. Let $\mathcal{E}$ be the set of explananda S offers an explanation for and $\mathcal{E}'$ the set of explananda S' offers an explanation for. We say that S is *explanatory more powerful than* S' ($S >_p S'$) if and only if $\mathcal{E} \supsetneq \mathcal{E}'$.

Definition 5. Let $\Delta = \langle \mathcal{A}, \mathcal{X}, \rightarrow, \dashrightarrow \rangle$ be an EAF and $S, S' \subseteq \mathcal{A}$. We say that S is *explanatory deeper than* S' ($S >_d S'$) if and only if for each explanation X' offered by S', there is an explanation X offered by S such that $X' \subseteq X$ and for at least one such X and X' pair, $X' \subsetneq X$.

Šešelja and Straßer (2013) define two procedures for selecting the most suitable sets of arguments. The first procedure (for the *argumentative core*) consists in selecting the most explanatory powerful conflict-free sets, from which the maximal most defended sets are then retained. The second procedure (for the *explanatory core*) selects the most explanatory powerful conflict-free sets, from which the most defended sets are taken, and then from those selects the minimal explanatory deepest sets. In our formalism, we will slightly alter and reformulate these procedures.

3 ASPIC-END

In this section, we define ASPIC-END and motivate the details of its definition.

Definition 6. An *argumentation theory* is a tuple $(\mathcal{L}, \mathcal{R}, n, <)$, where:

- $\mathcal{L}$ is a logical language containing a set of free variables $\mathcal{L}_v$ and closed under the binary connective disjunction ($\vee$), the unary connectives negation ($\neg$), the three types of assumability ($Assumable_\neg, Assumable_\vee, Assumable_\supset$), and the existential quantifiers (if $\varphi \in \mathcal{L}$ and $x \in \mathcal{L}_v$, then $\forall x.\varphi, \exists x.\varphi \in \mathcal{L}$) such that $\bot \in \mathcal{L}$.
- $\mathcal{R} = \mathcal{R}_{is} \cup \mathcal{R}_d$ is a set of intuitively strict ($\mathcal{R}_{is}$) and defeasible ($\mathcal{R}_d$) rules of the form $\varphi_1, \ldots, \varphi_n \rightsquigarrow \varphi$ and $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$ respectively, where $n \geq 0$ and $\varphi_i, \varphi \in \mathcal{L}$.
- $n : \mathcal{R} \to \mathcal{L}$ is a partial function.
- $\mathcal{R}_{ce} := \{(\bot \rightsquigarrow \alpha) \mid \alpha \in \mathcal{L}\} \subseteq \mathcal{R}_{is}$, and $\forall r \in \mathcal{R}_{ce}$, $n(r)$ is undefined.
- $<$ is an asymmetric and transitive relation over R_d which represents preference.

Note that we interpret $\bot$ not just as any contradiction but as the conjunction of all formulas in the language. We thus require that rules are present in the framework which allow one to derive any formula from $\bot$, which are effectively rules of conjunction elimination.

We now inductively define how to construct arguments. At the same time, we define five functions on arguments that specify certain features of any given argument: $\mathsf{Conc}(A)$ denotes the conclusion of argument A. $\mathsf{As}_\neg(A)$, $\mathsf{As}_\vee(A)$ and $\mathsf{As}_\supset(A)$ denote the set of assumptions under which argument A is operating: $\mathsf{As}_\neg(A)$ stands for the assumptions made for a proof by contradiction, or negation introduction, $\mathsf{As}_\vee(A)$ stands for the assumptions made for reasoning by cases, or disjunction elimination, and $\mathsf{As}_\supset(A)$ stands for the assumptions made for an implication introduction. As a short-hand, we will sometimes write $\mathsf{As}(A) := \mathsf{As}_\neg(A) \cup \mathsf{As}_\vee(A) \cup \mathsf{As}_\supset(A)$. So whenever $\mathsf{As}(A) \neq \emptyset$, A is a hypothetical argument. $\mathsf{Sub}(A)$ denotes the set of sub-arguments of A. $\mathsf{DefRules}(A)$ denotes the set of all defeasible rules used in A. $\mathsf{TopRule}(A)$ denotes the last inference rule which has been used in the argument if such a rule exists, and is undefined otherwise.

Definition 7. An *argument* A on the basis of an argumentation theory $\Sigma = (\mathcal{L}, \mathcal{R}, n, <)$ has one of the following forms:

1. $A_1, \ldots, A_n \rightsquigarrow \psi$, where $A_1, \ldots, A_n$ are arguments such that there exists an intuitively strict rule $\mathsf{Conc}(A_1), \ldots, \mathsf{Conc}(A_n) \rightsquigarrow \psi$ in $\mathcal{R}_{is}$.
   $\mathsf{Conc}(A) := \psi$, $\mathsf{As}_\neg(A) := \mathsf{As}_\neg(A_1) \cup \cdots \cup \mathsf{As}_\neg(A_n)$,
   $\mathsf{As}_\vee(A) := \mathsf{As}_\vee(A_1) \cup \cdots \cup \mathsf{As}_\vee(A_n)$, $\mathsf{As}_\supset(A) := \mathsf{As}_\supset(A_1) \cup \cdots \cup \mathsf{As}_\supset(A_n)$,
   $\mathsf{Sub}(A) := \mathsf{Sub}(A_1) \cup \cdots \cup \mathsf{Sub}(A_n) \cup \{A\}$,
   $\mathsf{DefRules}(A) := \mathsf{DefRules}(A_1) \cup \cdots \cup \mathsf{DefRules}(A_n)$,
   $\mathsf{TopRule}(A) := \mathsf{Conc}(A_1), \ldots, \mathsf{Conc}(A_n) \rightsquigarrow \psi$.
2. $A_1, \ldots, A_n \Rightarrow \psi$, where $A_1, \ldots, A_n$ are arguments s.t. $\mathsf{As}(A_1) \cup \ldots \cup \mathsf{As}(A_n) = \emptyset$ and there exists a defeasible rule $\mathsf{Conc}(A_1), \ldots, \mathsf{Conc}(A_n) \Rightarrow \psi$ in $\mathcal{R}_d$.
   $\mathsf{Conc}(A) := \psi$, $\mathsf{As}_\neg(A) := \emptyset$,
   $\mathsf{As}_\vee(A) := \emptyset$, $\mathsf{As}_\supset(A) := \emptyset$,

$\mathsf{Sub}(A) := \mathsf{Sub}(A_1) \cup \cdots \cup \mathsf{Sub}(A_n) \cup \{A\}$,
$\mathsf{DefRules}(A) := \mathsf{DefRules}(A_1) \cup \cdots \cup \mathsf{DefRules}(A_n) \cup$
$\{\mathsf{Conc}(A_1), \ldots, \mathsf{Conc}(A_n) \Rightarrow \psi\}$,
$\mathsf{TopRule}(A) := \mathsf{Conc}(A_1), \ldots, \mathsf{Conc}(A_n) \Rightarrow \psi$.

