Argument

Henry Prakken

October 4, 2017

Abstract

This chapter discusses how formal models of argumentation can clarify philosophical problems and issues. Some of these arise in the field of epistemology, where it has been argued that the principles by which knowledge can be acquired are defeasible. Other problems and issues originate from the fields of informal logic and argumentation theory, where it has been argued that outside mathematics the standards for the validity of arguments are context-dependent and procedural, and that what matters is not the syntactic form but the persuasive force of an argument.

Formal models of argumentation are of two kinds. Argumentation logics formalise the idea that an argument only warrants its conclusion if it can be defended against counterarguments. Dialogue systems for argumentation regulate how dialogue participants can resolve a conflict of opinion. This chapter discusses how argumentation logics can define non-deductive consequence notions and how their embedding in dialogue systems for argumentation can account for the context-dependent and procedural nature of argument evaluation and for the dependence of an argument's persuasive force on the audience in an argumentation dialogue.

Keywords: argumentation theory; critical discussion; defeasible reasoning; formal dialectics; nonmonotonic logic; persuasion.

1 Introduction

Introductions to logic often portray logically valid inference as 'foolproof' reasoning: an argument is valid if the truth of its premises guarantees the truth of its conclusion. However, we all construct arguments from time to time that are not foolproof in this sense but that merely make their conclusion plausible when their premises are true. For example, if we are told that Peter, a professor in economics, says that reducing taxes increases productivity, we conclude that reducing taxes increases productivity since we know that experts are usually right within their domain of expertise. Sometimes such arguments are defeated by counterarguments. For example, if we are also told that Peter has political ambitions, we have to retract our previous conclusion that he is right about the effect of taxes if we also believe that people with political ambitions are often unreliable when it comes to taxes. Or, to use an example of practical instead of epistemic reasoning, if we accept that reducing taxes increases productivity and that increasing productivity is good, then we conclude that the taxes should be reduced, unless we also accept that reducing taxes increases inequality, that this is bad and that equality is more important than productivity. However, as long as such counterarguments are not available, we are happy to live with the conclusions of our fallible arguments. The question is: are we then reasoning fallaciously or is there still logic in our reasoning?

An answer to this question has been given in the development of argumentation logics. In a nutshell, the answer is that there is such logic but that it is inherently dialectic: an argument only warrants its conclusion if it is acceptable, and an argument is acceptable if, firstly, it is properly constructed and, secondly, it can be defended against counterarguments. Thus argumentation

logics must define three things: how arguments can be constructed, how they can be attacked by counterarguments and how they can be defended against such attacks.

Argumentation logics are a form of nonmonotonic logic, since their notion of warrant is nonmonotonic: new information may give rise to new counterarguments defeating arguments that were originally acceptable. Besides a logical side, argumentation also has a dialogical side: notions like argument, attack and defence naturally apply when (human or artificial) agents try to persuade each other to adopt or give up a certain point of view.

This chapter[1] aims to show how formal models of argumentation can clarify philosophical problems and issues. Some of these arise in the field of epistemology. Pollock (1974) argued that the principles by which knowledge can be acquired are defeasible. Later he made this precise in a formal system (Pollock, 1995), which inspired the development of argumentation logics in artificial intelligence (AI). Rescher (1977) also stressed the dialectical nature of theories of knowledge and presented a disputational model of scientific inquiry.

Other issues and problems originate from the fields of informal logic and argumentation theory. In 1958, Stephen Toulmin launched his influential attack on the logic research of those days, accusing it of only studying mathematical reasoning while ignoring other forms of reasoning, such as commonsense reasoning and legal reasoning (Toulmin, 1958). He argued that outside mathematics the standards for the validity of arguments are context-dependent and procedural: according to him an argument is valid if it has been properly defended in a dispute, and different fields can have different rules for when this is the case. Moreover, in his famous argument scheme he drew attention to the fact that different premises can have different roles in an argument (data, warrant or backing) and he noted the possibility of exceptions to rules (rebuttals). Perelman argued that arguments in ordinary discourse should not be evaluated in terms of their syntactic form but on their rhetorical potential to persuade an audience (Perelman and Olbrechts-Tyteca, 1969). These criticisms gave rise to the fields of informal logic and argumentation theory, which developed notions like argument schemes with critical questions and dialogue systems for argumentation. Many scholars in these fields distrusted or even rejected formal methods, but one point of this chapter is that formal methods can also clarify these aspects of reasoning. Another claim often made in these fields is that arguments can only be evaluated in the context of a dialogue or procedure. A second point of this paper is that this can be respected by embedding logical in dialogical accounts of argumentation.

The philosophical problems to be discussed in this chapter then are:

- Can argumentation-based standards for non-deductive inference be defined?
- To what extent are these standards procedural?
- To what extent are they context-dependent?
- What is the nature of argument schemes?
- Can the use of arguments to persuade be formalised?

2 Dung's abstract argumentation frameworks

In 1995 Phan Minh Dung introduced a now standard abstract formalism for argumentation-based inference, which assumes as input nothing but a set (of arguments) ordered by a binary relation (by Dung called 'attack' but in this chapter the term 'defeat' will be used).

Definition 2.1 An *abstract argumentation framework* (AF) is a pair $\langle \mathcal{A}, Def \rangle$, where $\mathcal{A}$ is a set arguments and $Def \subseteq \mathcal{A} \times \mathcal{A}$ is a binary relation of defeat. We say that an argument A defeats an argument B iff $(A, B) \in Def$, and that A strictly defeats B if A defeats B while B does not defeat A. A set S of arguments is said to defeat an argument A iff some argument in S defeats A.

