Argumentation and Artificial Intelligence 11

Contents

F.H. van Eemeren et al., *Handbook of Argumentation Theory*,
DOI 10.1007/ 978-90-481-9473-5_11,

11.1 Research on Argumentation in Artificial Intelligence

The study of artificial intelligence (AI) is in many ways connected with the study of argumentation. Though both fields have developed separately, the last 20 years have witnessed an increase of mutual influence and exchange of ideas. From this development, both fields stand to profit: argumentation theory providing a rich source of ideas that may be used in the computerization of theoretical and practical reasoning and of argumentative interaction, and artificial intelligence providing the systems for testing these ideas. In fact, combining argumentation theory with AI offers argumentation theory a laboratory for examining implementations of its rules and concepts.

By their interdisciplinary nature, approaches to argumentation in AI integrate insights from different perspectives (see Fig. 11.1). In the theoretical systems perspective, the focus is on theoretical and formal models of argumentation, for instance, extending the long tradition of philosophical and formal logic. In the artificial systems perspective, the aim is to build computer programs that model or support argumentative tasks, for instance, in online dialogue games or in expert systems (computer programs that reproduce the reasoning of a professional expert, e.g., in the law or in medicine). The natural systems perspective helps to ground research by concentrating on argumentation in its natural form, for instance, in the human mind or in an actual debate.

Since the 1990s, the main areas of AI that have been of interest for argumentation theory are those of defeasible reasoning, multi-agent systems, and models of legal argumentation. A great many articles about these overlapping areas have appeared in journals in the realm of computation.[1] The biennial COMMA conference series focuses on the study of computational models of argument.[2]

[1] We mention a few of these journals: *Artificial Intelligence*, *Artificial Intelligence and Law*, *Autonomous Agents and Multi-Agent Systems*, *Computational Intelligence*, *International Journal of Cooperative Information Systems*, *International Journal of Human-Computer Studies*, *Journal of Logic and Computation*, and *The Knowledge Engineering Review*. Contributions have also been made to journals that deal primarily with argumentation, such as *Argumentation* and *Informal Logic*. A journal devoted explicitly to the interdisciplinary area of AI is *Argument and Computation*.

[2] The first COMMA conference was held in Liverpool in 2006, followed by conferences in Toulouse (2008), Desenzano del Garda (2010), and Vienna (2012). See http://www.comma-conf.org/. ArgMAS (Argumentation in Multi-Agent Systems) and CMNA (Computational Models of Natural Argument) are related workshops.

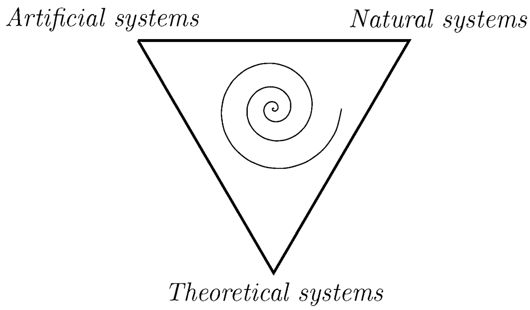


Fig. 11.1 Perspectives on argumentation

The impact of argumentation studies in the field of AI is illustrated by the fact that many of the best cited articles in the authoritative journal *Artificial Intelligence* are about argumentation.[3]

Research on argumentation in the field of AI often emphasizes formal and computational detail, sometimes making the papers concerned hardly accessible to a less formally or computationally oriented audience. In an attempt to disseminate AI's contribution to argumentation research, in this chapter, the focus is on the core ideas. This fits the feeling that progress in argumentation research can be accelerated by a cross-fertilization of ideas from different origins (see Fig. 11.1).

This chapter aims to be a starting point for the study of the contribution of artificial intelligence to argumentation research in general. As said, the focus is on the presentation of key ideas in the field, not on a representative description of all contributions by all contributors. The sheer scope and rapid growth of the field would make the latter impossible anyway.[4]

The first two sections that follow trace the historical roots of argumentation research in artificial intelligence, discussing work on non-monotonic logic (Sect. 11.2) and on defeasible reasoning (Sect. 11.3). Then follow a number of foundational topics in Sects. 11.4, 11.5, 11.6, and 11.7 on abstract argumentation, arguments with structure, argument schemes, and argumentation dialogues. In the Sects. 11.8, 11.9, 11.10, 11.11, 11.12 and 11.13 of this chapter, a number of specific topics are addressed that have been studied in AI approaches to argumentation: reasoning with rules, case-based reasoning, values and audiences, argumentation support software, burden of proof, evidence and argument strength, and applications and case studies.

[3] Nine of the top twenty best cited articles in Artificial Intelligence since 2007 deal with argumentation, five of the top ten, and three of the top five. Source: Scopus.com, June 2012.

[4] For a survey of the literature up till approximately 2002, we refer to the road map by Reed and Norman (2004a) and the more formally oriented overview by Prakken and Vreeswijk (2002). For more detail, including formal and computational elaboration, the interested reader may wish to consult the original sources referred to in this chapter. In addition, we refer to the collection of papers edited by Rahwan and Simari (Eds., 2009), which contains contributions by a great many researchers in the field of argumentation and artificial intelligence, and to the sources we mentioned in Notes 1 and 2. See also the special issue of the *Artificial Intelligence* journal edited by Bench-Capon and Dunne (2007).

11.2 Non-monotonic Logic

Today many artificial intelligence publications directly address issues related to argumentation. A relevant development predating such contemporary work is the area of *non-monotonic logic*.[5] A logic is non-monotonic when a conclusion that, according to the logic, follows from given premises need not also follow when premises are added. In contrast, classical logic is monotonic. For instance, in a standard classical analysis, from premises "Edith goes to Vienna or Rome" and "Edith does not go to Rome," it follows that "Edith goes to Vienna," irrespective of possible additional premises. In a non-monotonic logic, it is possible to draw tentative conclusions, while keeping open the possibility that additional information may lead to the retraction of such conclusions. The standard example of non-monotonicity used in the literature of the 1980s concerns the flying of birds. Typically, birds fly, so if you hear about a bird, you will conclude that it can fly.

11.2.1 Reiter's Logic for Default Reasoning

A prominent proposal in non-monotonic logic is Raymond Reiter's (1980) logic for default reasoning. In his system, non-monotonic inference steps are applications of a set of given *default rules*. Reiter's first example of a default rule expresses that birds typically fly:

$$\mathrm{BIRD}(x) : \mathrm{M\ FLY}(x)/\mathrm{FLY}(x)$$

Here the M should be read as "it is consistent to assume." The default rule expresses that if x is a bird, and it is consistent to assume that x can fly, then by default one can conclude that x can fly. One can then add exceptions, for instance, using this expression in classical logic that if x is a penguin, x cannot fly:

$$\mathrm{PENGUIN}(x) \rightarrow \neg\mathrm{FLY}(x)$$

The general default rule can be applied to a specific bird, by instantiating the variable x by an instance t. In this situation, from just the premise BIRD(t), one can conclude (by default) FLY(t), but when one has a second premise PENGUIN(t), the conclusion FLY(t) does not follow.

A more general form of a default rule is α : M β/γ, where the element α is the *prerequisite* of the rule, β the *justification*, and γ the *consequent*. A special case occurs when the justification and consequent coincide, as in the bird example above; then we speak of a "normal default rule."

[5] See the entry on nonmonotonic logic in the Stanford Encyclopedia of Philosophy at http://plato.stanford.edu/entries/logic-nonmonotonic/ (Antonelli 2010).

Other influential logical systems for non-monotonic reasoning include circumscription, auto-epistemic logic, and non-monotonic inheritance; each of them is discussed in the representative overview of the study of non-monotonic logic at its heyday by Gabbay et al. (1994).