3. $\mathsf{Assume}_{\neg}(\varphi)$, where $\varphi \in \mathcal{L}$.
   $\mathsf{Conc}(A) := \varphi$, $\mathsf{As}_{\neg}(A) := \{\varphi\}$,
   $\mathsf{As}_{\vee}(A) := \emptyset$, $\mathsf{As}_{\supset}(A) := \emptyset$,
   $\mathsf{Sub}(A) := \{\mathsf{Assume}_{\neg}(\varphi)\}$,
   $\mathsf{DefRules}(A) := \emptyset$, $\mathsf{TopRule}(A)$ is undefined.
4. $\mathsf{Assume}_{\vee}(\varphi)$, where $\varphi \in \mathcal{L}$.
   $\mathsf{Conc}(A) := \varphi$, $\mathsf{As}_{\neg}(A) := \emptyset$,
   $\mathsf{As}_{\vee}(A) := \{\varphi\}$, $\mathsf{As}_{\supset}(A) := \emptyset$,
   $\mathsf{Sub}(A) := \{\mathsf{Assume}_{\vee}(\varphi)\}$,
   $\mathsf{DefRules}(A) := \emptyset$, $\mathsf{TopRule}(A)$ is undefined.
5. $\mathsf{Assume}_{\supset}(\varphi)$, where $\varphi \in \mathcal{L}$.
   $\mathsf{Conc}(A) := \varphi$, $\mathsf{As}_{\neg}(A) := \emptyset$,
   $\mathsf{As}_{\vee}(A) := \emptyset$, $\mathsf{As}_{\supset}(A) := \{\varphi\}$,
   $\mathsf{Sub}(A) := \{\mathsf{Assume}_{\supset}(\varphi)\}$,
   $\mathsf{DefRules}(A) := \emptyset$, $\mathsf{TopRule}(A)$ is undefined.
6. $\mathsf{ProofByContrad}(\neg\varphi, A')$, where A' is an argument such that $\varphi \in \mathsf{As}_{\neg}(A')$ and $\mathsf{Conc}(A') = \bot$.
   $\mathsf{Conc}(A) := \neg\varphi$, $\mathsf{As}_{\neg}(A) := \mathsf{As}_{\neg}(A') \setminus \{\varphi\}$,
   $\mathsf{As}_{\vee}(A) := \mathsf{As}_{\vee}(A')$, $\mathsf{As}_{\supset}(A) := \mathsf{As}_{\supset}(A')$,
   $\mathsf{Sub}(A) := \mathsf{Sub}(A') \cup \{\mathsf{ProofByContrad}(\neg\varphi, A')\}$,
   $\mathsf{DefRules}(A) := \mathsf{DefRules}(A')$, $\mathsf{TopRule}(A)$ is undefined.
7. $\mathsf{ReasonByCases}(\psi, A_1, A_2, A_3)$, where:
   A_1 is an argument such that $\varphi \in \mathsf{As}_{\vee}(A_1)$ and $\mathsf{Conc}(A_1) = \psi$,
   A_2 is an argument such that $\varphi' \in \mathsf{As}_{\vee}(A_2)$ and $\mathsf{Conc}(A_2) = \psi$,
   A_3 is an argument such that $\mathsf{Conc}(A_3) = \varphi \vee \varphi'$.
   $\mathsf{Conc}(A) := \psi$,
   $\mathsf{As}_{\neg}(A) := \mathsf{As}_{\neg}(A_1) \cup \mathsf{As}_{\neg}(A_2) \cup \mathsf{As}_{\neg}(A_3)$,
   $\mathsf{As}_{\vee}(A) := (\mathsf{As}_{\vee}(A_1) \setminus \{\varphi\}) \cup (\mathsf{As}_{\vee}(A_2) \setminus \{\varphi'\}) \cup \mathsf{As}_{\vee}(A_3)$,
   $\mathsf{As}_{\supset}(A) := \mathsf{As}_{\supset}(A_1) \cup \mathsf{As}_{\supset}(A_2) \cup \mathsf{As}_{\supset}(A_3)$,
   $\mathsf{Sub}(A) := \mathsf{Sub}(A_1) \cup \mathsf{Sub}(A_2) \cup \mathsf{Sub}(A_3) \cup \{\mathsf{ReasonByCases}(\psi, A_1, A_2, A_3)\}$,
   $\mathsf{DefRules}(A) := \mathsf{DefRules}(A_1) \cup \mathsf{DefRules}(A_2) \cup \mathsf{DefRules}(A_3)$,
   $\mathsf{TopRule}(A)$ is undefined.
8. $\supset\text{-}\mathsf{intro}(\varphi \supset \psi, A')$, where A' is an argument such that $\varphi \in \mathsf{As}_{\supset}(A')$ and $\mathsf{Conc}(A') = \psi$.
   $\mathsf{Conc}(A) := \varphi \supset \psi$, $\mathsf{As}_{\neg}(A) := \mathsf{As}_{\neg}(A')$,
   $\mathsf{As}_{\vee}(A) := \mathsf{As}_{\vee}(A')$, $\mathsf{As}_{\supset}(A) := \mathsf{As}_{\supset}(A') \setminus \{\varphi\}$,
   $\mathsf{Sub}(A) := \mathsf{Sub}(A') \cup \{\supset\text{-}\mathsf{intro}(\varphi \supset \psi, A')\}$,
   $\mathsf{DefRules}(A) := \mathsf{DefRules}(A')$, $\mathsf{TopRule}(A)$ is undefined.
9. $\forall\text{-}\mathsf{intro}(\forall x.\varphi(x), A')$, where A' is an argument such that for some $x \in \mathcal{L}_v$, there is no $\psi \in \mathsf{As}(A')$ such that x is free in ψ, and $\mathsf{Conc}(A') = \varphi(x)$.
   $\mathsf{Conc}(A) := \forall x.\varphi(x)$, $\mathsf{As}_{\neg}(A) := \mathsf{As}_{\neg}(A')$,
   $\mathsf{As}_{\vee}(A) := \mathsf{As}_{\vee}(A')$, $\mathsf{As}_{\supset}(A) := \mathsf{As}_{\supset}(A')$,

$\mathsf{Sub}(A) := \mathsf{Sub}(A') \cup \{\forall\text{-}\mathsf{intro}(\forall x.\varphi(x), A')\}$,
$\mathsf{DefRules}(A) := \mathsf{DefRules}(A')$, $\mathsf{TopRule}(A)$ is undefined.

Notice that we do not allow for the use of defeasible rules within hypothetical arguments, as reflected in the condition of Def. 7 item 2 that the sub-arguments cannot have any assumptions. We do however allow for the conclusions of defeasible arguments to be imported inside of a hypothetical argument. This is motivated by the fact that allowing for proofs by contradiction amounts to allowing for transpositions of any rule that can be used within a proof by contradiction, and transpositions are usually assumed only for strict rules in structured argumentation (Caminada and Amgoud, 2007; Modgil and Prakken, 2013).

