Frans H. van Eemeren
Bart Garssen • Erik C. W. Krabbe
A. Francisca Snoeck Henkemans
Bart Verheij • Jean H. M. Wagemans

Handbook of Argumentation Theory

With 78 Figures and 1 Table

Springer Reference

Defeat by parallel strengthening is associated with what has been called the "accrual of reasons." When reasons can accrue, it is possible that different reasons for a conclusion are together stronger than each reason separately. For instance, having robbed someone and having injured someone can be separate reasons for convicting someone. But when the suspect is a minor first offender, these reasons may each by itself be rebutted. On the other hand when a suspect has both robbed someone and also injured that person, the reasons may accrue and outweigh the fact that the suspect is a minor first offender. The argument for not punishing the suspect based on the reason that he is a minor first offender is defeated by the "parallel strengthening" of the two arguments for punishing him.

Pollock considered the accrual of reasons to be a natural idea, but argued against it (1995, p. 101 f.). His main point is that it is a contingent fact about reasons whether they accrue or not. For instance, whereas separate testimonies can strengthen each other, the opposite is the case when they are not independent but the result of an agreement between the witnesses. More recent discussions of the accrual of reasons are to be found in Prakken (2005a), Gómez Lucero et al. (2009, 2013), and D'Avila Garcez et al. (2009, p. 155 f.).

11.4 Abstract Argumentation

In 1995, a paper appeared in the journal *Artificial Intelligence* which reformed the formal study of non-monotonic logic and defeasible reasoning: Phan Minh Dung's "On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games" (Dung 1995). By his focus on argument attack as an abstract formal relation, Dung gave the field of study a mathematical basis that inspired many new insights. Dung's approach and the work inspired by it are generally referred to as *abstract argumentation*.[13]

Dung's paper is strongly mathematically oriented and has led to intricate formal studies. However, the mathematical tools used by Dung are elementary. As a result of this, and because of the naturalness of Dung's basic concept of "argument attack," we shall be able, in this section, to explain various concepts studied by Dung without going into much formal detail. This section on abstract argumentation is nevertheless the most formally oriented of the present chapter.

11.4.1 Dung's Abstract Argumentation

The central innovation of Dung's 1995 paper is that he started the formal study of the attack relation between arguments, thereby separating the properties depending exclusively on argument attack from any concerns related to the structure of the

[13] The success of the paper is illustrated by its number of citations. By an imperfect but informative count in Google Scholar of July 22, 2013, there were 1938 citations.

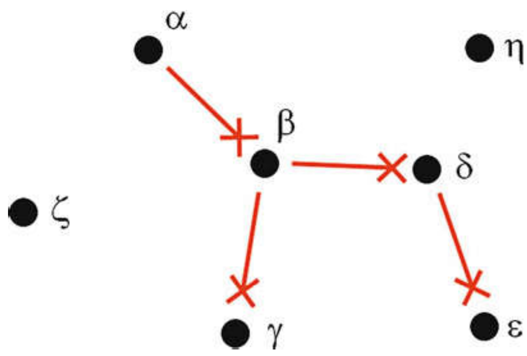


Fig. 11.4 An argumentation framework representing attack between arguments

arguments. Mathematically speaking, the argument attack relation is a directed graph, the nodes of which are the arguments, whereas the edges represent that one argument attacks another. Such a directed graph is called an *argumentation framework*. Figure 11.4 shows an example of an argumentation framework, with the dots representing arguments and the arrows (ending in a cross to emphasize the attacking nature of the connection[14]) representing argument attack.

In Fig. 11.4, the argument α attacks the argument β, which in turn attacks both γ and δ.

Dung's paper consists of two parts, corresponding to two steps in what he refers to as an "analysis of the nature of human argumentation in its full generality" (Dung 1995, p. 324). In the first step, Dung develops the theory of argument attack and how argument attack determines argument acceptability. In the second part, he evaluates his theory by two applications, one consisting of a study of the logical structure of human economic and social problems and the other comprising a reconstruction of a number of approaches to non-monotonic reasoning, among them Reiter's and Pollock's. Notwithstanding the relevance of the second part of the paper, the paper's influence is largely based on the first part about argument attack and acceptability.