[1] An earlier version of this chapter has appeared as Prakken (2011).

Dung (1995) defined four alternative semantics for AFs (over the years further semantics have been proposed; cf. Baroni et al. (2011)). A semantics for AFs characterises so-called argument extensions of AF's, that is, subsets of $\mathcal{A}$ that are in some sense coherent. One way to define extensions is with *labellings* of AFs, which assign to zero or more members of $Args$ either the label *in* or *out* (but not both) satisfying the following constraints:

1. an argument is *in* iff all arguments defeating it are *out*.
2. an argument is *out* iff it is defeated by an argument that is *in*.

Stable semantics labels all arguments, while *grounded semantics* minimises and *preferred semantics* maximises the set of arguments that are labelled *in*, and *complete semantics* allows all labellings satisfying the two constraints. Let $S \in \{\text{stable, preferred, grounded, complete}\}$ and (In, Out) an S-status assignment. Then In is defined to be an S-extension.[2]

Some known facts (also holding for the corresponding extensions) are that each grounded, preferred or stable labelling of an AF is also a complete labelling of that AF; the grounded labelling is unique but all other semantics allow for multiple labellings of an AF; each AF has a grounded and at least one preferred and complete labelling, but there are AFs without stable labellings; and the grounded labelling of an AF is contained in all other labellings of that AF.

Then the acceptability status of arguments can be defined as follows:

Definition 2.2 For grounded semantics an argument A is *justified* iff A is in the grounded extension; *overruled* iff A is not in the grounded extension but defeated by a member of the grounded extension; *defensible* otherwise. For stable and preferred semantics an argument A is *justified* iff A is in all stable/preferred extensions; *overruled* iff A is in no stable/preferred extension; *defensible* otherwise.

Figure 1 illustrates the definitions with some example argumentation frameworks, where defeat relations are graphically depicted as arrows.

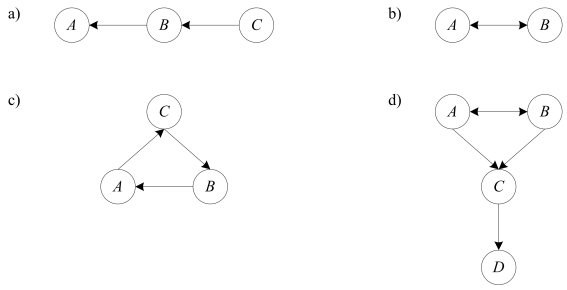

Figure 1: four argumentation frameworks

In AF (a) all semantics produce the same unique labelling. Argument C is *in* by constraint (1) since it has no defeaters, so B is *out* by constraint (2) since it is defeated by C, so A is *in* by constraint (1) since C defeats B. So all semantics produce the same, unique extension, namely, $\{A, C\}$. Hence in all semantics A and C are justified while B is overruled. It is sometimes said that C *reinstates*, or *defends* A by defeating its defeater B.

In AF (b) grounded semantics does not label any of the arguments while preferred and stable semantics produce two alternative labellings: one in which A is *in* and B is *out* and one in which B is *in* and A is *out*. Hence the grounded extension is empty while the preferred-and-stable extensions are $\{A\}$ and $\{B\}$. All these extensions are also complete. Hence in all semantics both A and B are defensible.

[2]This definition is different from but equivalent to Dung's (1995) definition of extensions.

AF (c) has no stable extensions since no argument can be labelled both *in* and *out* while there is a unique grounded, preferred and complete extension, which is empty, generated by a labelling which does not label any argument. Note that if a fourth argument D is added with no defeat relations with the other three arguments, there is still no stable extension while the unique grounded, preferred and complete extension is $\{D\}$.

Finally, AF (d) shows a difference between grounded and preferred semantics. The grounded extension is empty, since A and B can be left unlabelled so that C and D are also unlabelled, while the two preferred (and stable) extensions are $\{A, D\}$ and $\{B, D\}$. Thus while in grounded semantics all arguments are defensible, in preferred and stable semantics A and B are defensible, D is justified and C is overruled.

The above definitions characterise *sets* of arguments that are in some sense acceptable. In addition, procedures have been studied for determining whether a given argument is a member of such a set. Some take the form of an *argument game* between two players, a proponent and an opponent of an argument. The precise rules of the game depend on the semantics the game is meant to capture. The rules should be chosen such that the existence of a winning strategy (in the usual game-theoretic sense) for the proponent of an argument corresponds to the investigated semantic status of the argument, for example, 'justified in grounded semantics' or 'defensible in preferred semantics'.

Because of space limitations we can give only briefly one example game. The following game is sound and complete for grounded semantics in that the proponent of argument A has a winning strategy just in case A is in the grounded extension. The proponent starts a game with an argument and then the players take turns, trying to defeat the previous move of the other player. In doing so, the proponent must strictly defeat the opponent's arguments while he is not allowed to repeat his own arguments. A game is terminated if it cannot be extended with further moves. The player who moves last in a terminated game wins the game. Thus the proponent has a winning strategy if he has a way to make the opponent run out of moves (from the implicitly assumed AF) whatever choice the opponent makes.

As remarked in the introduction, argumentation logics must define three things: how arguments can be constructed, how they can be attacked and how they can be defended against attacks. Dung's abstract formalism only answers the third question. To answer the first two questions, accounts are needed of argument construction and the nature of attack and defeat. We next discuss a general framework for formulating such accounts.