Example 1. Consider an argumentation theory $AT_1 = (\mathcal{L}, \mathcal{R}, n, <)$, where $\mathcal{L}$ is the smallest set containing $\{p, q, r, s, u\}$ and satisfying Definition 6 item 1, $\mathcal{R}_{is} = \{p \rightsquigarrow q; q \rightsquigarrow \bot; \rightsquigarrow r\}$, $\mathcal{R}_d = \{\neg p, r \Rightarrow s; u \Rightarrow q\}$ and $<$ is the empty relation. We can then construct an argument for s as follows:

- A_1:= $\mathsf{Assume}_\neg(p)$, with $\mathsf{As}_\neg(A_1) = \{p\}, \mathsf{Conc}(A_1) = p$
- A_2:= $A_1 \rightsquigarrow q$, with $\mathsf{As}_\neg(A_2) = \{p\}, \mathsf{Conc}(A_2) = q$
- A_3:= $A_2 \rightsquigarrow \bot$, with $\mathsf{As}_\neg(A_3) = \{p\}, \mathsf{Conc}(A_3) = \bot$
- A_4:= $\mathsf{ProofByContrad}(\neg p, A_3)$, with $\mathsf{As}_\neg(A_4) = \emptyset, \mathsf{Conc}(A_4) = \neg p$
- A_5:= $\rightsquigarrow r$, with $\mathsf{As}_\neg(A_5) = \emptyset, \mathsf{Conc}(A_5) = r$
- A_6:= $A_4, A_5 \Rightarrow s$, with $\mathsf{As}_\neg(A_6) = \emptyset, \mathsf{Conc}(A_6) = s$

We can see that A_1 introduces the assumption p, and from there the arguments A_2 and A_3 manage to derive a contradiction, which allows the construction of argument A_4 with conclusion $\neg p$ under no assumption. We can then use this together with the premise r to form an argument for s. Note however that we cannot form an argument for $\neg u$ using a proof by contradiction, because to derive an inconsistency from u we would have to use d_2. However, defeasible rules can only be applied under no assumption, hence we would be unable to apply it in the proof by contradiction for $\neg u$.

We now need to define the attack relation in our framework. Notice that in ASPIC-END, we also allow for an argument A to attack an argument B which makes an assumption φ if A concludes that φ is not assumable. For example, if one were to assume that the number 5 is yellow, since numbers do not have colors, it should be possible to attack the argument that introduces this assumption and any argument making an inference from this assumption. We also separate the assumption-attack into the three different kinds of assumptions, so that one can, for example, deny a formula's assumability for reasoning by cases but still allow it to be assumed for implication-introduction. Additionally, if one wishes, for example, to refute the well-foundedness of a construction such as proof by contradiction while still accepting reasoning by cases, one simply needs to attack the $\neg$-assumability of all formulas.

Definition 8. Let $\Sigma = (\mathcal{L}, \mathcal{R}, n, <)$ be an argumentation theory and A, B two arguments on the basis of Σ. We say that A *attacks* B iff A *rebuts*, *undercuts* or *assumption-attacks* B, where:

- *A rebuts* argument B (on B') iff $\mathsf{Conc}(A) = \neg\varphi$ or $\neg\mathsf{Conc}(A) = \varphi$ for some $B' \in \mathsf{Sub}(B)$ of the form $B'_1, \ldots, B'_n \Rightarrow \varphi$ and $\mathsf{As}(A) = \emptyset$.
- *A undercuts* argument B (on B') iff $\mathsf{Conc}(A) = \neg n(r)$ or $\neg\mathsf{Conc}(A) = n(r)$ for some $B' \in \mathsf{Sub}(B)$ such that $\mathsf{TopRule}(B') = r$, there is no $\varphi \in \mathsf{As}(B')$ such that $\neg\varphi = \mathsf{Conc}(A')$ or $\varphi = \neg\mathsf{Conc}(A')$ for some $A' \in \mathsf{Sub}(A)$, and there are arguments $B_1, ..., B_n$ such that $B_1 = B'$, $B_n = B$, $B_i \in \mathsf{Sub}(B_{i+1})$ for $1 \leq i < n$ and $\mathsf{As}(A) \subseteq \mathsf{As}(B_1) \cup \cdots \cup \mathsf{As}(B_n)$.
- *A assumption-attacks* B (on B') iff for some $B' \in \mathsf{Sub}(B)$ such that $\mathsf{As}(A) = \emptyset$ and one of the following holds:
  - $B' = \mathsf{Assume}_{\neg}(\varphi)$ and $\mathsf{Conc}(A) = \neg Assumable_{\neg}(\varphi)$;
  - $B' = \mathsf{Assume}_{\vee}(\varphi)$ and $\mathsf{Conc}(A) = \neg Assumable_{\vee}(\varphi)$;
  - $B' = \mathsf{Assume}_{\supset}(\varphi)$ and $\mathsf{Conc}(A) = \neg Assumable_{\supset}(\varphi)$.

We require that any attacking argument A is making fewer assumptions than the B' it attacks, as to prevent arguments from attacking outside of their assumption scope. Note that in the case of rebuttal, since the attacked argument cannot have assumptions, we require that the attacking argument have none either.

In the case of undercutting, we also have the requirement that A does not use the contrary of any assumptions made by B' in any of its inferences, since the attack would not stand in the scope of B'. Additionally, we allow A to make use of any assumptions appearing in the chain of arguments leading B' to B, as these assumptions, even if they have been retracted, still constitute valid grounds on which to form an attack.

Similarly as in ASPIC+, one can also define a notion of successful attack by lifting the preference relation from rules to arguments as follows:

Definition 9. Let $\Sigma = (\mathcal{L}, \mathcal{R}, n, <)$ be an argumentation theory and A, B be two arguments on the basis of Σ. We define the *lifting of* $<$ *to arguments* $\prec$ to be such that $A \prec B$ iff there exists $r_a \in \mathsf{DefRules}(A)$, such that for all $r_b \in \mathsf{DefRules}(B)$, we have $r_a < r_b$.

Notice that this lifting corresponds to elitist weakest-link as described by Modgil and Prakken (2014). We believe that this ordering is best suited for modeling philosophical and scientific arguments.

We now define what it means for an attack to be successful:

Definition 10. Let $\Sigma = (\mathcal{L}, \mathcal{R}, n, <)$ be an argumentation theory, A, B be two arguments on the basis of Σ. We say that A *successfully rebuts* B iff A rebuts B on B' for some argument B' and $A \not\prec B'$, and that A *defeats* B iff A assumption-attacks, undercuts or successfully rebuts B.

The aim of our system is to generate an EAF as defined in Section 2. For this three things need to be specified: A set $\mathcal{X}$ of explananda, a condition under which an argument explains an explanandum, and a condition under which an argument explains another argument. The first two of these three details are domain-specific, and are thus to be specified in an instantiation of the ASPIC-END framework. The third one, on the other hand, should be

the same in all domains. The reason for this can be found in the informal clarification that Šešelja and Straßer (2013) provided for what it means to say that an argument b explains an argument a: "argument b can be used to explain one of the premises of argument a [...] or the link between the premises and the conclusion."

In the context of structured argumentation, this informal clarification can be turned into a formal definition:

Definition 11. Let A, B be arguments. We say that B *explains* A (on A') iff $A' \in \mathsf{Sub}(A)$, $\mathsf{As}(B) \subseteq \mathsf{As}(A')$ and at least one of the following two cases holds:

- $A' \notin \mathsf{Sub}(B)$ and either $A' = (\rightsquigarrow \mathsf{Conc}(B))$ or $A' = (\Rightarrow \mathsf{Conc}(B))$.
- $\mathsf{Conc}(B) = n(\mathsf{TopRule}(A'))$ and $\nexists B' \in \mathsf{Sub}(B)$ such that $\mathsf{TopRule}(B') = \mathsf{TopRule}(A')$.

Intuitively, the idea behind this definition is that an argument B explains another argument A if B non-trivially concludes one of A's premises or one of the inference rules used by A.

We now have all the elements needed to build an EAF.