In Dung's approach, the notion of an "admissible set of arguments" is central. A set of arguments is *admissible* if two conditions obtain:

1. The set of arguments is *conflict-free*, i.e., does not contain an argument that attacks another argument in the set.
2. Each argument in the set is *acceptable* with respect to the set, i.e., when an argument in the set is attacked by another argument (which by (1) cannot be in the set itself), the set contains an argument that attacks the attacker.

In other words, a set of arguments is admissible if it contains no conflicts and if the set also can defend itself against all attacks. An example of an admissible set of arguments for the framework in Fig. 11.4 is {α, γ}. Since α and γ do not attack one another, the set is conflict-free. The argument α is acceptable with respect to the set since it is not attacked, so that it needs no defense. The argument γ is also

[14] This is especially helpful when also supporting connections are considered; see Sect. 11.5.

acceptable with respect to {α, γ}: the argument γ needs a defense against the attack by β, which defense is provided by the argument α, α being in the set. The set {α, β} is not admissible since it is not conflict-free. The set {γ} is not admissible since it does not contain a defense against the argument β, which attacks argument γ.

Admissible sets of arguments can be used to define argumentation notions of what counts as a proof or a refutation.[15] An argument is "(admissibly) provable" when there is an admissible set of arguments that contains the argument. A minimal such set can be regarded as a kind of "proof" of the argument, in the sense that the arguments in such a set are just enough to successfully defend the argument against counterarguments. An argument is "(admissibly) refutable" when there is an admissible set of arguments that contains an argument that attacks the former argument. A minimal such set can be regarded as a kind of "refutation" of the attacked argument.

Dung speaks of the basic principle of argument acceptability using an informal slogan: the one who has the last word laughs best. The argumentative meaning of this slogan can be explained as follows. When someone makes a claim and that is the end of the discussion, the claim stands. But when there is an opponent raising a counterargument attacking the claim, the claim is no longer accepted – unless the proponent of the claim provides a counterattack in the form of an argument attacking the counterargument raised by the opponent. Whoever has raised the last argument in a sequence of arguments, counterarguments, counter-counterarguments, etc. is the one who has won the argumentative discussion.

Formally, Dung's argumentation principle "the one who has the last word laughs best" can be illustrated using the notion of an "admissible set of arguments." In Fig. 11.4, a proponent of the argument γ clearly has the last word and laughs best, since the only counterargument β is attacked by the counter-counterargument α. Formally, this is captured by the admissibility of the set {α, γ}.

Although the principle of argument acceptability and the concept of an admissible set of arguments seem straightforward enough, it turns out that intricate formal puzzles loom. This has to do with two important formal facts:

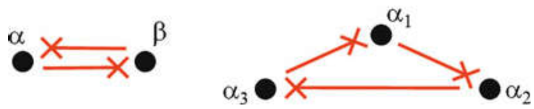
1. It can happen that an argument is both admissibly provable and refutable.
2. It can happen that an argument is neither admissibly provable nor refutable.

The two argumentation frameworks shown in Fig. 11.5 provide examples of these two facts. In the cycle of attacks on the left, consisting of two arguments α and β, each of the arguments is both admissibly provable and admissibly refutable. This is a consequence of the fact that the two sets {α} and {β} are each admissible. For instance, {α} is admissible since it is conflict-free and can defend itself against attacks: the argument α itself defends against its attacker β. By the admissibility of the set {α}, the argument α is admissibly probable, and the argument β admissibly refutable.

The cycle of attacks on the right containing three arguments, α_1, α_2, and α_3, is an example of the second fact above, the fact that it can happen that an argument is neither admissibly provable nor refutable. This follows from the fact that there is no

[15] In the following, we make use of terminology proposed by Verheij (2007).