3 An abstract framework for structured argumentation

The *ASPIC*$^+$ framework (Prakken, 2010; Modgil and Prakken, 2013, 2014) aims to integrate and further develop the main current formal models of structured argumentation. While some of its design choices can perhaps be debated, the framework is still representative of work in the field, for which reason we present it here. *ASPIC*$^+$ gives structure to Dung's arguments and defeat relation. It defines arguments as inference trees formed by applying strict ($\rightarrow$) or defeasible ($\Rightarrow$) inference rules to premises formulated in some logical language. Informally, if an inference rule's antecedents are accepted, then if the rule is strict, its consequent must be accepted *no matter what*, while if the rule is defeasible, its consequent must be accepted *if there are no good reasons not to accept it*. Arguments can be attacked on their 'ordinary' premises and on their applications of defeasible inference rules. Some attacks succeed as *defeats*; whether this is so is partly determined by preferences. The acceptability status of arguments is then defined by applying any of Dung (1995) semantics for abstract argumentation frameworks to the resulting set of arguments with its defeat relation.

ASPIC$^+$ is not a system but a framework for specifying systems. To start with, it defines the notion of an abstract *argumentation system* as a structure consisting of a logical language $\mathcal{L}$

with a negation symbol $\neg$[3], a set $\mathcal{R}$ consisting of two subsets $\mathcal{R}_s$ and $\mathcal{R}_d$ of strict and defeasible inference rules, and a naming convention n in $\mathcal{L}$ for defeasible rules in order to talk about the applicability of defeasible rules in $\mathcal{L}$. Thus, informally, $n(r)$ is a wff in $\mathcal{L}$ which says that rule $r \in \mathcal{R}$ is applicable. (as is usual, the inference rules in $\mathcal{R}$ are defined *over* the language $\mathcal{L}$ and are not elements *in* the language.)

Definition 3.1 An *argumentation system* is a triple $AS = (\mathcal{L}, \mathcal{R}, n)$ where:

- $\mathcal{L}$ is a logical language with a negation symbol $\neg$.
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules of the form $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$ and $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$ respectively (where φ_i, φ are meta-variables ranging over wff in $\mathcal{L}$), and $\mathcal{R}_s \cap \mathcal{R}_d = \varnothing$.
- $n : \mathcal{R}_d \longrightarrow \mathcal{L}$ is a naming convention for defeasible rules.

We write $\psi = -\varphi$ just in case $\psi = \neg\varphi$ or $\varphi = \neg\psi$ (we will sometimes informally say that formulas φ and $-\varphi$ are each other's negation).

Henceforth, a set $S \subseteq \mathcal{L}$ is said to be *directly consistent* iff $\nexists\ \psi, \varphi \in S$ such that $\psi = -\varphi$, otherwise S is *directly inconsistent*. And S is said to be *indirectly (in)consistent* if its closure under application of strict inference rules is directly (in)consistent.

Definition 3.2 A *knowledge base* in an $AS = (\mathcal{L}, \mathcal{R}, n)$ is a set $\mathcal{K} \subseteq \mathcal{L}$ consisting of two disjoint subsets $\mathcal{K}_n$ (the *axioms*) and $\mathcal{K}_p$ (the *ordinary premises*).

Intuitively, the axioms are certain knowledge and thus cannot be attacked, whereas the ordinary premises are uncertain and thus can be attacked.

Definition 3.3 An *argumentation theory* is a tuple AT = $(AS, \mathcal{K})$ where AS is an argumentation system and $\mathcal{K}$ is a knowledge base in AS.

ASPIC$^+$ arguments are now defined relative to an argumentation theory $AT = (AS, \mathcal{K})$, and chain applications of the inference rules from AS into inference graphs (which are trees if no premise is used more than once), starting with elements from the knowlege base $\mathcal{K}$. Arguments thus contain subarguments, which are the structures that support intermediate conclusions (plus the argument itself and its premises as limiting cases). In what follows, for a given argument the function `Prem` returns all its premises, `Conc` returns its conclusion, `Sub` returns all its subarguments, `DefRules` returns all defeasible rules of an argument and `TopRule` returns the final rule applied in the argument.

Definition 3.4 An *argument* A on the basis of an argumentation theory with a knowledge base $\mathcal{K}$ and an argumentation system $(\mathcal{L}, \mathcal{R}, n)$ is any structure obtainable by applying one or more of the following steps finitely many times:

1. φ if $\varphi \in \mathcal{K}$ with $\texttt{Prem}(A) = \{\varphi\}$; $\texttt{Conc}(A) = \varphi$; $\texttt{Sub}(A) = \{\varphi\}$; $\texttt{DefRules}(A) = \varnothing$; $\texttt{TopRule}(A)$ = undefined.

2. $A_1, \ldots A_n \rightarrow\!\!/\Rightarrow \psi$[4] if $A_1, \ldots, A_n$ are arguments such that there exists a strict/defeasible rule $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow\!\!/\Rightarrow \psi$ in $\mathcal{R}_s$/$\mathcal{R}_d$.
   $\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$,
   $\texttt{Conc}(A) = \psi$,
   $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$.
   $\texttt{DefRules}(A)$ = $\texttt{DefRules}(A_1) \cup \ldots \cup \texttt{DefRules}(A_n)$;
   $\texttt{TopRule}(A)$ = $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow\!\!/\Rightarrow \psi$.