Definition 12. Let $\Sigma = (\mathcal{L}, \mathcal{R}, n, <)$ be an argumentation theory. Let $\mathcal{X}$ be a set of explananda, and let $\mathcal{C}$ be a criterion for determining whether an argument constructed from Σ explains a given explanandum $E \in \mathcal{X}$. The *explanatory argumentation framework* (EAF) *defined by* $(\Sigma, \mathcal{X}, \mathcal{C})$ is a tuple $\langle \mathcal{A}, \mathcal{X}, \rightarrow, \dashrightarrow \rangle$, where:

- $\mathcal{A}$ is the set of all arguments that can be constructed from Σ satisfying Definition 7;
- $(A, B) \in \rightarrow$ iff A defeats B, where $A, B \in \mathcal{A}$;
- $(A, E) \in \dashrightarrow$ iff criterion $\mathcal{C}$ is satisfied with respect to A and E, where $A \in \mathcal{A}$ and $E \in \mathcal{X}$;
- $(A, B) \in \dashrightarrow$ iff A explains B according to Definition 11, where $A, B \in \mathcal{A}$.

Once such a framework has been generated, we want to be able to extract the most interesting sets of arguments. Such a set should be able to explain as many explananda in as much detail as possible, while being self-consistent and plausible.

We define two kinds of extensions corresponding to the two selection procedures defined by Šešelja and Straßer (2013). As suggested in the informal discussion in their paper, we chose to give higher importance to the criterion of defense compared to the criterion of explanatory power. This prevents some absurd theories which manage to explain all explananda but cannot defend themselves against all attacks from beating plausible theories which fail to explain some of the explananda but are sound and fully defended.

Definition 13. Let $\Sigma = (\mathcal{L}, \mathcal{R}, n, <)$ be an argumentation theory, $\Delta = \langle \mathcal{A}, \mathcal{X}, \rightarrow, \dashrightarrow \rangle$ the EAF defined by Σ and $S \subseteq \mathcal{A}$ a set of arguments.

1. We say that S is *satisfactory* iff S is admissible and there is no $S' \subseteq \mathcal{A}$ such that $S' >_p S$ and S' is admissible.

2. We say that S is *insightful* iff S is satisfactory and there is no $S' \subseteq \mathcal{A}$ such that $S' >_d S$ and S' is satisfactory.
3. We say that S is an *argumentative core extension* (*AC-extension*) of Δ iff S is satisfactory and there is no $S' \supset S$ such that S' is satisfactory.
4. We say that S is an *explanatory core extension* (*EC-extension*) of Δ iff S is insightful and there is no $S' \subset S$ such that S' is insightful.

The AC-extensions are sets of arguments which represent the theories explaining the most explananda, together with all other compatible beliefs present in the framework. EC-extensions represent the core of those theories and only include the arguments which defend or provide details for them.

We define the conclusions of the arguments in a given extension as follows:

Definition 14. Let $\Sigma = (\mathcal{L}, \mathcal{R}, n, <)$ be an argumentation theory, $\Delta = \langle \mathcal{A}, \mathcal{X}, \rightarrow, \dashrightarrow \rangle$ be the EAF defined by Σ and S be an extension of Δ. Then, we define the *conclusions of* S, denoted $\mathsf{Concs}(S)$, to be $\mathsf{Concs}(S) = \{\mathsf{Conc}(A) | A \in S \text{ s.t. } \mathsf{As}(A) = \emptyset\}$.

4 Modelling explanations of semantic paradoxes in ASPIC-END

In this section, we discuss how ASPIC-END can be applied to modeling argumentation about explanations of semantic paradoxes, and illustrate this potential application with a simple example. We start by briefly motivating this application of structured argumentation theory.

Philosophy is an academic discipline in which good argumentative skills are a central part of every student's training. Philosophical texts are often much richer in explicit formulation of arguments than texts from other academic disciplines. For these reasons, we believe that modeling arguments from philosophical textbooks, monographs and papers can be an interesting test case for structured argumentation theory.

Different areas of philosophy vary with respect to how much logical rigor is commonly applied in the presentation of arguments. Even logically rigorous argumentation poses many interesting problems, as the rich literature on abstract and structured argumentation attests. In order to not confound these interesting problems with issues arising from the lack of logical rigor, it is a good idea to concentrate on the study of logically rigorous argumentation. Philosophical logic is an area of logic where logically rigorous arguments abound. One topic that has gained a lot of attention in philosophical logic is the study of semantic paradoxes such as the Liar paradox and Curry's paradox (Beall et al, 2016; Field, 2008). We therefore use the argumentation about the various explanations of the paradoxes that have been proposed in the philosophical literature as a test case for structured argumentation theory.

In our application of ASPIC-END to argumentation about explanations of semantic paradoxes, the explananda are the paradoxes (i.e. arguments that derive an absurdity under no assumption without using defeasible rules), which other arguments can explain by attacking the said derivation. So we instantiate

the set $\mathcal{X}$ of explananda and criterion $\mathcal{C}$ for an explanation of an explanandum by an argument as specified in the following two definitions:

Definition 15. Let $\Sigma = (\mathcal{L}, \mathcal{R}, n, <)$ be an argumentation theory. For every argument A on the basis of Σ such that $\mathsf{DefRules}(A) = \emptyset, \mathsf{As}(A) = \emptyset$ and $\mathsf{Conc}(A) = \bot$, we stipulate an explanandum E_A, and say that $\mathsf{Source}(E_A) = A$. We define the set $\mathcal{X}$ of explananda based on Σ to be the set of all explananda E_A that we have thus stipulated.

Definition 16. Let $\Sigma = (\mathcal{L}, \mathcal{R}, n, <)$ be an argumentation theory, A an argument and E an explanandum based on Σ. We say that criterion $\mathcal{C}$ is satisfied with respect to A and E iff A defeats $\mathsf{Source}(E)$.

The following example illustrates an application of ASPIC-END to a version of the Liar paradox and two very simple explanations of it:[3]

Example: Define L to be the sentence "L is false". If L is true, i.e. "L is false" is true, then L is false, which is a contradiction. So L is not true, i.e. L is false. So "L is false" is true, i.e. L is true. So we have the contradiction that L is both true and false from no assumption.

A truth-value gap explanation: In this paradox, the only inference steps that are not justified by the laws of classical logic are the steps that involve reasoning about the meaning of "true" and "false". Since classical logic is a well-studied system for formalizing rational reasoning, we should accept it. Thus we need to give up some inference rules based on the meaning of "true" and "false". This can be achieved by giving up the assumption that every sentence is either true or false for problematically self-referential sentences such as L. In the paradox, this assumption is used when concluding that L is false because L is not true, so this inference should be rejected.

A paracomplete explanation: If we give up some of the natural inference rules that are based on the meaning of "true" and "false", our formalism no longer correctly captures the meaning of "true" and "false", so we should not give up these rules. In order to avoid the paradox, we therefore need to limit some rules of classical logic. This can be achieved by allowing a proof by contradiction based on assumption ϕ only in case the law of excluded middle holds for ϕ, i.e. in case $\phi \vee \neg\phi$. The law of excluded middle should not be accepted for problematically self-referential statements like L, and thus also

[3] Note that our aim here is not to present a detailed case study of how a debate about a semantic paradox can be formalized in ASPIC-END, but only to illustrate the way ASPIC-END works and could be used for such a case study in future work. For this reason, we restrict ourselves to a simple exposition of the Liar paradox and two very simple explanations of it, a truth-value gap explanation and a paracomplete explanation. See Field (2008) for comprehensive presentations of truth-value gap and paracomplete explanations, besides many others. Additionally note that, for the sake of simplicity, we only include in our model those instances of rules that are actually used in the explanations that we formalize, so we leave out other instances of the general rules (rule schemes) that lie behind these instances. A detailed case study would have to consider what happens when all instances of these rules are included; for this purpose, other paradoxes like Curry's paradox and various revenge versions of the Liar paradox would need to be considered as well, as the instances of these rules applied to the paradoxical sentences from these other paradoxes would be included in the model.

not for the statement "L is true". So "L is true" cannot be assumed for a proof by contradiction, i.e. the derivation of "L is not true" based on deriving a contradiction from the assumption the L is true is not valid.