Fig. 11.5 Arguments attacking each other in cycles



admissible set that contains (at least) one of the arguments, α_1, α_2, or α_3. Suppose that the argument α_3 is in an admissible set. Then the set should defend α_3 against the argument α_2, which attacks α_3. This means that α_1 should also be in the set, since it is the only argument that can defend α_3 against α_2. But this is not possible, because then α_1 and α_3 are both in the set, introducing a conflict in the set. As a result, there is only one admissible set: the empty set that contains no arguments at all. We conclude that no argument is admissibly provable or admissibly refutable.

The framework on the left can be interpreted informally as a situation where there are two reasonable options, as in the case when it has to be decided where to go for one's summer holidays. For instance, for someone living in the Netherlands, it is reasonable to argue that one should go to the south of France, e.g., because of the expected nice weather (argument α), but also that one should go to the north of Norway, e.g., because of a chance to see the Northern Lights (argument β). Arguing for doing both in one and the same holiday period would not normally be considered reasonable, which fact is formally expressed as the arguments attacking each other.

An informal interpretation of the framework on the right can be given in a sports situation involving three teams, where it may be unclear which team is the best one. For instance, consider the Dutch soccer teams Ajax, Feyenoord, and PSV. When Ajax has recently won most matches against Feyenoord, one has reason to think that Ajax is the best team (argument α_3). But when PSV has won most recent matches against Ajax, one has reason to think that PSV is the best (argument α_2), an argument attacking α_3 that Ajax is the best. When it also happens to be the case that Feyenoord has won most recent matches against PSV, there is a reason to think that Feyenoord is the best (argument α_1), attacking argument α_3. Clearly, in this situation (not corresponding to the actual recent match results between the three teams), there is no answer to the question which team is the best. Formally, this corresponds to the fact that none of the three arguments is provable or refutable.

A related formal issue is that, when two sets of arguments are admissible, it need not be the case that their union is admissible. The framework on the left in Fig. 11.5 is an example. As we saw, the two sets {α} and {β} are both admissible, but their union {α, β} is not, since it contains a conflict. This has led Dung to propose the notion of a "preferred extension" of an argumentation framework, which is an admissible set that is as large as possible, in the sense that adding elements to the set makes it not admissible. The framework in Fig. 11.4 has one preferred extension: the set {α, γ, δ, ζ, η}. The framework in Fig. 11.5 on the left has two preferred extensions, {α} and {β}, and the one on the right has one, the empty set.

Some preferred extensions have a special property, namely, that each argument that is not in the set is attacked by an argument in the set. Such an extension is called a *stable extension*. Stable extensions are formally defined as conflict-free sets that attack each argument not in the set. It follows from this definition that a stable extension is also a preferred extension.

The preferred extension {α, γ, δ, ζ, η} of the framework in Fig. 11.4, for instance, is stable, since the arguments β and ε, which are the only ones that are not in the set, are attacked by arguments in the set, α and δ, respectively. The preferred extensions {α} and {β} of Fig. 11.5 (left) are also stable. The preferred extension of Fig. 11.5 (right), the empty set, is not stable, since none of the arguments α_1, α_2, and α_3 is attacked by an argument in the set. This example shows that there exist preferred extensions that are not stable. It also shows that there are argumentation frameworks that do not have a stable extension. In contrast, every argumentation framework has at least one preferred extension (which can be the empty set).

The concepts of preferred and stable extension of an argumentation framework can be regarded as different ways to interpret a framework, and therefore they are often referred to as "preferred semantics" and "stable semantics." Dung (1995) proposed two other kinds of semantics: "grounded semantics" and "complete semantics," and following his paper several additional kinds of semantics have been proposed (see Baroni et al. 2011, for an overview). By the abstract nature of argumentation frameworks, formal questions about the computational complexity of related algorithms and formal connections with other theoretical paradigms came within reach (see, e.g., Dunne and Bench-Capon 2003; Dunne 2007; Egly et al. 2010).