[3] In most papers on *ASPIC*$^+$ negation can be non-symmetric. In this paper we present the special case with symmetric negation.

[4] $\rightarrow\!\!/\Rightarrow$ means that the rule is a strict, respectively, defeasible rule.

Then A is: *strict* if $\texttt{DefRules}(A) = \varnothing$; *defeasible* if $\texttt{DefRules}(A) \neq \varnothing$; *firm* if $\texttt{Prem}(A) \subseteq \mathcal{K}_n$; *plausible* if $\texttt{Prem}(A) \subseteq \mathcal{K}_p$.

Example 3.5 Consider a knowledge base in an argumentation system with $\mathcal{R}_s = \{p, q \rightarrow s;\ u, v \rightarrow w\}$; $\mathcal{R}_d = \{p \Rightarrow t;\ s, r, t \Rightarrow v\}$; $\mathcal{K}_n = \{q\}$; $\mathcal{K}_p = \{p, r, u\}$. An argument for w is displayed in Figure 2. The type of a premise is indicated with a superscript and defeasible inferences and attackable premises and conclusions are displayed with dotted lines. Formally

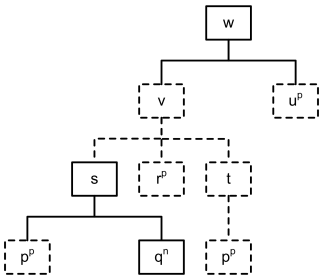

Figure 2: An argument

the argument and its subarguments are written as follows:

A_1: p	A_5: $A_1 \Rightarrow t$
A_2: q	A_6: $A_1, A_2 \rightarrow s$
A_3: r	A_7: $A_5, A_3, A_6 \Rightarrow v$
A_4: u	A_8: $A_7, A_4 \rightarrow w$

We have that

$\texttt{Prem}(A_8) =$	$\{p, q, r, u\}$
$\texttt{Conc}(A_8) =$	w
$\texttt{Sub}(A_8) =$	$\{A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8\}$
$\texttt{DefRules}(A_8) =$	$\{p \Rightarrow t;\ s, r, t \Rightarrow v\}$
$\texttt{TopRule}(A_8) =$	$u, v \rightarrow w$

Arguments can be attacked in three ways: on their premises (undermining attack), on their conclusion (rebutting attack) or on an inference step (undercutting attack). The latter two are only possible on applications of defeasible inference rules.

Definition 3.6 *A attacks B* iff *A undercuts*, *rebuts* or *undermines* B, where:
• *A undercuts* argument B (on B') iff $\texttt{Conc}(A) = -n(r)$ for some $B' \in \texttt{Sub}(B)$ such that B''s top rule r is defeasible.
• *A rebuts* argument B (on B') iff $\texttt{Conc}(A) = -\varphi$ for some $B' \in \texttt{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \varphi$.
• Argument *A undermines* B (on B') iff $\texttt{Conc}(A) = -\varphi$ for some $B' = \varphi$, $\varphi \in \mathcal{K}_p$.

In Example 3.5 argument A_8 can be undercut on two of its subarguments, namely, A_5 and A_7. An undercutter of A_5 must have a conclusion $-\varphi$ where $n(p \Rightarrow t) = \varphi$ while an undercutter of A_5 must have a conclusion $-\varphi$ where $n(s,\ r,\ t \Rightarrow w) = \varphi$. Argument A_8 can be rebutted on A_5 with an argument for $-t$ and on A_7 with an argument for $-v$. Moreover, if the rebuttal of A_5 has a defeasible top rule, then A_5 in turn rebuts the argument for $-t$. However, A_8 itself

does not rebut that argument, except in the special case where $w = - - t$. Finally, argument A_8 can be undermined with an argument that has conclusion $-p$, $-r$ or $-u$.

Attack relations between arguments can be resolved with an ordering on arguments. To formalise this, the notion of a structured argumentation framework is introduced.

Definition 3.7 Let AT be an *argumentation theory* (AS, KB). A *structured argumentation framework (*SAF*)* defined by AT is a triple $\langle \mathcal{A}, Att, \preceq \rangle$ where

- $\mathcal{A}$ is the all arguments on the basis of AT;
- $\preceq$ is an ordering on $\mathcal{A}$;
- $(X, Y) \in Att$ iff X attacks Y.

Modgil and Prakken (2013) also study a variant of this definition in which arguments are required to have indirectly consistent premises.

Now attacks combined with the argument ordering yield three kinds of defeat. For undercutting attack no preferences are needed to make it succeed, since undercutters are explicit exceptions to the rule they undercut. Rebutting and undermining attacks succeed only if the attacked argument is not stronger than the attacking argument.

Definition 3.8 *A defeats B* iff: A undercuts B, or; A rebuts/undermines B on B' and $A \not\prec B'$.[5] A strictly *defeats* B iff A *defeats* B and B does not *defeat* A

The success of rebutting and undermining attacks thus involves comparing the conflicting arguments at the points where they conflict. The definition of successful undermining exploits the fact that an argument premise is also a subargument.

The *ASPIC*$^+$ framework assumes the argument ordering as given. It may depend on all sorts of standards, such as statistical strength of generalisations, reliability of information sources, preferences over outcomes of actions, or norm hierarchies. In many contexts such standards can themselves be argued about. One way to formalise this is by using Modgil's (2009) idea to decompose the defeat relation of Dung (1995)'s abstract argumentation frameworks into a more basic *attack* relation and to allow *attacks on attacks* in addition to attacks on arguments. Combined with *ASPIC*$^+$, the idea is that if argument C claims that argument B is preferred to argument A, and A attacks B, then C undermines the success of A's attack on B (i.e., A does not *defeat* B) by pref-attacking A's attack on B.