11.2.2 Logic Programming

The general idea underlying logic programming is that a computer can be programmed using logical techniques. In this view, computer programs are considered not only procedurally as recipes for how to achieve the program's aims but also declaratively, in the sense that the program can be read like a text, for instance, as the rule-like knowledge needed to answer a question. In the logic programming language Prolog (the result of a collaboration between Colmerauer and Kowalski; see Kowalski 2011), examples of some facts and rules are the following (Bratko 2001):

```
parent(pam, bob)
parent(tom, bob)
parent(bob, pat)
female(pam)
female(pat)
male(bob)
male(tom)
mother(X, Y) :- parent(X, Y), female(X)
grandparent(X, Z) :- parent(X, Y), parent(Y, Z)
```

This small logic program represents (among other things) the facts that Pam and Tom are Bob's parents and that Pam is female and Tom male. It also represents the rules that someone's mother is a female parent and that a grandparent is the parent of a parent. Given this Prolog program, a computer can as expected derive that Pam is Bob's mother and that Pam is Pat's grandparent. Interaction with a Prolog program usually takes the form of a dialogue, where the user asks the program a question. For instance, the question whether Pam is Pat's grandparent takes this form:

```
?- grandparent(pam, pat)
```

which will be answered "Yes."

In the interpretation of logic programs, the closed world assumption plays a key role: a logic program is assumed to describe all facts and rules about the world. For instance, in the program above, it is assumed that all parent relations are given, so that the question "?- parent(bob, pam)," will be answered negatively. The closed world assumption is related to the idea of negation as failure. When a program cannot find a derivation of a statement, it will consider the statement to be false.

An example of a Prolog rule using negation as failure is the following (Bratko 2001):

```
likes(mary, X) :- animal(X), not snake(X)
```

This Prolog rule expresses that Mary likes animals, except snakes. The interpretation of the "not"-operator is not the same as the classical negation of formal logic. Since the not-operator models negation as failure, Mary likes any animal of which it *cannot be derived* that it is a snake. If the program only contains

```
animal(viper)
```

as a fact, it can be derived that "likes(mary, viper)." When the program has the following two factual clauses

```
animal(viper)
snake(viper)
```

the question

```
?- likes(mary, viper)
```

will be answered "No." The example shows that logic programming is related to non-monotonic logic: adding facts can make a derivable fact underivable.

There are technical difficulties involved in the interpretation of the closed world assumption and negation as failure. The so-called stable model semantics of a logic program (Gelfond and Lifschitz 1988) formalizes the interpretation of logic programs with negation as failure. In later sections of this chapter (in particular in Sects. 11.3 and 11.5), we shall see how the stable model semantics of logic programming has influenced argumentation research.

11.2.3 Themes in the Study of Non-monotonic Logics

The study of non-monotonic logics gave hope that logical tools would become more relevant for the study of reasoning and argumentation. To some extent this hope has been fulfilled, since certain themes in reasoning and argumentation that before were at the boundaries of logic are now placed in the center of attention. Examples of such themes are defeasible inference, consistency preservation, and uncertainty. We shall briefly discuss these themes as they are addressed in the chapters of the handbook edited by Gabbay et al. (1994).

An inference is *defeasible* when it can be blocked or defeated in some way (Nute 1994, p. 354). Donald Nute speaks of the presentation of sets of beliefs as reasons for holding other beliefs as advancing *arguments*. When such arguments correspond to a defeasible inference, the argument is defeasible, and blockers or defeaters for an inference are blockers or defeaters for the corresponding argument.

Consistency preservation is the property that the conclusions drawn on the basis of certain premises can only be inconsistent in case the premises are inconsistent (Makinson 1994, p. 51). Makinson reviews general patterns of non-monotonic reasoning, explaining which patterns hold for which systems of non-monotonic reasoning. For instance, a pattern that holds for all systems listed by David Makinson (p. 88) is called *inclusion.* According to this pattern, the conclusions that can defeasibly be inferred from certain premises include those premises themselves. The property of consistency preservation is much more restrictive: the property fails for many non-monotonic systems, meaning that in those systems certain consistent premises can lead to inconsistent conclusions. The property does hold for Reiter's logic for default reasoning, when only normal defaults are allowed. This corresponds to the intuitive meaning of a normal default of the form α : M β/β, namely, that β follows from α, when it is consistent to assume β.

Henry Kyburg (1994, p. 400) distinguishes three kinds of inference involving *uncertainty*. The first is classical, deductive, valid inference about uncertainty. An example of this kind of inference is that when tossing a fair coin, we can conclude that the chance of three times heads in a row is 1/8. The second kind of inference involving uncertainty Kyburg refers to as "inductive" (Kyburg's quotation marks): a categorical conclusion is accepted on the basis of premises that do not logically imply the conclusion, in the sense that the conclusion can be false even when the premises are true. Kyburg uses the flying bird example, discussed above, in which we conclude of a given bird that it flies, though it can happen that it does not. The third kind of inference with uncertainty gives probabilities of particular statements. Kyburg mentions the example "Given what I believe about coins, the chance is 1/8 of getting three heads on the *next* three tosses" (1994, p. 400, Kyburg's italics).

11.2.4 Impact of the Study of Non-monotonic Logic

The study of non-monotonic logic has been very successful as a research enterprise and led to innovations in computer programming in the form of logic-based languages such as Prolog and to commercial applications: expert systems (see Sect. 11.1) often include some form of non-monotonic reasoning.

At the same time, non-monotonic logic did not fulfil all expectations of the artificial intelligence community in which it was initiated. Matthew Ginsberg (1994), for instance, notes – somewhat disappointedly – that the field put itself "in a position where it is almost impossible for our work to be validated by anyone other than a member of our small subcommunity of Artificial Intelligence as a whole" (1994, pp. 28–29). His diagnosis of this issue is that attention shifted from the key objective of building an intelligent artifact to the study of simple examples and mathematics. This leads him to plead for a more experimental, scientific attitude as opposed to a theoretical, mathematical focus.

Ginsberg's position can be connected to adequacy criteria for a system of non-monotonic logic (derived from the issues discussed by Antonelli 2010). Ideally a system of non-monotonic logic scores well on each of the three criteria: material, formal, and computational adequacy. A system is *materially adequate* when it can express a broad range of relevant examples. It is *formally adequate* when it has formal properties that are in line with our expectations (see in particular Makinson 1994). It is *computationally adequate* when the system models forms of inference that can be computed using a reasonable amount of resources (especially time and memory). A key lesson of the research on non-monotonic logic has been that for their fulfillment, these criteria depend on each other and that meeting them all is a complex matter of balancing considerations. One way of interpreting Ginsberg's disappointment is that the focus of the field had shifted too strongly to formal adequacy, paying insufficient attention to material and computational adequacy. As we shall see, the argumentation perspective helped emphasize both the material and the computational adequacy of the systems studied.

11.3 Defeasible Reasoning

In 1987, the publication of John Pollock's paper Defeasible reasoning in cognitive science marked a turning point. The paper emphasized that the philosophical notion of "defeasible reasoning" coincides with what in AI is called "non-monotonic reasoning."[6] Before turning to Pollock's contribution, we discuss some precursors.

11.3.1 Defeasible Reasoning: Origins

As philosophical heritage for the study of defeasible reasoning, Pollock (1987) refers to works by Roderick Chisholm (going back to 1957) and himself (earliest reference in 1967). In an insightful and scholarly historical essay, Ronald Loui (1995) places the origins of the notion of "defeasibility" a decade earlier, namely, in 1948 when the legal positivist H. L. A. Hart presented the paper "The ascription of responsibility and rights at the Aristotelian Society" (Hart 1951). Here is what Hart says:

> [. . .] the accusations and claims upon which law courts adjudicate can usually be challenged or opposed in two ways. First, by a denial of the facts upon which they are based [. . .] and secondly by something quite different, namely a plea that although all the circumstances on which a claim could succeed are present, yet in the particular case, the claim or accusation should not succeed because other circumstances are present which brings the case under some recognized head of exception, the effect of which is either to defeat the claim or accusation altogether, or to "reduce" it so that only a weaker claim can be sustained. (Hart 1951, pp. 147–148; also quoted by Loui 1995, p. 22)

In this quote, Hart not only distinguishes the denial of the premises on which an argument is based from the denial of the inference from the premises to the conclusion, but he also points out that premises that would normally be sufficient may fail because "other circumstances are present."