We now proceed to the ASPIC-END model of the reasoning and argumentation involved in the paradox and the two explananda. We use T and F to mean *true* and *false* respectively; the other abbreviations we use should be self-explanatory from the context. The rules in our model are such that $\mathcal{R}_{is}$ is the smallest set satisfying Def 6 item 1 and including the rules listed below. For each intuitively strict rule, we provide either a brief explanation of where the rule comes from, or we refer to the name of the corresponding rule in Field (2008), of which the rule in question is an instance:

$T(L) \rightsquigarrow T(F(L))$	(by definition, as L is defined to mean $F(L)$)
$T(F(L)) \rightsquigarrow F(L)$;	(T-Elim)
$T(L), F(L) \rightsquigarrow \bot$;	(a sentence cannot be both true and false)
$\neg T(L) \rightsquigarrow F(L)$;	(a sentence that is not true is considered false)
$F(L) \rightsquigarrow T(F(L))$;	(T-Introd)
$T(F(L)) \rightsquigarrow T(L)$;	(by definition, as L is defined to mean $F(L)$)
$\rightsquigarrow \forall r.(\mathit{used\_in\_paradox}(r) \wedge \neg \mathit{T\text{-}F\text{-}rule}(r) \supset r \in \mathit{classical\_logic})$	(all inference rules that are used in the derivation of the paradox and that are not based on the meaning of "true" and "false" are admissible in classical logic)

The naming function is defined by $n(\neg T(L) \rightsquigarrow F(L)) = r_1$. The set $\mathcal{R}_d$ of defeasible rules is defined as follows:

- $\Rightarrow \mathit{formalizes\_rational\_reasoning}(\mathit{classical\_logic})$;
- $\mathit{formalizes\_rational\_reasoning}(\mathit{classical\_logic}) \Rightarrow \mathit{accept}(\mathit{classical\_logic})$;
- $\forall r.(\mathit{used\_in\_paradox}(r) \wedge \neg \mathit{T\text{-}F\text{-}rule}(r) \supset r \in \mathit{classical\_logic})$, $\mathit{accept}(\mathit{classical\_logic}) \Rightarrow \exists r.(\mathit{T\text{-}F\text{-}rule}(r) \wedge \mathit{give\_up}(r))$;
- $\Rightarrow \mathit{problematically\_self\text{-}referential}(L)$;
- $\mathit{problematically\_self\text{-}referential}(L), \exists r.(\mathit{T\text{-}F\text{-}rule}(r) \wedge \mathit{give\_up}(r)) \Rightarrow \neg r_1$;
- $\Rightarrow \mathit{correctly\_capture}(\mathit{TF\text{-}meaning})$;
- $\mathit{correctly\_capture}(\mathit{TF\text{-}meaning}) \Rightarrow \neg\exists r.(\mathit{T\text{-}F\text{-}rule}(r) \wedge \mathit{give\_up}(r))$;
- $\forall r.(\mathit{used\_in\_paradox}(r) \wedge \neg \mathit{T\text{-}F\text{-}rule}(r) \supset r \in \mathit{classical\_logic})$, $\neg\exists r.(\mathit{T\text{-}F\text{-}rule}(r) \wedge \mathit{give\_up}(r)) \Rightarrow \neg \mathit{accept}(\mathit{classical\_logic})$;
- $\mathit{problematically\_self\text{-}referential}(L), \mathit{accept}(\mathit{classical\_logic}) \Rightarrow \neg \mathit{accept}(T(L) \vee \neg T(L))$;
- $\neg \mathit{accept}(T(L) \vee \neg T(L)) \Rightarrow \neg \mathit{Assumable}_{\neg}(T(L))$

Infinitely many arguments can be constructed from this argumentation theory. However, the following set of arguments is the set of most relevant arguments, in the sense that other arguments will not defeat these arguments

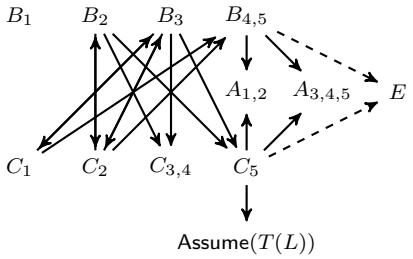

Fig. 2 The relevant arguments, explanandum, attacks and explanations from the example

and will not add relevant new conclusions.

$$
\begin{aligned}
A_{1,2} &= \mathsf{ProofByContrad}(\neg T(L), (\mathsf{Assume}_{\neg}(T(L)),\\
&\quad ((\mathsf{Assume}_{\neg}(T(L)) \rightsquigarrow T(F(L))) \rightsquigarrow F(L)) \rightsquigarrow \bot)) \rightsquigarrow F(L)\\
A_{3,4,5} &= ((A_1 \rightsquigarrow T(F(L))) \rightsquigarrow T(L)), A_1 \rightsquigarrow \bot\\
B_1 &= (\rightsquigarrow \forall r.(\mathit{used\_in\_paradox}(r) \wedge \neg \mathit{T\text{-}F\text{-}rule}(r) \supset r \in \mathit{classical\_logic}))\\
B_2 &= B_1, (\Rightarrow \mathit{formalizes\_rational\_reasoning}(\mathit{classical\_logic})) \Rightarrow \mathit{accept}(\mathit{classical\_logic})\\
B_3 &= B_2 \Rightarrow \exists r.(\mathit{T\text{-}F\text{-}rule}(r) \wedge \mathit{give\_up}(r))\\
B_{4,5} &= (\Rightarrow \mathit{problematically\_self\text{-}referential}(L)), B_3 \Rightarrow \neg r_1\\
C_1 &= (\Rightarrow \mathit{correctly\_capture}(\mathit{TF\text{-}meaning})) \Rightarrow \neg\exists r.(\mathit{T\text{-}F\text{-}rule}(r) \wedge \mathit{give\_up}(r))\\
C_2 &= B_1, C_1 \Rightarrow \neg \mathit{accept}(\mathit{classical\_logic})\\
C_{3,4} &= (\Rightarrow \mathit{problematically\_self\text{-}referential}(L)), C_2 \Rightarrow \neg \mathit{accept}(T(L) \vee \neg T(L))\\
C_5 &= C_{3,4} \Rightarrow \neg \mathit{Assumable}_{\neg}(T(L))
\end{aligned}
$$

We get the explanandum E with $\mathsf{Source}(E) = A_{3,4,5}$. $B_{4,5}$ defeats A_2 on A_1 and C_5 defeats A_2 on $\mathsf{Assume}(T(L))$, thus they both explain E. The AC-extensions are $\{B_1, B_2, B_3, B_{4,5}\}$ and $\{B_1, C_1, C_2, C_{3,4}, C_5\}$, and the EC-extensions are $\{B_3, B_{4,5}\}$ and $\{C_2, C_5\}$.