Recall that argumentation logics must define three things: how arguments can be constructed, how they can be defeated and how they can be defended against defeating counterarguments. While Dung's abstract argumentation semantics addresses the last issue, we can now combine it with the *ASPIC*$^+$ framework to address the first two issues.

Definition 3.9 An *abstract argumentation framework (AF) corresponding to a* SAF = $\langle \mathcal{A}, Att, \preceq \rangle$ is a pair $(\mathcal{A}, Def)$ such that Def is the defeat relation on $\mathcal{A}$ determined by $\langle \mathcal{A}, Att, \preceq \rangle$.

The justified arguments of the above defined AF are then defined under various semantics, as in Definition 2.2. We now see that an argument can be defended against attacks in two ways: by showing that the attacker is inferior to it or by defeating the attacker with a counterattack that reinstates the original argument.

We can now finally define an argumentation-based consequence notion for well-formed formulas (relative to an AT and with respect to any given semantics):

Definition 3.10 A wff $\varphi \in \mathcal{L}$ is *justified* if φ is the conclusion of a justified argument, and *defensible* if φ is not justified and is the conclusion of a defensible argument.

An alternative definition of a justified wff is to say that every extension contains an argument with the wff as its conclusion. Unlike the above definition, this alternative definition allows

[5] $X \prec Y$ means as ususal that $X \preceq Y$ and $Y \not\preceq X$.

different extensions containing different arguments for a justified conclusion. This is similar to the different treatments that semantics for abstract argumentation give to Figure 1d.

One possible analysis of this difference is that some semantics, or some definitions of justification, are better than others, but an alternative analysis is that different definitions capture different senses or strengths of justification, which each may have their use in certain contexts. For example, in the law, criminal cases require higher proof standards than civil cases. And while in domains like the law and medicine defeasible arguments are acceptable, in the field of mathematics all arguments must, of course, be deductive. Thus we see how our formal framework for argumentation can make sense of Toulmin's claim that the standards for the validity of arguments are context-dependent.

In addition, the kind of reasoning can be relevant, such as the distinction between epistemic and practical reasoning. If, for instance, two incompatible actions (say reducing and increasing taxes) have two different good consequences (say increasing productivity and increasing equality in society) and there is no reason to prefer one consequence over the other, then an arbitrary choice is (all other things being equal) rational. If, on the other hand, two experts disagree about whether reducing taxes increases productivity, then an arbitrary choice for one of them seems irrational. So it might be argued that in practical reasoning a defensible conclusion can be good enough while in epistemic reasoning we should aim for justified conclusions.

4 The nature of inference rules

While we now have a general framework for the definition of argumentation logics, much more can be said. To start with, the framework can be instantiated in many ways, so there is a need for principles that can be used in assessing the quality of instantiations. Caminada and Amgoud (2007) formulated several so-called rationality postulates, namely, that each extension should be closed under subarguments and under strict rule application, and be directly and indirectly consistent. *ASPIC*$^+$ unconditionally satisfies the two closure postulates while Prakken (2010) and Modgil and Prakken (2013) identify conditions under which some broad classes of instantiations satisfy the two consistency postulates.

The next question is, what are 'good' collections of strict and defeasible inference rules? In AI there is a tradition to let inference rules express domain-specific information, such as *all penguins are birds* or *birds typically fly*. This runs counter to the usual practice in logic, in which inference rules express general patterns of reasoning, such as modus ponens, universal instantiation and so on. This practice is also followed in systems for so-called classical argumentation (Besnard and Hunter, 2008), in which arguments from a possibly inconsistent knowledge base are classical proofs from consistent subsets of the knowledge base. These systems are in fact a special case of the *ASPIC*$^+$ framework with $\mathcal{L}$ being the language of standard propositional or first-order logic, the strict rules being all valid propositional or first-order inferences, with no defeasible rules and no axiom premises, and with the premises of all arguments required to be indirectly consistent. In this approach (which can be generalised to other deductive logics) arguments can thus only be sensibly attacked on their premises.

While this approach has some merits, it is doubtful whether all argumentation can be reduced to inconsistency handling in some deductive logic. In particular John Pollock strongly emphasized the importance of *defeasible* reasons in argumentation. He was quite insistent that defeasible reasoning is not just some exotic, exceptional, add-on to deductive reasoning but is, instead, an essential ingredient of our cognitive life:

> ... we cannot get around in the world just reasoning deductively from our prior beliefs together with new perceptual input. This is obvious when we look at the varieties of reasoning we actually employ. We tend to trust perception, assuming that things are the way they appear to us, even though we know that sometimes they are not. And we tend to assume that facts we have learned perceptually will

> remain true, as least for a while, when we are no longer perceiving them, but of course, they might not. And, importantly, we combine our individual observations inductively to form beliefs about both statistical and exceptionless generalizations. None of this reasoning is deductively valid. (Pollock, 2009, p. 173)

Here the philosophical distinction between *plausible* and *defeasible* reasoning is relevant; see Rescher (1976, 1977) and Vreeswijk (1993, Ch. 8). Plausible reasoning is valid deductive reasoning from an uncertain basis while defeasible reasoning is deductively invalid (but still rational) reasoning from a solid basis. In these terms, models of deductive argumentation formalize plausible reasoning, while Pollock modeled defeasible reasoning and the *ASPIC*$^+$ framework gives a unified account of these two kinds of reasoning.