Although Toulmin (2003) rarely uses the term *defeasible* in *The Uses of Argument* (see Chap. 4, "Toulmin's Model of Argumentation" of this volume), he is obviously an early adopter of the idea of defeasible reasoning, not mentioned by Pollock (1987). Toulmin himself is aware of the connection to Hart (acknowledged by him as inspiration for elements of his model of argument). In elegant modesty,

[6] See the opening sentence of the paper's abstract: "What philosophers call defeasible reasoning is roughly the same as non-monotonic reasoning in AI" (Pollock 1987, p. 481).

Toulmin says that his key distinctions (claims, data, warrants, modal qualifiers, conditions of rebuttal, statements about the applicability or inapplicability of warrants) "will not be particularly novel to those who have studied explicitly the logic of special types of practical argument" (Toulmin 2003, p. 131). Toulmin notes that Hart has shown that the notion of defeasibility is relevant for jurisprudence, free will, and responsibility and that another philosopher, David Ross, has applied it to ethics, recognizing that moral rules may hold prima facie, but can have exceptions.

11.3.2 Pollock's Undercutting and Rebutting Defeaters

In Pollock's approach (1987), "reasoning" is conceived as a process that proceeds in terms of *reasons*. Pollock's reasons correspond to the constellations of premises and a conclusion which argumentation theorists and logicians call (*elementary*) *arguments*. The process is governed by internalized rules that together form the procedural knowledge that allows us to reason correctly. Philosophers, in particular epistemologists such as Chisholm and Pollock himself, have – in Pollock's opinion – a good understanding of the forms of defeasible reasoning, and the construction of computer programs that perform defeasible reasoning provides a good test setting for theories of reasoning. When a theory of reasoning is good, it should be possible to construct a computer program that implements it. By evaluating the program's behavior, the successes and failures (Pollock speaks of *counterexamples*) can be studied.

As said, Pollock's theory of reasoning is built around the notion of "reasons," the building blocks of arguments. Pollock distinguishes two kinds of reasons: (1) a reason is *non-defeasible* when it logically implies its conclusion and (2) a reason P for Q is prima facie when there is a circumstance R such that $P \wedge R$ is not a reason for the reasoner to believe Q.[7] R is then a *defeater* of P as a reason for Q.

Note how closely related the idea of a prima facie reason is to non-monotonic inference: Q can be concluded from P, but not when there is additional information R.

Pollock's standard example is about an object that looks red. "X looks red to John" is a reason for John to believe that X is red, but there can be defeating circumstances, e.g., when there is a red light illuminating the object. See Fig. 11.2.

Pollock has argued for the existence of two kinds of defeaters: "rebutting" and "undercutting defeaters." A defeater is *rebutting*, when it is a reason for the opposite conclusion (Fig. 11.3, left). *Undercutting* defeaters attack the connection between the reason and the conclusion and not the conclusion itself (Fig. 11.3, right). The example about looking red concerns an undercutting defeater since

[7] In this volume, logical symbols are introduced in Sect. 3.3.5 and in Sect. 6.2.3. The symbol "∧" stands for conjunction ("and").

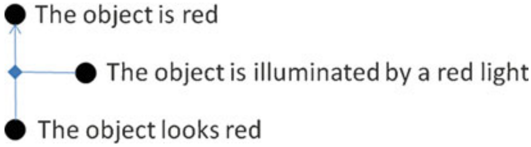


Fig. 11.2 Pollock's red light example

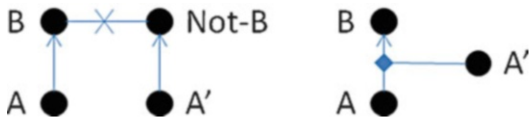


Fig. 11.3 A rebutting defeater and an undercutting defeater

when there is a red light, it is not attacked that the object is red, but merely that the object's looking red is a reason for its being red.

A key element in Pollock's work on defeasible reasoning is the development of a theory of warrant. Pollock uses the term *warrant* as follows: a proposition is *warranted* in an epistemic situation if and only if an ideal reasoner starting in that situation would be justified in believing the proposition. Here justification is based on the existence of an undefeated argument with the proposition as conclusion.

Pollock has developed his theory of warrant in a series of publications which formed the basis of his 1995 book *Cognitive Carpentry*. Time and again, Pollock discovered special situations and examples that led him to revise the criteria determining warrant. He studied, for instance, self-defeating arguments and epistemological paradoxes, such as the lottery paradox. An argument is self-defeating if it contains propositions that are defeaters of other propositions in the argument. In the lottery paradox, there is a fair lottery of a million tickets, so for each specific ticket, there is a good reason to believe that it will not be the winning ticket. When these reasons are combined, one has a reason to believe that no ticket will win; a contradiction. A technically more problematic example goes as follows: P is a prima facie reason for Q, and Q a prima facie reason for R, but R is an undercutting defeater for P as a reason for Q. So if P justifies one's belief in Q, Q itself justifies R. But then P cannot justify Q because of the undercutter R, a contradiction. On the other hand, if P does not justify Q, there must be an argument defeating P as a reason for Q, which requires that R is justified, assuming that it is the only potential defeater available. This requires that Q is justified (assuming that Q is the only potential justification for R available), which is not possible now that P is not justifying (again assuming that P is the only potential justification for Q available).

As a background for his approach to the structure of defeasible reasoning, Pollock provides a list of important classes of specific reasons:

1. *Deductive reasons*. These are the conclusive reasons as they are in particular studied in standard classical logic. For instance, $P \wedge Q$ is a reason for P and for Q, and P and Q taken together are a reason for $P \wedge Q$.
2. *Perception*. When we perceive our world, the resulting perceptual state provides us with prima facie reasons pertaining to our world. Pollock says that no

intellectual mechanism such as reasoning is needed to bring us into such perceptual states. The perceptual state corresponding to the fact that *P* is the case (philosophers speak of "being appeared to" as if *P* were the case) provides a prima facie reason for believing that *P* is the case. In connection with perception, Pollock mentions a general kind of defeater that holds for all prima facie reasons: reliability defeaters.

3. *Memory*. Justified beliefs can also be arrived at by reasoning. The results of reasoning can be rejected when a belief used in the reasoning is later rejected by us. Pollock explains that people often have difficulty to remember the reasons used to arrive at a belief and only remember the beliefs that are the results of the reasoning process. As a consequence, recollection provides a class of prima facie reasons: reasoner S's recalling of *P* is a prima facie reason for S to believe *P*. One undercutting defeater for this class of reasons is that one of the beliefs used in the original reasoning toward *P* is no longer believed and another that reasoner S misremembers. In the latter case, one remembers to have reasoned toward *P*, but that is not the case.
4. *Statistical syllogism*. Pollock describes the statistical syllogism as the simplest form of probabilistic reasoning: from "Most *F*'s are *G*" and "This is an *F'*," we can conclude prima facie "This is a G." The strength of the reason depends on the probability of F's being G. (Pollock notes that qualifications are in place, but we shall not discuss these here.[8]) The use of the statistical syllogism requires that all relevant information is taken into account. As an example, Pollock discusses the probability of arriving home when one is driving. Normally this may be a probability of 0.99, whereas when one is too drunk to stand, this probability may only be 0.5. Pollock explains that "if we know that Jones is driving home and is so drunk he cannot stand, the first probability gives us a prima facie reason for thinking he will get home. But the second probability gives us an undercutting defeater for that instance, leaving us unjustified in drawing any conclusion about whether Jones will get home."
5. *Induction*. Pollock discusses two kinds of induction: (a) in *enumerative induction*, we conclude that all *F*'s are *G* when all *F*'s observed until now have been *G*; (b) in *statistical induction*, we conclude that the proportion of *F*'s being *G* is approximately *r* when the proportion of observed *F*'s being *G* is *r*. About defeaters for inductive reasoning, Pollock remarks that they are complicated and sometimes problematic.