5 Modelling argumentation on Axiom of Choice

Additionally to the relatively simple model presented in the previous section, we have also applied ASPIC-END to produce a more extensive model of a debate in the formal sciences, namely a model of parts of the debate that mathematicians had about the Axiom of Choice (AC) in the early 20th century (see Moore, 1982). Given that this model is too extensive to be presented within this paper, we will here only present some fragments of the model, briefly describe some features of the overall model, and discuss some of the the insight into the strengths and drawbacks of the modeling capacities of ASPIC-END that we have gained from producing this model. A complete description of our model can be found in our technical online appendix (Cramer and Dauphin, 2018).

In 1904, the German mathematician Ernst Zermelo published a proof of the Well-Ordering Theorem, in which he explicitly referred to a set-theoretic principle that came to be known as the Axiom of Choice (Zermelo, 1904). The Axiom of Choice states that for each set M whose elements are non-empty sets, there is a function f that maps each element $m \in M$ to an element $f(m) \in m$. In the first years after its publication, Zermelo's proof received a lot of critique, a significant part of which questioned the general validity of the Axiom of Choice (see (Moore, 1982)). In the long run, however, the proof got accepted, as the Axiom of Choice got accepted as a valid part of the de-facto standard foundational theory for mathematics, *Zermelo-Fraenkel set theory with the Axiom of Choice* (*ZFC*).

The two critiques of Zermelo's Axiom of Choice that we consider in our model are those of Peano (1906) and Lebesgue (see Hadamard et al, 1905). Furthermore, we consider the counterarguments to these critiques put forward by Zermelo (1908) and by Hadamard (see Hadamard et al, 1905). When constructing the formal model, we have made a number of design choices that enabled us to keep the model relatively simple and concise:

- We have only considered the contributions of Zermelo, Peano, Lebesgue and Hadamard to this debate, leaving out some of the other contributions to the debate that are discussed in Moore (1982). The choice of which contributions to include was partially based on the importance of those contributions from the point of view of the history and philosophy of mathematics, and partially based on considerations about which contributions best illustrate the interesting formal features of the ASPIC-END framework.
- In the case of some arguments, we have opted not to formalize the internal details of the argument, but instead include the conclusion of the argument as a defeasible premise in our model, as this significantly simplifies the model. This solution allows the effect of the argument on the overall debate to be faithfully represented even when the internal details of the argument are not made explicit by the model.
- An additional way in which we kept our model simple was by not formalizing in any detail the uncontested mathematical reasoning that is related to the debate, e.g. parts of the proof of the Well-Ordering Principle that do not make use of the Axiom of choice or the proof of the Partition Principle that Zermelo refers to in one of his arguments.

Due to these simplifications, we consider our model to only be a preliminary model that we plan to extend in the future. However, the model already gives some insight into the strengths and drawbacks of the modeling capacities of ASPIC-END, as well as inspiration for further research into this direction.

In our model, the purely mathematical and purely logical demonstrations and reasoning are formalized using intuitively strict rules, while the philosophical and metamathematical argumentation and reasoning is formalized using defeasible rules. Most of the attacks between arguments attack defeasible arguments, i.e. philosophical or metamathematical arguments. But given that some

of the mathematical and logical principles that were applied in the mathematical and logical reasoning that we model, e.g. the Axiom of Choice and the non-constructivist parts of classical logic, are attacked by some philosophical or metamathematical arguments, there are also some arguments using only intuitively strict rules that get attacked. By the design of ASPIC-END, all such attacks have to be undercuts.

The debate about the Axiom of Choice that we have formalized in our model concerns the purported justification that Zermelo has given for the Axiom of Choice as well as attacks on this purported justification, but it does not involve any mathematical explanations. For this reason, our model of this debate does not make use of the explanatory machinery included in ASPIC-END, but it does make use of other two novel features of ASPIC-END, i.e. hypothetical reasoning and undercuts of intuitively strict rules.

In order to give a flavor of our formal model, we now present some fragments of it and describe some feature of the overall model. We start by looking at the first argument Zermelo presented for the Axiom of Choice in 1904:

> "this logical principle cannot be reduced to a still simpler one, but is used everywhere in mathematical deduction without hesitation. So for example the general validity of the theorem that the number of subsets into which a set is partitioned is less than or equal to the number of its elements, cannot be demonstrated otherwise than by assigning to each subset one of its elements." (Zermelo, 1904, p. 516)

Here are the formal ASPIC-END arguments that we construct to represent this argument and its subarguments:

$$Z_1^{04} = (\Rightarrow \text{simple}(AC))$$
$$Z_2^{04} = (\Rightarrow \neg\exists x.\ \text{calls\_to\_doubt}(x, \text{usage}(AC)))$$
$$Z_3^{04} = (\Rightarrow \exists p.\ \text{demonstrates}(p, PP))$$
$$Z_4^{04} = (\Rightarrow \forall p.\ (\text{demonstrates}(p, PP) \supset \text{uses}(p, AC)))$$
$$Z_5^{04} = \mathsf{Assume}_{\supset}(\text{demonstrates}(p, PP))$$
$$Z_6^{04} = (Z_4^{04}, Z_5^{04} \vdash \exists p, t.\ (\text{demonstrates}(p, t) \wedge \text{uses}(p, AC)))$$
$$Z_7^{04} = \supset\text{-}\mathsf{intro}(\text{demonstrates}(p, PP) \supset \exists p, t.\ (\text{demonstrates}(p, t) \wedge \text{uses}(p, AC)))$$
$$Z_8^{04} = \forall\text{-}\mathsf{intro}(\forall p.\ (\text{demonstrates}(p, PP) \supset \exists p, t.\ (\text{demonstrates}(p, t) \wedge \text{uses}(p, AC))))$$
$$Z_9^{04} = (Z_3^{04}, Z_8^{04} \rightsquigarrow \exists p, t.\ (\text{demonstrates}(p, t) \wedge \text{uses}(p, AC)))$$
$$Z_{10}^{04} = (Z_9^{06} \Rightarrow \text{widely\_used}(AC))$$
$$Z_{11}^{04} = (Z_1^{06}, Z_6^{06}, Z_{10}^{06} \Rightarrow \text{accept}(AC))$$

The rules that are needed to construct these arguments can actually be read off from the arguments, and are explicitly stated in the technical online appendix (Cramer and Dauphin, 2018, p. 4-6). The notation $(A_1, \ldots, A_n \vdash \psi)$ used in argument Z_6^{04} stands for an argument that uses multiple rules of

intuitionistic logic to get from the conclusions of arguments $A_1, \ldots, A_n$ to the conclusion ψ. Of course, all these rules are included in our model. Note that argument Z_7^{04} makes use of $\supset$-Introduction, Z_8^{04} makes use of $\forall$-Introduction.

In a letter to Borel that shortly afterwards got published in the Bulletin de la Société mathématique de France (Hadamard et al, 1905), Lebesgue made a constructivist argument against the Axiom of Choice:

> "I believe that we can only build solidly by granting that it is impossible to demonstrate the existence of an object without defining it."