There is also semantic support for the idea of defeasible inference rules. Consider, for example, the statistical generalisation *men usually have no beard*. Concluding from this that *people with a beard are usually not men* is a so-called 'base rate fallacy' (Tversky and Kahneman, 1974). If (epistemic) defeasible reasoning is reduced to inconsistency handling in deductive logic, such fallacies are easily committed. Likewise, it has been argued that reasons of practical and normative reasoning are inherently defeasible; cf. e.g. Raz (1975).

While the case for defeasible inference rules thus seems convincing, the question remains what are 'good' defeasible inference rules, especially if they are to express general patterns of inference. Here two bodies of philosophical work are relevant, namely, Pollock's (1974; 1995) notion of *defeasible reasons* and argumentation-theory's notion of *argument schemes* (Walton et al., 2008). Pollock's defeasible reasons are general patterns of epistemic defeasible reasoning. He formalised reasons for perception, memory, induction, temporal persistence and the statistical syllogism, as well as undercutters for these reasons. In the *ASPIC*$^+$ framework Pollock's defeasible reasons can be expressed as schemes (in the logical sense, with metavariables ranging over $\mathcal{L}$) for defeasible inference rules. The same analysis applies to argument schemes, which are stereotypical non-deductive patterns of reasoning. Uses of argument schemes are evaluated in terms of critical questions specific to the scheme. In the literature on argumentation theory many collections of argument schemes have been proposed, both for epistemic, practical and evaluative reasoning. An example of an epistemic argument scheme is the scheme from expert opinion (Walton et al., 2008, p. 310):

> E is an expert in domain D, E asserts that P is true, P is within D, therefore presumably P is true

Walton et al. (2008) give this scheme six critical questions: (1) Is E credible as an expert source? (2) Is E an expert in domain D? (3) What did E assert that implies P? (4) Is E personally reliable as a source? (5) Is P consistent with what other experts assert? (6) Is E's assertion of P based on evidence?

A practical argument scheme is the scheme from good (bad) consequences (here in a formulation that deviates from Walton et al. (2008) to stress its abductive nature):

> Action A results in P, P is good (bad), therefore all other things being equal A should (not) be done.

This scheme is usually given two critical questions: (1) Does A result in P? (2) Does A also result in something which is bad (good)? (3) (When P is concluded to be good) Is there another way to realise P?

In *ASPIC*$^+$, argument schemes can also be formalised as schemes for defeasible inference rules; then critical questions are pointers to counterarguments. In the scheme from expert opinion questions (2) and (3) point to underminers (of, respectively, the first and second premise), questions (4), (1) and (6) point to undercutters (the exceptions that the expert is biased or incredible for other reasons and that he makes scientifically unfounded statements) while question (5) points to rebutting applications of the expert opinion scheme. In the scheme from good (bad)

consequences question (1) points to underminers of the first premise, question (2) points to rebuttals using the opposite version of the scheme while question (3) points to undercutters.

This account of argument schemes can also clarify Toulmin's (1958) distinction between *warrants* (rule-like premises) and *backings* of warrants. For example, a warrant can be that smoking causes cancer while its backing can be an expert opinion: then the defeasible inference rule expressing the scheme from expert opinion allows to infer the warrant from the backing.

Let us illustrate the just-proposed modelling of defeasible reasons and argument schemes with an example. The logical language $\mathcal{L}$ is informally assumed to be a first-order language augmented with a conditional for defeasible generalisations, $\mathcal{R}_s$ consists of all deductively valid inferences over $\mathcal{L}$ and $\mathcal{R}_d$ consists of the above schemes from expert opinion (e) and from good (gc) and bad (bc) consequences, plus a modus ponens scheme (dmp) for defeasible generalisations. Consider then the following arguments (where premise arguments are assumed to be in $\mathcal{K}_p$ and defeasible inferences are labelled with the inference rule they apply).

A_1: *P says "lowering taxes increases productivity"*
A_2: *P is an expert in economics*
A_3: *"lowering ... productivity" is about economics*
A_4: $A_1, A_2, A_3 \Rightarrow_e$ *lowering ... productivity*
A_5: *Increased productivity is good*
A_6: $A_4, A_5 \Rightarrow_{gc}$ *taxes should be lowered*

B_1: *lowering taxes increases inequality*
B_2: *Increased inequality is bad*
B_3: $B_1, B_2 \Rightarrow_{bc}$ *taxes should not be lowered*

C_1: *P has political ambitions*
C_2: *people with political ambitions are usually not reliable about taxes*
C_3: $C_1, C_2 \Rightarrow_{dmp}$ *P is not reliable about taxes*
C_4: *Rule e does not apply to unreliable people*
C_5: $C_3, C_4 \rightarrow$ *Rule e does not apply to P*

D_1: *P is never on TV*
D_2: *people who are never on TV usually have no political ambitions*
D_3: $D_1, D_2 \Rightarrow_{dmp}$ *P has no political ambitions*

Arguments A_6 and B_3 rebut each other. Assume $B_3 \prec A_6$ so A_6 strictly defeats B_3. Assuming the obvious naming convention, argument C_5 undercuts A_6 on A_4 and so defeats both, while D_3 undermines C_5 on C_1 and C_1 in turn rebuts D_3. At this point we know that all unattacked premise arguments are justified in any semantics, since they have no defeaters. For the remaining arguments, suppose first $D_3 \prec C_1$. Then C_1 strictly defeats D_3, so in any semantics D_3, A_4 and A_6 are overruled, while all C_i and B_3 are justified. Suppose next $C_1 \prec D_3$. Then D_3 strictly defeats C_3 and C_5 by strictly defeating C_1, so in any semantics D_3 and all A_i are justified, while C_1, C_3, C_5 and B_3 are overruled. Suppose finally that neither $C_1 \prec D_3$ nor $D_3 \prec C_1$. Then C_1 and D_3 defeat each other so, even though D_3 still strictly defeats C_3 and C_5, in any semantics all non-premise arguments plus C_1 are defensible.