Pollock's theory is embedded in what he called the OSCAR project (Pollock 1995). This project aimed at the implementation of a rational agent. In the project Pollock addressed both theoretical (epistemic) and practical reasoning.[9]

[8] Pollock aims for a theory of projectible properties. See also Pollock (1995, p. 66f).

[9] See Hitchcock (2001, 2002a) for a survey and a discussion of the OSCAR project for those interested in argumentation. Hitchcock also gives further information about Pollock's work on practical reasoning, i.e., reasoning concerning what to do.

11.3.3 Forms of Argument Defeat

In a theory of defeasible reasoning based on arguments that can defeat each other, the question needs to be considered which forms of argument defeat exist.

Above we saw that both Hart and Pollock distinguished different forms of argument defeat. Hart distinguished the denial of the argument's premises and the denial of the inference from reason to conclusion; Pollock distinguished rebutting defeaters that include a reason for an opposite conclusion and undercutting defeaters that only attack the connection between reason and conclusion. We can conclude that Hart's denial of an inference and Pollock's undercutting defeater are closely related notions. As a result, three forms of argument defeat can be distinguished:

1. An argument can be *undermined*. In this form of defeat, the premises or assumptions of an argument are attacked.[10] This form of defeat corresponds to Hart's denial of the premises.
2. An argument can be *undercut*. In this form of defeat, the connection between a (set of) reason(s) and a conclusion in an argument is attacked.
3. An argument can be *rebutted*. In this form of defeat, an argument is attacked by giving an argument for an opposite conclusion.

Precisely these three forms of argument defeat are used in a recent state-of-the-art system for the formal modelling of defeasible argumentation, ASPIC+ (Prakken 2010),[11] building on experiences in the ASPIC project.[12]

Bart Verheij (1996a, p. 122 f.) distinguishes two further forms of argument defeat: "defeat by sequential weakening" and "defeat by parallel strengthening." In *defeat by sequential weakening*, each step in an argument is correct, but the argument breaks down when the steps are chained. An example is an argument based on the sorites paradox:

This body of grains of sand is a heap.
So, this body of grains of sand minus 1 grain is a heap.
So, this body of grains of sand minus 2 grains is a heap.
. . .
So, this body of grains of sand minus *n* grains is a heap.

At some point, the argument breaks down, in particular when *n* exceeds the total amount of grains of sand to start with.

[10] This form of defeat is the basis of Bondarenko et al. (1997). We shall here not elaborate on the distinction between premises and assumptions. One way of thinking about assumptions is to see them as defeasible premises. See Sect. 11.5.3.

[11] Prakken (2010) speaks of ways of attack, where argument defeat is the result of argument attack.

[12] The ASPIC project (full name: Argumentation Service Platform with Integrated Components) was supported by the EU 6th Framework Programme and ran from January 2004 to September 2007. In the project, academic and industry partners cooperated in developing argumentation-based software systems.

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.

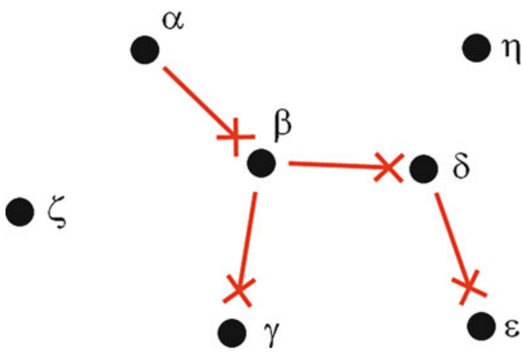


Fig. 11.4 An argumentation framework representing attack between arguments

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 $\{\alpha, \gamma\}$: 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 $\{\alpha, \beta\}$ is not admissible since it is not conflict-free. The set $\{\gamma\}$ 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 $\{\alpha, \gamma\}$.

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 $\{\alpha\}$ and $\{\beta\}$ are each admissible. For instance, $\{\alpha\}$ 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 $\{\alpha\}$, 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).

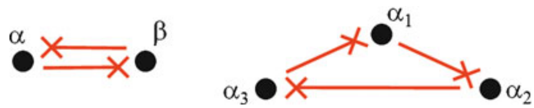


Fig. 11.5 Arguments attacking each other in cycles

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.

11.5 Arguments with Structure

Abstract argumentation, discussed in Sect. 11.4, focuses on the attack relation between arguments, abstracting from the structure of arguments. In this section we will discuss various ways of considering the structure of arguments for and against conclusions. The section is organized thematically in order to present general ideas rather than concrete systems. The themes discussed are arguments and specificity, the comparison of conclusive force, arguments with prima facie assumptions, arguments and classical logic, and the combination of support and attack.

11.5.1 Arguments and Specificity

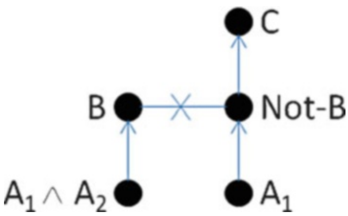
An early theme in the formal study of argumentation was that of "argument specificity" in relation to the resolution of a conflict between arguments. The key idea connecting arguments and specificity is that when two arguments are conflicting, with one of them being based on more specific information, the more specific argument wins the conflict and defeats the more general argument.

Guillermo Simari and Ronald Loui (1992) have provided a mathematical formalization of this connection between arguments and specificity. Their work was inspired by Poole's (1985) work on specificity in the field of non-monotonic logic. Poole proposed to consider default hypotheses as explanations akin to scientific theories that need to be compared, such that more specific information is preferred to more general information. Simari and Loui (1992) aimed to combine specificity with a theory of argument, connecting to Pollock's work on argumentative warrant. In their proposal, an argument is a pair (T, *h*), with T being a set of defeasible rules that are applied to arrive at the argument's conclusion *h* given the argument's premises (formalized in the background knowledge). Arguments are assumed to be consistent, in the sense that no contradiction can be derived (not even defeasibly). Also arguments are assumed to be minimal, in the sense that all rules are needed to arrive at the conclusion. Formally, for an argument (T, *h*), it holds that when T' is the result of omitting one or more rules in T, the pair (T', *h*) is not an argument. Two arguments (T, *h*) and (T', *h*') *disagree* when *h* and *h*' are logically incompatible, given the background knowledge. An argument (T, *h*) *counterargues* an argument (T', *h*') if (T, *h*) disagrees with an argument (T", h") that is a sub-argument of (T', *h*'), i.e., T" is a subset of T'. An argument (T, *h*) *defeats* an argument (T', *h*') when (T, *h*) disagrees with a sub-argument of (T', *h*') that is strictly less specific.

For instance, given defeasible rules $A_1 \wedge A_2 \Rightarrow B$, $A_1 \Rightarrow$ not-B, and not-$B \Rightarrow C$, and premises $A_1 \wedge A_2$ and A_1, the argument ($\{A_1 \wedge A_2 \Rightarrow B\}$, B) disagrees with the strictly less specific argument ($\{A_1 \Rightarrow$ not-$B\}$, not-B), so counterargues and defeats ($\{A_1 \Rightarrow$ not-B, not-$B \Rightarrow C\}$, C). The graphical structure of the argumentation is shown in Fig. 11.6.