We formalize Lebesgue's argument through a defeasible premise according to which an existence proof requires definition and a strict rule that allows to reject the Axiom of Choice based on this defeasible premise:

$$L_1^{05} = (\Rightarrow \text{existence\_proof\_requires\_definition})$$
$$L_2^{05} = (L_1^{05} \rightsquigarrow \neg\text{accept}(AC))$$

The rules that we included in the model in order to formalize the arguments that have been explicitly mentioned in the historical debate on the Axiom of Choice can also be used to construct *implicit arguments* that were not explicitly mentioned in the historical debate. It should not come as a surprise that at the current level of development of our methodology, the model has not given rise to philosophically insightful implicit arguments. However, there is an implicit argument that plays an important role with respect to the formal behavior of our model: It is an argument that makes use of the proof by contradiction to construct an attack on Lebesgue's argument L_1^{05} based on Zermelo's 1908 argument Z_{29}^{08} for the Axiom of Choice:

$$I_1 = (\mathsf{Assume}_{\neg}(\text{existence\_proof\_requires\_definition}))$$
$$I_2 = (I_1 \rightsquigarrow \neg\text{accept}(AC))$$
$$I_3 = (Z_{29}^{08}, I_2 \rightsquigarrow \bot)$$
$$I_4 = (\mathsf{ProofbyContrad}(I_3, \neg\text{existence\_proof\_requires\_definition}))$$

The idea is that assuming a premise ("existence_proof_requires_definition") of Lebesgue's argument against the Axiom of Choice, we can derive that the Axiom of Choice should not be accepted, which in combination with Zermelo's argument for the acceptance of the Axiom of Choice leads to a contradiction. So we have a proof by contradiction for ¬existence_proof_requires_definition, which thus attacks Lebesgue's argument. The relevance of this argument to the formal properties of our model is explained in Section 1.7 of the technical online appendix (Cramer and Dauphin, 2018).

While the model described here has not led to philosophically relevant implicit arguments, we believe that the methodology we are proposing has the potential to bring to light such arguments once more sophisticated formal models of debates in the formal sciences are constructed. We expect the use

of automated theorem provers to be helpful in order to discover philosophically relevant implicit arguments in more sophisticated models, just like they already have been used by Benzmüller and Woltzenlogel Paleo (2016) and Benzmüller et al (2017) to discover philosophically relevant mistakes and insights in axiomatic theories of metaphysics, as explained in the last paragraph of the Introduction. This would allow for the discovery of mistakes and new insights at the meta-level of debates about formal theories rather than just at the object level of the theories themselves.

We consider it one of the strengths of our methodological approach that it allows to identify such implicit arguments that no one has put forward, but that could be put forward and that could have a relevant influence on the outcome of the debate.

Without imposing preferences on the set of rules, all attacks in our model other than the just mentioned undercuts would become *practically* bidirectional. By this we mean that even though there can be a unidirectional attack from some argument A to some argument B, in such a case there will always be an attack back onto A from some argument B' that is closely related to B and accepted in the same circumstances as B. In order to make the model more interesting and more realistic, we have therefore include in it a preference order on the rules, which by Definition 9 gives rise to a preference order on the arguments. One drawback of our methodology is that it gives no methodological guidance on how to select a preference order on the rules, which is the main determining factor for which extensions are finally accepted. In our model, we followed our common sense of the relative strength of different arguments from the historical debate in order to specify the preference order between the rules.

The set of rules of our model allow for infinitely many arguments to be constructed, so that the EAF corresponding to the model will also be infinite. However, only a small finite subset of this infinite EAF contains attacks that are relevant for the overall status of the acceptability of the Axiom of Choice, which was the focus of attention of the debate that we have formally modeled. In Figure 3, we depict the small subset of relevant arguments and the defeats between them. In this depiction, the letter in the argument name (Z, P, L or H) refers to either Zermelo, Peano, Lebesgue or Hadamard as the source of the argument, and the subscript indicates the year in which the argument was presented (with the 19 dropped, as they were all presented between 1904 and 1908). For the precise content of the argument and the details of their formalization in ASPIC-END, please refer to the technical online appendix (Cramer and Dauphin, 2018). Here we concisely sketch the content of the arguments that have not yet been specified above:

- P_2^{06}: Peano points out that in an 1890 publication he had already considered and rejected the assumption that infinitely many arbitrary choices can be made in an argument.
- P_{14}^{06}: Peano points out that while a single arbitrary choice and thus any finite number of arbitrary choices can be formalized in his *Formulario Math-*

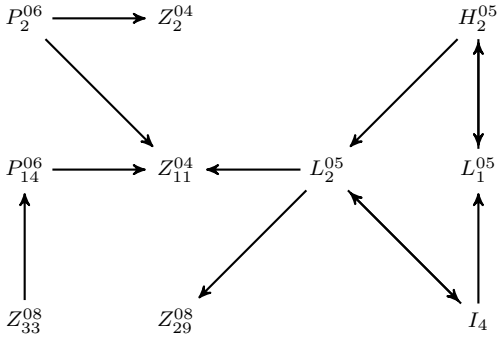

Fig. 3 The relevant arguments and attacks from the example

ematico, an infinite number of arbitrary choices would require an infinitely long argument, which is not allowed in his *Formulario Mathematico*. This argument has the implicit premise that an argument can be accepted if and only if it can be formalized in the *Formulario Mathematico*.

- Z_{29}^{08}: Zermelo points out that Peano himself arrived at the fundamental principles of his *Formulario Mathematico* by analyzing the rules of inference that have historically been recognized as valid and by referring both to the intuitive evidence for the rules and to their necessity for science. He then argues that the Axiom of Choice can be justified in the same way: Multiple set theorists have implicitly applied it, which supports both the claim that it has historically been recognized as valid and that it is intuitively evident. Furthermore, Zermelo lists seven theorems which he believed not to be provable without the Axiom of Choice, and concludes that the Axiom of Choice is necessary for science.
- Z_{33}^{08}: The implicit assumption in P_{14}^{06} (see above) is incorrect, because by Z_{29}^{08} arguments using the Axiom of Choice can be accepted even though they cannot be formalized in Peano's *Formulario Mathematico*.
- H_2^{05}: Hadamard argues against Lebesgue's premise that an existence proof requires definition by pointing out that historical progress in mathematics was achieved by annexing notions which had previously been considered to be outside mathematics because it was impossible to describe them.

Restricted to this set of relevant arguments, there are two argumentative core (AC) extensions: $S_1 = \{P_2^{06}, Z_{33}^{08}, Z_{29}^{08}, H_2^{05}, I_4\}$, and $S_2 = \{P_2^{06}, Z_{33}^{08}, L_1^{05}, L_2^{05}\}$. This means that arguments P_2^{06} and Z_{33}^{08} are accepted in every AC-extension of our model, while P_{14}^{06}, Z_2^{04} and Z_{11}^{04} are rejected in every AC-extension, and the status of the arguments Z_{29}^{08}, I_4, L_1^{05} and L_2^{05} depends on the choice of AC-extension. This set of relevant arguments contains two arguments with conclusion accept(AC), namely Z_{11}^{04} and Z_{29}^{08}. While the first one gets rejected in both extensions, the second one gets accepted in one and rejected in the other extension, so that overall, the status of the claim accept(AC) depends on the choice of the AC-extension.

These properties of our formal model intuitively correspond to the situation that on the one hand there are compelling arguments both in favor and against the Axiom of Choice, and purely formal methods will not decide which of the two stands is "correct" (if there even is a single "correct" answer here), while on the other hand certain arguments in favor or against the Axiom of Choice are so weak that they do not hold up against the scrutiny provided by certain counterarguments against them.