5 Argumentation as a form of dialogue

As stated in the introduction, argumentation theorists often claim that arguments can only be evaluated in the context of a dialogue or procedure. More specifically, Walton (1996) regards argument schemes as dialogical devices, determining dialectical obligations and burdens of proof. An argument is a move in a dialogue and the scheme that it instantiates determines the allowed and required responses to that move. At first sight, our account of argument schemes as defeasible inference rules would seem to be incompatible with Walton's dialogical account. However, these two accounts can be reconciled by embedding argumentation logics in dialogue systems for argumentation.

While argumentation logics define notions of consequence from a given body of information, dialogue systems for argumentation (Walton and Krabbe, 1995) regulate disputes between real agents, who each have their own body of information, and who may be willing to learn from each other so that their information state may change. Moreover, during the dialogue they

may construct a joint theory on the issue in dispute, which also evolves over time. Essentially, dialogue systems define a communication language (the well-formed utterances) and a protocol (when a well-formed utterance may be made and when the dialogue terminates).

Consider the following simple example, with a dialogue system that allows players to move arguments and to challenge, concede or retract premises and conclusions of these arguments. Each challenge must be answered with a ground for the challenged statement or else the statement must be retracted. The two agents have their own knowledge base but a shared *ASPIC*$^+$ argumentation system with a propositional language and three defeasible inference rules: $p \Rightarrow q$, $r \Rightarrow p$ and $s \Rightarrow \neg r$. Paul's and Olga's knowledge bases contain, respectively, single ordinary premises p and r. Let us assume that all arguments are of equal preference. Paul wants to persuade Olga that q is the case. He can internally construct the following argument for q: A_1: r, A_2: $A_1 \Rightarrow p$, A_3: $A_2 \Rightarrow q$. However, a well-known argumentation heuristic says that arguments in dialogue should be made as sparse as possible in order to avoid attacks. Therefore, Paul only utters the last step in the argument, hoping that Olga will accept p so that Paul does not have to defend r. This leads to the following dialogue.

P_1:	q *since* p	O_1:	*why* p
P_2:	p *since* r	O_2:	$\neg r$ *since* s
P_3:	*retract* r, *retract* q		

What has happened here? If Olga had been a trusting person who concedes a statement if she cannot construct an argument for the opposite, then she would have conceded p and q after P_1. But q is not a justified conclusion from the joint knowledge bases, so this outcome is undesirable. In fact, Olga was less trusting and first asked Paul for his reasons for p. Since Paul was honest, he gave his true reasons, which allowed Olga to discover that she could attack Paul with an undermining counterargument. Paul could not defend himself against this attack, so he realised that he cannot persuade Olga that q is true; he therefore retracted r and q.

Argumentation logic applies here in several ways. It can model the agents' internal reasoning but it can also be applied at each dialogue stage to the joint theory that the agents have created at that stage. For example, after O_2 the logic says that q is overruled on the basis of $\mathcal{K}_n = \varnothing, \mathcal{K}_p = \{p, r, s\}$ while after P_4 the logic says that no argument for q can be constructed on the basis of $\mathcal{K}_n = \varnothing, \mathcal{K}_p = \{p, s\}$. Argumentation logic can also be used as a component of notions of soundness and completeness of protocols, such as:

- A protocol is *sound* if whenever at termination p is accepted, p is justified by the participants' joint knowledge bases.
- A protocol is *weakly* complete if whenever p is justified by the participants' joint knowledge bases, there is a legal dialogue at which at termination p is accepted.
- A protocol is *strongly* complete if whenever p is justified by the participants' joint knowledge bases, all legal dialogues terminate with acceptance of p.

These notions can also be defined relative to the joint theory constructed during a dialogue, or made conditional on particular agent strategies and heuristics (for example, a protocol could be sound and complete on the condition that all agents are honest but not trusting).

We can now without giving up the idea of an argumentation logic make sense of the claim that arguments should be evaluated in the context of a dialogue or procedure. The dialogue provides the relevant statements and arguments at each stage of the dialogue. The logic then determines the justified arguments at that stage. The logic also points at the importance of investigation. Since arguments can be defeated by counterarguments, the search for information that gives rise to counterarguments is an essential part of testing an argument's viability: the more thorough this search has been, the more confident we can be that an argument is justified if we cannot find defeaters. The ultimate justification of an argument is then determined by applying the logic to the final information state. Thus the ultimate justification of an argument depends on both logic and dialogue, or more generally on both logic and investigation.