Simari and Loui's approach has been developed further – with applications in artificial intelligence, multi-agent systems, and logic – by the Bahia Blanca group,

Fig. 11.6 An argument defeated by a more specific counterargument



led by Simari (e.g., García and Simari 2004; Chesñevar et al. 2004; Falappa et al. 2002). García and Simari (2004) show the close connection between argumentation and logic programming that was also an inspiration for Dung (1995) (see also the Sect. 11.4.1, above, and Sect. 11.2.2). In their DeLP system of *defeasible logic programming*, they have developed close connections with logic programming. In a defeasible logic program, facts, strict rules and defeasible rules are represented. In DeLP, arguments are constructed by constructing derivations starting from the facts and using the strict and defeasible rules. Arguments cannot support opposing literals and obey a minimality constraint. Which arguments are counterarguments is determined using the notion of "disagreeing literals," i.e., elementary claims and their negations.

It has been argued that specificity can only be one among several domain-dependent conflict resolution strategies. For example, in the law, a conflict between arguments based on two rules can be resolved not only by the specificity of the rules but also by their recency or authority (Hage 1997; Prakken 1997; see also Sect. 11.7). Pollock has argued against the general applicability of specificity defeat in logically complex situations.[18]

11.5.2 Comparing Conclusive Force

Another criterion that can determine which conflicting arguments survive the conflict is conclusive force. Arguments that have more conclusive force will survive more easily than arguments with less conclusive force.

One idea that connects conclusive force with argument structure is the *weakest link principle*, which Pollock characterizes as follows:

> The degree of support of the conclusion of a deductive argument is the minimum of the degrees of support of its premises. (1995, p. 99)

Pollock presents the weakest link principle as an alternative to a Bayesian approach, which he rejects.

Gerard Vreeswijk (1997) has proposed an abstract model of argumentation with defeasible arguments that focuses on the comparison of the conclusive force of arguments. In his model, conclusive force is not modelled directly but as an abstract

[18] He believes that a projectibility constraint is required (1995, pp. 105–106). See Note 8.

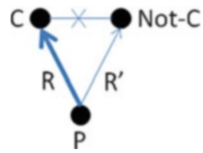


Fig. 11.7 Two conflicting arguments from the same premises, one with stronger conclusive force

comparison relation that expresses which arguments have more conclusive force than which other arguments. Vreeswijk had been looking for general, nontrivial, principles for determining the relative strength among arguments, but found that syntactic principles are not enough. By abstracting from such principles, the comparison of the conclusive force of arguments is no longer dealt with by the formal model itself, but becomes part of the domain knowledge as it can be represented using the formalism. According to Vreeswijk, this abstraction "saves us from the responsibility of telling how and why a particular argument should overrule any other particular argument" it "frees from the involvements with specificity and conclusive force in which it was enmeshed" (1997, p. 229). Instead, the focus can now be on the relations between argument structure, comparative conclusive force, and argument defeat.

Vreeswijk defines an *abstract argumentation system* as a triple (L, R, $\leq$), where L is a set of sentences expressing the claims made in an argument, R is a set of defeasible rules allowing the construction of arguments, and $\leq$ represents the conclusive force relation between arguments. The rules come in two flavors: strict and defeasible. Arguments are constructed by chaining rules. A set of arguments Σ is a *defeater* of an argument α if Σ and α are incompatible (i.e., imply an inconsistency), and α is not an underminer of Σ. An argument α is an *underminer* of a set of arguments Σ if Σ contains an argument β that has strictly lower conclusive force than α.

Vreeswijk's model of abstract conclusive force can, for instance, be readily applied to Pollock's notion of rebutting defeaters. Assume that, given the set of premises P, we have both a reason *R* for *C* and a reason *R*' for not-*C*. If now an argument based on *R* has higher conclusive force than an argument based on *R*', the former is a defeater of the latter (Fig. 11.7).

Whereas Dung's (1995) system is abstract, since it only considers argument attack, Vreeswijk's proposal is abstract in particular because the conclusive force relation is left unspecified. Vreeswijk gives the following examples of conclusive force relations:

1. *Basic order*. In this order, a strict argument has more conclusive force than a defeasible argument. In a strict argument, no defeasible rule is used.
2. *Number of defeasible steps*. An argument has more conclusive force than another argument if it uses less defeasible steps. Vreeswijk remarks that this is not a very natural criterion, but it can be used to give formal examples and counterexamples.
3. *Weakest link*. Here the conclusive force relation on arguments is derived from an ordering relation on the rules. An argument has more conclusive force than another if its weakest link is stronger than the weakest link of the other.

4. *Preferring the most specific argument*. Of two defeasible arguments, one has more conclusive force than the other if the first has the premises of the second among its conclusions.

11.5.3 Arguments with Prima Facie Assumptions

In other proposals, the defeat of arguments is the result of prima facie assumptions that are successfully attacked. In their abstract, argumentation-theoretic approach to default reasoning, Bondarenko et al. (1997) use such an approach. Using a given deductive system (L, R) that consists of a language L and a set of rules R, so-called deductions are built by the application of rules. Given a deductive system (L, R), an *assumption-based framework* is then a triple (T, Ab, Contrary), where T is a set of sentences expressing the current beliefs, Ab expresses assumptions that can be used to extend T, and Contrary is a mapping from the language to itself that expresses which sentences are contraries of which other sentences. Bondarenko and colleagues define a number of semantics (similar to Dung's 1995 in the context of abstract argumentation). For instance, a *stable extension* is a set of assumptions Δ such that the following properties hold:

1. Δ is closed, meaning that Δ contains all assumptions that are logical consequences of the beliefs in T and Δ itself.
2. Δ does not attack itself, meaning that there is no deduction from the beliefs in T and Δ with a contrary of an element of Δ as conclusion.
3. Δ attacks each assumption not in Δ, meaning that, for every assumption outside Δ, there is a deduction from T and Δ with a contrary of that assumption as conclusion.

As an example, Bondarenko and colleagues use the principle that a person is innocent unless proved guilty. When, for instance, the formula "¬guilty" "→ innocent" (in classical logic) is the only belief in T and "¬guilty" is the only assumption, then there is one stable extension, consisting of the elements of T, "¬guilty", and their logical consequences (p. 71).

Bondarenko and colleagues show that several systems of non-monotonic logic can be modelled in their assumption-based framework. Its argumentative nature stems from the fact that it is built around the notion of attack, specifically attack on the assumptions that can be added to one's beliefs.

Verheij (2003a) has also developed an assumption-based model of defeasible argumentation. A difference with Bondarenko et al. (1997) is that the rules that are applied when drawing defeasible conclusions are themselves part of the assumptions. Technically, the rules have become conditionals in the underlying language. As a result, it can be the issue of an argument whether some proposition supports another proposition. In this way, Pollock's undercutting defeaters can be modelled as an attack on a conditional. Pollock's example of an object that looks red (Sect. 11.3.2) is formalized using two conditional sentences:

looks_red → is_red
red_light → ×(looks_red → is_red)

The first expresses the conditional prima facie assumption that if something looks red, it is red. The second expresses an attack on this prima facie assumption: when there is a red light illuminating the object, it no longer holds that if the object looks red, it is red. The sentences illustrate the two connectives of the language: one to express the conditional (→) and the other to express what is called *dialectical negation* (×). The two conditional sentences correspond exactly to two graphical elements in Fig. 11.2: the first to the arrow connecting the reason and the conclusion and the second, nested, conditional to the arrow (ending in a diamond) that expresses the attack on the first conditional. This isomorphism between formal structures of the language and graphical elements has been used for the diagrams supported by the argumentation software ArguMed (Verheij 2005b; see Sect. 11.11).