Of course, the fact that the status of the Axiom of Choice in our formal model of the debate is not determined but depends on the choice of the AC-extension is to a certain extent an artifact of the choice of arguments that we formalized and of the preference order that we imposed. We could have gotten a different result, for example if we had chosen to formalize only strong arguments in favor of the Axiom of Choice and weak arguments against it, or if we had just made significantly different judgments about the preference order on the rules involved in our model. So at the current level of development, such a model cannot be seriously defended as a method for deciding which side in a debate is right. What it can do, however, is to help us discover relevant implicit arguments like argument I_4 in our model (and hopefully with a more developed model also philosophically more relevant implicit arguments), to help us get a more precise understanding of what assumptions are made and what is at stake in a given debate, and to point towards weaknesses of the current methodology of structured argumentation theory, like the lack of a methodological guidance for choosing a preference order on the rules.

6 Conclusion and Future Work

We have proposed the application of the structured argumentation methodology to formally model informal and semi-formal debates in the formal sciences. For this purpose, we have proposed a modification of ASPIC+ called ASPIC-END, which incorporates a formal model of explanations, and features natural-deduction style arguments. We have then discussed two instantiations of ASPIC-END, one that models relatively simple arguments about two solutions the Liar Paradox, and one that constitutes a more extensive model of part of the debate that mathematicians had about the Axiom of Choice in the early 20th century.

In a technical online appendix (Cramer and Dauphin, 2018) we have proved four rationality postulates for ASPIC-END that are analogous to the four postulates that Modgil and Prakken (2013) have established for ASPIC+, as well as two new postulates motivated by the application of structured argumentation to debates in the formal sciences. One problem that ASPIC-END shares with ASPIC+ and that we have left for future work is that it does not satisfy the non-interference postulate (see Caminada et al, 2012).

As explained in the introduction, we believe the methodological approach proposed in this paper to be of significant potential for further research. The model of the debate about the Axiom of Choice sketched in Section 5 could

be extended to a model covering a wider range of topics related to the foundational questions in mathematics as well as active research questions in philosophical logic. Given that with increasing size of the model it becomes more and more difficult to produce the model manually and to find all relevant arguments and attacks, we propose that interactive theorem provers like Isabelle (Nipkow et al, 2002) or HOL Light (Harrison, 2009) be used for producing and studying such extensive formal models. Furthermore, combining the methodology of structured argumentation theory with insights from natural language semantics could lead to formal models that are more faithful to the logical form implicit in natural language, which could strengthen the link between the formalization of a debate and the original natural language form of the debate.

References

Baroni P, Caminada M, Giacomin M (2011) An introduction to argumentation semantics. The Knowledge Engineering Review 26(4):365–410

Beall J, Glanzberg M, Ripley D (2016) Liar Paradox. In: Zalta EN (ed) The Stanford Encyclopedia of Philosophy, winter 2016 edn, Metaphysics Research Lab, Stanford University

Beirlaen M, Heyninck J, Straßer C (2017) Reasoning by Cases in Structured Argumentation. In: Proceedings of SAC/KRR 2017, pp 989–994

Beirlaen M, Heyninck J, Straßer C (2018) A critical assessment of Pollock's work on logic-based argumentation with suppositions. In: Proceedings of the 17th International Workshop on Non-Monotonic Reasoning, forthcoming.

Benzmüller C, Woltzenlogel Paleo B (2016) The Inconsistency in Gödel's Ontological Argument: A Success Story for AI in Metaphysics. In: Kambhampati S (ed) IJCAI 2016, AAAI Press, vol 1-3, pp 936–942, URL `http://www.ijcai.org/Proceedings/16/Papers/137.pdf`

Benzmüller C, Weber L, Woltzenlogel Paleo B (2017) Computer-Assisted Analysis of the Anderson-Hájek Controversy. Logica Universalis 11(1):139–151, DOI 10.1007/s11787-017-0160-9, URL `http://christoph-benzmueller.de/papers/J32.pdf`

Besnard P, Garcia A, Hunter A, Modgil S, Prakken H, Simari G, Toni F (2014) Introduction to structured argumentation. Argument & Computation 5(1):1–4

Caminada M, Amgoud L (2007) On the evaluation of argumentation formalisms. Artificial Intelligence 171(5-6):286–310

Caminada M, Modgil S, Oren N (2014) Preferences and Unrestricted Rebut. In: Computational Models of Argument - Proceedings of COMMA 2014, pp 209–220

Caminada MWA, Carnielli WA, Dunne PE (2012) Semi-stable semantics. Journal of Logic and Computation 22(5):1207–1254, DOI 10.1093/logcom/exr033, URL `http://dx.doi.org/10.1093/logcom/exr033`,

/oup/backfile/content_public/journal/logcom/22/5/10.1093/logcom/exr033/2/exr033.pdf

Cramer M, Dauphin J (2018) Technical online appendix to "A Structured Argumentation Framework for Modeling Debates in the Formal Sciences". URL http://orbilu.uni.lu/retrieve/56794/65813/appendix.pdf

Dauphin J, Cramer M (2017) ASPIC-END: Structured Argumentation with Explanations and Natural Deduction. In: Theory and Applications of Formal Argumentation (TAFA) 2017, Revised Selected Papers, LNAI 10757, pp 51–66

Dung PM (1995) On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence 77(2):321–357

Feferman S (1991) Reflecting on incompleteness. The Journal of Symbolic Logic 56(01):1–49

Field H (2008) Saving Truth from Paradox. Oxford University Press

Fitelson B, Zalta EN (2007) Steps toward a Computational Metaphysics. Journal of Philosophical Logic 36(2):227–247, URL http://www.jstor.org/stable/30226964

Hadamard J, Baire R, Lebesgue H, Borel E (1905) Cinq lettres sur la théorie des ensembles. Bulletin de la Société mathématique de France 33:261–273

Harrison J (2009) HOL Light: An Overview. In: TPHOLs, Springer, vol 5674, pp 60–66

Modgil S, Prakken H (2013) A general account of argumentation with preferences. Artificial Intelligence 195:361–397

Modgil S, Prakken H (2014) The ASPIC+ framework for structured argumentation: a tutorial. Argument & Computation 5(1):31–62

Moore G (1982) Zermelo's axiom of choice: its origins, development, and influence. Studies in the history of mathematics and physical sciences, Springer-Verlag

Nipkow T, Paulson LC, Wenzel M (2002) Isabelle/HOL: a proof assistant for higher-order logic, vol 2283. Springer Science & Business Media

Peano G (1906) Additione. Revista de mathematica 8:143–157

Pollock JL (1987) Defeasible reasoning. Cognitive science 11(4):481–518

Pollock JL (1995) Cognitive carpentry: A blueprint for how to build a person. Mit Press

Prakken H (2010) An abstract framework for argumentation with structured arguments. Argument & Computation 1(2):93–124

Reinhardt WN (1986) Some remarks on extending and interpreting theories with a partial predicate for truth. Journal of Philosophical Logic 15(2):219–251

Šešelja D, Straßer C (2013) Abstract argumentation and explanation applied to scientific debates. Synthese 190(12):2195–2217

Zalta E (2012) Abstract Objects: An Introduction to Axiomatic Metaphysics. Synthese Library, Springer Netherlands, URL https://books.google.de/books?id=dHTvCAAAQBAJ

Zermelo E (1904) Beweis, daß jede Menge wohlgeordnet werden kann. Mathematische Annalen 59(4):514–516
Zermelo E (1908) Neuer Beweis für die Möglichkeit einer Wohlordnung. Mathematische Annalen 65(1):107–128