11.4.2 Labelling Arguments

Dung's original definitions are in terms of mathematical sets. An alternative way of studying argument attack is in terms of labelling. Arguments are marked with a label, such as "Justified" or "Defeated" (or IN/OUT, +/−, 1/0, "Warranted"/ "Unwarranted," etc.), and the properties of different kinds of labelling are studied in the field. For instance, the notion of a stable extension corresponds to the following notion in terms of labelling:

> A *stable labelling* is a function that assigns one label "Justified" or "Defeated" to each argument in the argumentation framework such that the following property holds: an argument α is labelled "Defeated" if and only if there is an argument β that attacks α and that is labelled "Justified."

A stable extension gives rise to a stable labelling by labelling all arguments in the extension "Justified" and all other arguments "Defeated." A stable labelling gives rise to a stable extension by considering the set of arguments labelled "Justified."

The idea of labelling arguments can be thought of in analogy with the truth functions of propositional logic, where propositions are labelled with truth-values "true" and "false" (or 1/0, t/f, etc.). In the formal study of argumentation, labelling techniques predate Dung's abstract argumentation (1995). Pollock (1994) uses labelling techniques in order to develop a new version of a criterion that determines warrant.

Verheij (1996b) applied the labelling approach to Dung's abstract argumentation frameworks. He uses argument labelling also as a technique to formally model which arguments are taken into account: in an interpretation of an abstract argumentation framework, the arguments that are assigned a label can be regarded as the ones taken into account, whereas the unlabelled arguments are not considered. Using this idea, Verheij defines two new kinds of semantics: the "stage semantics" and the "semi-stable semantics."[16] Other authors using a labelling approach are Jakobovits and Vermeir (1999) and Caminada (2006). The latter author translated each of Dung's extension types into a mode of labelling.

As an illustration of the labelling approach, we give a labelling treatment of the *grounded extension* of an argumentation framework as defined by Dung.[17] Consider the following procedure in which gradually labels are assigned to the arguments of an argumentation framework:

1. Apply the following to each unlabelled argument α in the framework: if the argument α is only attacked by arguments that have been labelled "Defeated" (or perhaps not attacked at all), label the argument α as "Justified."
2. Apply the following to each unlabelled argument α in the framework: if the argument α is attacked by an argument that has been labelled "Justified," label the argument α as "Defeated."
3. If step 1 and/or step 2 has led to new labelling, go back to step 1; otherwise stop.

When this procedure is completed (which always happens after a finite number of steps when the argumentation framework is finite), the arguments labelled "Justified" constitute the grounded extension of the argumentation framework. Consider, for instance, the framework of Fig. 11.4. In the first step, the arguments α, ζ, and η are labelled "Justified." The condition that all arguments attacking them have been "Defeated" is vacuously fulfilled, since there are no arguments attacking them. In the second step the argument β is labelled "Defeated," since α has been labelled "Justified." Then a second pass of step 1 occurs and the arguments γ and δ are labelled "Justified," since their only attacker β has been labelled "Defeated." Finally, the argument ε is labelled "Defeated," since δ has been labelled "Justified." The arguments α, γ, δ, ζ, and η (i.e., those labelled "Justified") together form the grounded extension of the framework. Every argumentation framework has a unique grounded extension. In the framework of Fig. 11.4, the grounded extension coincides with the unique preferred extension that is also the unique stable extension. The framework in Fig. 11.5 (left) shows that the grounded extension is not always a stable or preferred extension. Its grounded extension is here the empty set, but its two preferred and stable extensions are not empty.

[16] In establishing the concept, Verheij (1996b) used the term *admissible stage extensions*. The now standard term *semi-stable extension* was proposed by Caminada (2006).

[17] Dung's own definition of grounded extension, which does not use labelling, is not discussed here.