The use of assumptions raises the question how they are related to an argument's ordinary premises. Assumptions can be thought of as the defeasible premises of an argument, and as such they are akin to defeasible rules[19] with an empty antecedent. The Carneades framework (Gordon et al. 2007) distinguishes three kinds of argument premises: ordinary premises, presumptions (much like the prima facie assumptions discussed in this subsection), and exceptions (which are like the contraries of assumptions).

11.5.4 Arguments and Classical Logic

The relation between classical logic and defeasible argumentation remains a puzzle. Above we already saw different attempts at combining elements of classical logic and defeasible argumentation. In Pollock's system, classical logic is one source of reasons. Often conditional sentences ("rules") are used to construct arguments by chaining them (e.g., Vreeswijk 1997). Chaining rules is closely related to the inference rule *modus ponens* of classical logic. Verheij's (2003a) system gives conditionals which validate *modus ponens* a central place. Bondarenko et al. (1997) allow generalized rules of inference by their use of a contingent deductive system as starting point.

Besnard and Hunter (2008) have proposed to formalize arguments in classical logic entirely. For them, an argument is a pair (Φ, α), such that Φ is a set of sentences and α is a sentence and such that Φ is logically consistent, Φ logically entails α (in the classical sense), and Φ is a minimal such set. (Note the analogy with the proposal by Simari and Loui 1992; see Sect. 11.5.1, above.) Φ is the support of

[19] Some would object to the use of the term *rules* here. Rules are here thought of in analogy with the inference rules of classical logic. An issue is then that, as such, they are not expressed in the logical object language, but in a metalanguage. In the context of defeasible reasoning and argumentation (and also in non-monotonic logic), this distinction becomes less clear. Often there is one logical language to express ordinary sentences, a second formal language (with less structure and/or less semantics and therefore not usually referred to as "logical") used to express the rules, and the actual metalanguage that is used to define the formal system.

the argument, and α the claim. They define defeaters as arguments that refute the support of another argument. More formally, a *defeater* for an argument (Φ, α) is an argument (Ψ, β), such that β logically entails the negation of the conjunction of some of the elements of Φ. An *undercut* for an argument (Φ, α) is an argument (Ψ, β) where β is equal to (and not just entails) the negation of the conjunction of some of the elements of Φ. A *rebuttal* for an argument (Φ, α) is an argument (Ψ, β) such that β↔¬α is a tautology. Besnard and Hunter give the following example (p. 46):

p Simon Jones is a Member of Parliament.
$p \rightarrow \neg q$ If Simon Jones is a Member of Parliament, then we need not keep quiet about details of his private life.
r Simon Jones just resigned from the House of Commons.
$r \rightarrow \neg p$ If Simon Jones just resigned from the House of Commons, then he is not a Member of Parliament.
$\neg p \rightarrow q$ If Simon Jones is not a Member of Parliament, then we need to keep quiet about details of his private life.

Then $(\{p, p \rightarrow \neg q\}, \neg q)$ is an argument with the argument $(\{r, r \rightarrow \neg p\}, \neg p)$ as an undercut and the argument $(\{r, r \rightarrow \neg \mathrm{p}, \neg p \rightarrow q\}, q)$ as a rebuttal.

Besnard and Hunter focus on structural properties of arguments, in part because of the diversity of proposals for semantics (see Sect. 11.4). For instance, when they discuss these systems, they note that the semantic conceptualization of such systems is not as clear as the semantics of classical logic, which is the basis of their framework (p. 221, also p. 226). At the same time, they note that knowledge representation can be simpler in systems based on defeasible logic (see the next subsection) or inference rules.

11.5.5 Combining Support and Attack

In this subsection, we discuss ways to combine support and attack when modelling argumentation. In several proposals, support and attack are combined in separated steps. In the first step, argumentative support is established by constructing arguments for conclusions from a given set of possible reasons or rules (of inference). The second step determines argumentative attack. Attack is, for instance, based on defeaters or on the structure of the supporting arguments in combination with a preference relation on arguments. In the third and final step, it is determined which arguments are warranted or undefeated. We already saw that several criteria have been proposed (e.g., Pollock's gradual development of criteria for argumentative warrant and Dung's abstract argumentation semantics).

An example of this modelling style is depicted in Fig. 11.8. Three supporting arguments are shown. The first on the left shows that *A* supports *B*, which in turn supports *C*. In the middle of the figure, this argument is attacked by a second argument, which reasons from *A*' to Not-B (hence against *B*). This argument is in turn attacked by a third argument, which reasons from *A*" against the support relation R between *A*' and Not-B. Using the terminology of Sect. 11.3.2, the first sub-argument of the first argument is rebutted by the second, which is undercut by

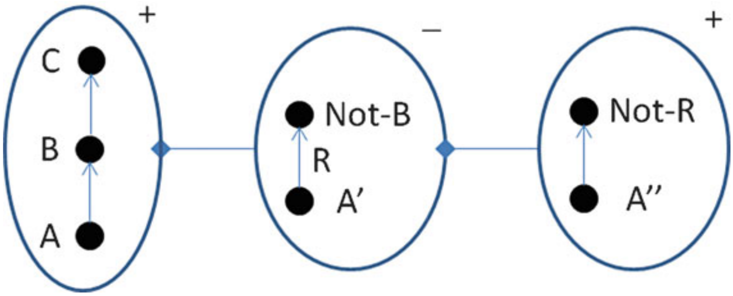


Fig. 11.8 Supporting arguments that attack each other

Fig. 11.9 The abstract argumentation framework associated with the example of Fig. 11.8

the third. The arguments are marked with a + sign when they are warranted and a – sign when they are defeated (which can be thought of as a variant of the labelling approaches of Sect. 11.4.2). The argument on the right is warranted, since it is not attacked. As a result, the middle argument is defeated, since it is attacked by a warranted argument. The left argument is then also warranted, since its only attacker is defeated. (See the procedure for computing the grounded extension of an argumentation framework discussed in Sect. 11.4.2.)

In this approach, the relation with Dung’s abstract argumentation is that we can abstract from the structure of the supporting arguments resulting in an abstract argumentation framework. For the example in Fig. 11.8, we get the abstract framework shown in Fig. 11.9. In this example, the argumentation semantics is unproblematic at the abstract argument attack level since the grounded extension coincides with the unique preferred extension that is also stable. Special care is needed to handle parts of arguments. For instance, the middle argument has the premise *A*’, which is not attacked and should therefore remain undefeated.

This type of combining support and attack is used in the ASPIC + model (Prakken 2010). As discussed in Sect. 11.3.3, the ASPIC + model incorporates the three main forms of argument defeat: undermining, undercutting, and rebutting.

A second approach does not separate support and attack when combining them. Arguments are constructed from reasons for and against conclusions, which in turn determine whether a conclusion follows or not. Figure 11.10 models the same argumentative information as Fig. 11.8, but now using this second approach.

Here the reason *A*” undercuts the argument from *A*’ to Not-B, so Not-B is not supported (indicated by the open circle). As a result, Not-B does not actually attack *B*, which is therefore justified by *A* and in turn justifies *C*.

In this approach, for instance, conditional sentences are used to express which reasons support or attack which conclusions. An example is Nute’s *defeasible logic*

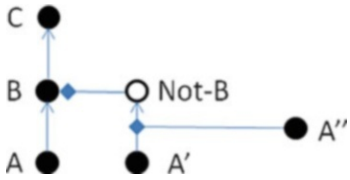


Fig. 11.10 Arguments supporting and attacking conclusions

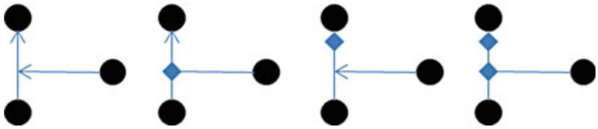


Fig. 11.11 The four ways of arguing about support and attack

(Nute 1994; Antoniou et al. 2001), which uses conditional sentences for the representation of strict rules and defeasible rules and for defeater rules, which can block an inference based on a defeasible rule. Algorithms for defeasible logic have been designed with good computational properties.