On this account the critical questions of argument schemes have a dual role. On the one hand they define possible counterarguments to arguments constructed with the scheme (logic) while on the other hand they point at investigations that could be done to find such counterarguments (dialogue and procedure). This account also gives a further explanation why argument evaluation is context dependent, since different contexts may require different protocols for dialogue: when a decision has to be reached in reasonable time (as in a business meeting), a protocol may be more restrictive than in settings like academic debate. For example, the right to give alternative replies to a move may be restricted so that agents are forced to think what is their best reply.

Finally, on this account persuasiveness of arguments can be modelled as follows. Each dialogical agent has an internal argumentation theory and evaluates incoming arguments in terms of how they fit with the AF that it can internally generate. Given an *acceptance attitude* the agent will either accept the argument's premises and/or conclusion, or attack it with a counterargument, or ask for further grounds for a premise. Personality models can help modelling which types of arguments an agent of a certain type tends to accept. This gives a third way in which argument evaluation is context-dependent: the persuasive force of an argument depends on the listener. Current work of this kind is still preliminary but fascinating and promising (see e.g. the proceedings of the annual *ArgMas* workshops on argumentation in multi-agent systems). In fact this work provides a formal or even computational account of Perelman's New Rhetoric (Perelman and Olbrechts-Tyteca, 1969).

6 Conclusion

In this chapter we discussed five philosophical problems concerning argumentation. We first showed how argumentation-based standards for non-deductive inference can be defined, by presenting an abstract framework for argument evaluation given a set of arguments and their attack and defeat relations, and by supplementing it with accounts of argument construction and the nature of attack and defeat. We then clarified how a dialogical account of argument evaluation can be given in formal terms, by discussing the embedding of argumentation logics in dialogue systems for argumentation. This embedding also clarified the nature of argument schemes: argument schemes can be seen as defeasible inference rules and their critical questions as pointers to counterarguments. We also clarified how the use of arguments to persuade can be formalised, by adding the notions of argumentation strategies and heuristics and suggesting the use of personality models of argumentative agents. Finally, we gave several reasons why argument evaluation is context-dependent: different domains may have different sets of argument schemes, different contexts may require more or less strict semantics and/or protocols for dialogue and the persuasive force of arguments may depend on the listener.

References

*** Baroni, P., Caminada, M. and Giacomin, M. (2011). An introduction to argumentation semantics, *The Knowledge Engineering Review* **26**: 365–410. [A comprehensive survey of semantics for abstract argumentation.]

Besnard, P. and Hunter, A. (2008). *Elements of Argumentation*, MIT Press, Cambridge, MA.

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

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

*** Hunter, A. (ed.) (2014). *Argument and Computation*, Vol. 5. Special issue with Tutorials on Structured Argumentation. [Contains tutorial introductions to four alternative accounts of structured argumentation.]

Modgil, S. (2009). Reasoning about preferences in argumentation frameworks, *Artificial Intelligence* **173**: 901–934.

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

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

Perelman, C. and Olbrechts-Tyteca, L. (1969). *The New Rhetoric. A Treatise on Argumentation*, University of Notre Dame Press, Notre Dame, Indiana.

Pollock, J. (1974). *Knowledge and Justification*, Princeton University Press, Princeton.

*** Pollock, J. (1995). *Cognitive Carpentry. A Blueprint for How to Build a Person*, MIT Press, Cambridge, MA. [A classic philosophical account of defeasible argumentation.]

Pollock, J. (2009). A recursive semantics for defeasible reasoning, *in* I. Rahwan and G. Simari (eds), *Argumentation in Artificial Intelligence*, Springer, Berlin, pp. 173–197.

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

Prakken, H. (2011). An overview of formal models of argumentation and their application in philosophy, *Studies in Logic* **4**: 65–86.

*** Prakken, H. (2017). Historical overview of formal argumentation, *in* P. Baroni and D. Gabbay and M. Giacomin and L. van der Torre (eds), *Handbook of Formal Argumentation*, Vol. 1, College Publications, London.

*** Prakken, H. and Vreeswijk, G. (2002). Logics for defeasible argumentation, *in* D. Gabbay and F. Günthner (eds), *Handbook of Philosophical Logic*, second edn, Vol. 4, Kluwer Academic Publishers, Dordrecht/Boston/London, pp. 219–318. [A systematic, although somewhat outdated introduction to argumentation logics.]

*** Rahwan, I. and Simari, G. (eds) (2009). *Argumentation in Artificial Intelligence*, Springer, Berlin. [A collection of survey papers on all aspects of formal and computational argumentation.]

Raz, J. (1975). *Practical Reason and Norms*, Princeton University Press, Princeton.

Rescher, N. (1976). *Plausible Reasoning*, Van Gorcum, Assen.

Rescher, N. (1977). *Dialectics: a Controversy-oriented Approach to the Theory of Knowledge*, State University of New York Press, Albany, N.Y.

Toulmin, S. (1958). *The Uses of Argument*, Cambridge University Press, Cambridge.

Tversky, A. and Kahneman, D. (1974). Judgement under uncertainty: heuristics and biases, *Science* **185**: 1124–1131.

Vreeswijk, G. (1993). *Studies in Defeasible Argumentation*, Doctoral dissertation Free University Amsterdam.

Walton, D. (1996). *Argumentation Schemes for Presumptive Reasoning*, Lawrence Erlbaum Associates, Mahwah, NJ.

Walton, D. and Krabbe, E. (1995). *Commitment in Dialogue. Basic Concepts of Interpersonal Reasoning*, State University of New York Press, Albany, NY.

Walton, D., Reed, C. and Macagno, F. (2008). *Argumentation Schemes*, Cambridge University Press, Cambridge.