Another example of the approach is Verheij's DefLog (2003a), in which a conditional for the representation of support is combined with a negation operator for the representation of attack. A related proposal extending Dung's abstract argumentation frameworks by expressing both support and attack is bipolar argumentation (Cayrol and Lagasquie-Schiex 2005; Amgoud et al. 2008). For DefLog and bipolar argumentation, generalizations of Dung's stable and preferred semantics are presented. DefLog has been used to formalize Toulmin's argument model (Verheij 2005b).

A special case of the combination of support and attack occurs when the support and attack relations can themselves be supported or attacked. Indeed it can be at issue whether a reason supports or attacks a conclusion. The four ways of arguing about support and attack are illustrated in Fig. 11.11, from left to right: support of a support relation, attack of a support relation, support of an attack relation, and attack of an attack relation, respectively.

For instance, Pollock's undercutting defeaters can be thought of as attacks of a support relation (second from the left in Fig. 11.11). In Verheij's DefLog (2003a, 2005b), the four ways are expressed using nested conditional sentences, in a way that extends the expressiveness of Dung's frameworks. Modgil (2005) has studied attacks of attacks (rightmost in Fig. 11.11) in a system that also extends Dung's expressiveness.

11.6 Argument Schemes

Argumentation formalisms can only come to life when arguments are built from meaningful reasons. We already saw (in subsection 11.3.2) that Pollock made explicit which kinds of reasons he considered: deductive reasons, perception, memory, statistical syllogism, and induction.

An approach to the specification of meaningful kinds of reasons to construct arguments from is that of *argument schemes*, as they have been studied in argumentation theory. Argument schemes were already distinguished by Perelman and Olbrechts-Tyteca (1969).[20] In today's artificial intelligence research on argumentation, Douglas Walton's approach to *argumentation schemes* (his terminology) has been widely adopted (e.g., Walton et al. 2008).

Argument schemes can be thought of as analogues of the rules of inference of classical logic. An example of a rule of inference is, for instance, the following version of modus ponens:

P
If *P*, then *Q*
Therefore: *Q*

Whereas logical rules of inference, such as *modus ponens*, are abstract, strict, and (usually) considered to have universal validity, argument schemes are concrete, defeasible, and context dependent. An example is the following scheme for witness testimony:

Witness A has testified that *P*.
Therefore: *P*

The use of this scheme is defeasible, as can be made explicit by asking critical questions, for instance:

Wasn't A mistaken?
Wasn't A lying?

A key reason why argument schemes have been taken up in artificial intelligence is that the critical questions associated with them correspond to defeating circumstances. For instance, the question whether A was mistaken gives rise to the defeater "A was mistaken."

Bex et al. (2003) applied the concept of "argumentation schemes" to the formalization of legal reasoning from evidence. An example of a scheme in that paper is the following:

Argument from expert opinion
Source E is an expert in domain D.
E asserts that proposition *A* is known to be true (false).
A is within D.
Therefore, *A* may plausibly be taken to be true (false).

This scheme has the following critical questions:

1. Expertise question: How credible is E as an expert source?
2. Field question: Is E an expert in D?
3. Opinion question: What did E assert that implies *A*?

[20] Although the term *schème argumentative* [argumentative scheme] was already used by Perelman and Olbrechts-Tyteca, according to Garssen (2001), van Eemeren et al. (1978, 1984) used the notion of argument(ation) scheme for the first time in its present sense. See also van Eemeren and Kruiger (1987), van Eemeren and Grootendorst (1992a), Kienpointner (1992), and Walton et al. (2008).

4. Trustworthiness question: Is E personally reliable as a source?
5. Consistency question: Is *A* consistent with what other experts assert?
6. Backup evidence question: Is E's assertion based on evidence?

The authors elaborate on how these and other argumentation schemes related to evidential reasoning can be formalized.

From the perspective of artificial intelligence, the work on argumentation schemes of Walton and his colleagues can be regarded as contributions to the theory of knowledge representation. Gradually, a collection of argumentation schemes is being developed. When appropriate, a scheme is added, and existing schemes are adapted, e.g., by refining the scheme's premises or critical questions. This knowledge representation point of view is developed by Verheij (2003b), who, like Bex et al. (2003), formalizes argumentation schemes as defeasible rules of inference. He notes that in Walton's work, argumentation schemes sometimes take the form of small derivations, or sequences of argumentation schemes, or even of a small prototypical dialogue. To streamline the work on knowledge representation, Verheij proposes to treat argumentation schemes as consisting of four elements: conclusion, premises, conditions of use, and exceptions. The exceptions correspond to answers to the critical questions of an argumentation scheme. By this representation format, it is also possible to consider different roles of critical questions: critical questions may concern a conclusion, a premise, a condition of use, or an exception.

Reed and Rowe (2004) have incorporated argumentation schemes in their Araucaria tool for the analysis of argumentative texts. Rahwan et al. (2007) have proposed formats for the integration of argumentation schemes in what is called the *Semantic Web*. The vision underlying the Semantic Web is that, when information on the Internet is properly tagged, it becomes possible to add meaning to such information that can be handled by a machine. For instance, when the conclusion, premises, conditions of use, and exceptions of an argumentation scheme are marked as such, software can be built that can handle these different elements of a scheme appropriately. Gordon et al. (2007) have integrated argumentation schemes in their Carneades model.

A fundamental issue concerning argumentation schemes is how to evaluate a scheme or set of schemes: When is a scheme good, and under which circumstances? When is an adaptation appropriate? This issue is, for instance, discussed in Reed and Tindale (2010).

11.7 Argumentation Dialogues

One reason why Toulmin's (2003, 1958) *The Uses of Argument* remains a thought-provoking study is his starting point that argument should be considered in its natural, critical, and procedural context. This starting point led him to propose that logic, in the sense of the theory of good argument, should be treated as "generalized jurisprudence," where a critical and procedural perspective on good argument is the norm. The critical and procedural sides of arguments come together in the study of argumentation dialogues.

The following is a fragment, taken from McBurney and Parsons (2002a), of an argumentation dialogue concerning the sale of a used car between a buyer (B) and seller (S), illustrating the study of argumentative dialogue in a computational setting:

> S: BEGIN(PERSUASION(Make);PERSUASION(Condition_of_Engine); PERSUASION (Number_of_Owners))
> S requests a sequence of three Persuasion dialogues over the purchase, criteria Make, Condition of the Engine, and Number of Owners.
> B: AGREE(PERSUASION(Make); PERSUASION(Condition_of_Engine); PERSUASION(Number_of_Owners))
> PERSUASION Dialogue 1 in the sequence of three opens.
> S: Argues that "Make" is the most important purchase criterion, within any budget, because a typical car of one Make may remain in better condition than a typical car of another Make, even though older.
> B: Accepts this argument.
> PERSUASION Dialogue 1 closes upon acceptance of the proposition by B. PERSUASION Dialogue 2 opens.
> S: Argues that "Condition_of_Engine" is the next most important purchase criterion.
> B: Does not accept this. Argues that he cannot tell the engine condition of any car without pulling it apart. Only S, as the Seller, is able to tell this. Hence, B must use "Mileage" as a surrogate for "Condition_of_Engine."
> PERSUASION Dialogue 2 closes with neither side changing its views: B does not accept "Condition_of_Engine" as the second criterion, and S does not accept "Mileage" as the second criterion. PERSUASION Dialogue 3 opens.

The fragment shows how dialogues about certain topics are opened and closed in relation to the arguments provided.

The formal and computational study of argumentation dialogues has primarily been performed in the fields of AI and law and of multi-agent systems, as addressed in the following two subsections.

11.7.1 Argumentation Dialogues in AI and Law

In the field of AI and law, argumentation dialogues have been studied extensively (see Bench-Capon et al 2004, 2009). Ashley's (1990) HYPO, to be discussed further in Sect. 11.9, takes a 3-ply dialogue model as starting point, in which a proponent makes a claim, which can be attacked by an opponent and then defended by the proponent. An early AI and law conception of argumentation dialogue is Thomas Gordon's (1993, 1995) *Pleadings game*. Gordon formalizes the pleading in a civil law process, which he considers to be aimed at determining the legal and factual issues of a case. In the Pleadings game, a proponent and opponent (in this setting referred to as "plaintiff" and "defendant") can concede, deny, and defend claims and also declare defeasible rules. Players can discuss the validity of a defeasible rule. Players are committed to the consequences of their claims, as prescribed by a non-monotonic logic underlying the Pleadings game.

Other dialogue models of argumentation in AI and law have been proposed by Prakken and Sartor (1996, 1998), Hage et al. (1993), and Lodder (1999). In Prakken and Sartor's approach (1996, 1998), dialogue models are presented as a kind of proof theory for their argumentation model. Prakken and Sartor interpret a proof as a dialogue between a proponent and opponent. An argument is justified when there is a winning strategy for the proponent of the argument. Hage et al. (1993) and Lodder (1999) propose a model of argumentation dialogues with the purpose of establishing the law in a concrete case. They are inspired by the idea of law as a pure procedure (though not endorsing it): when the law is purely procedural, there is no criterion for a good outcome of a legal procedure other than the procedure itself.

Some models emphasize that the rules of argumentative dialogue can themselves be the subject of debate. An actual example is a parliamentary discussion about the way in which legislation is to be discussed. In philosophy, Suber has taken the idea of self-amending games to its extreme by proposing the game of Nomic, in which the players can gradually change the rules.[21] Proposals to formalize such meta-argumentation include Vreeswijk (2000) and Brewka (2001), who have proposed formal models of argumentative dialogues allowing self-amendments.[22]

In an attempt to clarify how logic, defeasibility, dialogue, and procedure are related, Henry Prakken (1997, p. 270 f.) proposed to distinguish four layers of argumentation models. The first is the logical layer, which determines contradiction and support. The second layer is dialectical, which defines what counts as attack, counterargument, and also when an argument is defeated. The third layer is procedural and contains the rules constraining a dialogue, for instance, which moves parties can make, e.g., propose a claim or present a counterargument, when parties can make a move, e.g., when it is their turn and when the dialogue is finished. The fourth and final layer is strategic. At this layer, one finds the heuristics used by a good, effective arguer.

Jaap Hage (2000) addresses the question of why dialogue models of argumentation became popular in the field of AI and law. He gives two reasons. The first is that legal reasoning is defeasible, and dialogue models are a good tool to study defeasibility. The second reason is that dialogue models are useful when investigating the process of establishing the law in a concrete case. Hage recalls the legal theoretic discussion about the law as an open system, in the sense that there can be disagreement about the starting points of legal arguments. As a result, the outcome of a legal procedure is indeterminate. A better understanding of this predicament can be achieved by considering the legal procedure as an argumentative dialogue.

Hage (2000) then discusses three functions of dialogue models of argumentation in AI and law. The first function is to define argument justification, in analogy with dialogical definitions of logical validity as can be found in the work by Lorenzen and Lorenz (1978). In this connection, Hage refers to Barth and Krabbe's notion of the "dialectical garb" of a logic as opposed to an axiomatic,

[21] http://en.wikipedia.org/wiki/Nomic. See also Hofstadter (1996, chapter 4).

[22] See also the study of Nomic by Vreeswijk (1995a).

inferential, or model-theoretic garb (Barth and Krabbe 1982, pp. 7–8). See also Sect. 6.5 of this volume. Hage generalizes the idea of dialectical garb to what he refers to as *battle of argument* models of defeasible reasoning in which arguments attack each other, such as Loui's (1987), Pollock's (1987, 1994), Vreeswijk's (1993), Dung's (1995), and Prakken and Sartor's (1996). Battle of argument models can or cannot be presented in a dialectical garb. In their dialectical garb, such models define the justification of an argument in terms of the existence of a winning strategy in an argumentative dialogue game.

The second function of dialogue models of argumentation that is distinguished by Hage is to establish shared premises. Proponent and opponent enter into a dialogue that leads to a shared set of premises. The conclusions that follow from these shared premises can be regarded as justified. In this category, Hage discusses Gordon's Pleadings game, which we discussed above. Hage makes connections to legal theory, in particular Alexy's (1978) procedural approach to legal justification, and the philosophy of truth and justification, in particular Habermas's (1973) consensus theory of truth and Schwemmer's approach to justification, in which the basis of justification is only assumed as long as it is not actually questioned (Schwemmer and Lorenzen 1973).

As a third and final function of dialogue models of argumentation in AI and law, Hage discusses the procedural establishment of law in a concrete case. In this connection, he discusses mediating systems, which are systems that support dialogues, instead of evaluating them. He uses Zeno (Gordon and Karacapilidis 1997), Room 5 (Loui et al. 1997) (see also Sect. 11.11), and DiaLaw (Lodder 1999) as examples. Hage argues that regarding the law as purely procedural is somewhat counterintuitive, since there exist cases in which there is a clear answer, which can be known even without actually going through the whole procedure. Hage speaks therefore of the law as an imperfect procedure, in which the correctness of the outcome is not guaranteed.

11.7.2 Argumentation Dialogues in Multi-agent Systems

Outside the field of AI and law, one further function of dialogue models of argumentation has been emphasized, namely, that a dialogue perspective on argumentation can have computational advantages. For instance, argumentative dialogue can be used to optimize search, e.g., by cutting off dead ends or focusing on the most relevant issues. Vreeswijk (1995b) takes this assumption as the starting point of a paper:

> If dialectical concepts like argument, debate, and resolution of dispute are seemingly so important in practical reasoning, there must be some reason as to why these techniques survived as rulers of commonsense argument. Perhaps the reason is that they are just most suited for the job. (Vreeswijk 1995b, p. 307)

Vreeswijk takes inspiration from a paper by Loui (1998), which circulated in an earlier version since 1992. Loui emphasizes the relevance of protocol, the assignment of burdens to parties, termination conditions, and strategy. A key idea is that argumentation dialogues are well suited for reasoning in a setting of bounded resources (see also Loui and Norman 1995).

Inspired by the computational perspective on argumentation, approaches to argumentative dialogue have been taken up in the field of multi-agent systems.[23] The focus in that field is on the interaction between autonomous software agents that pursue their own goals or goals shared with other agents. Since the actions of one agent can affect those of another, beyond control of an individual agent or the system as a whole, the kinds of problems when designing multi-agent software systems are of a different nature than those in the design of software where control can be assumed to be centralized. Computational models of argumentation have inspired the development of interaction protocols for the resolution of conflicts among agents and for belief formation. The typology of argumentative dialogue that has been proposed by Douglas Walton and Erik Krabbe (1995) has been especially influential (see also Sect. 7.8 of this volume).[24]

In particular, the persuasion dialogue, starting with a conflict of opinion and aimed at resolving the issue by persuading a participant, has been extensively studied. An early persuasion system (predating Walton and Krabbe's typology) is Sycara's Persuader system (1989). Persuader, developed in the field of what was then called *Distributed AI*, uses the domain of labor negotiation as an illustration. An agent forms a model of another agent's beliefs and goals and determines its actions in such a way that it influences the other agent. For instance, agents can choose a so-called threatening argument, i.e., an argument that is aimed at persuading another agent to give up a goal. Here it is notable that in Walton and Krabbe's typology, negotiation is a dialogue type different from persuasion.

Prakken (2006, 2009) gives an overview and analysis of dialogue models of persuasion. In a dialogue system, dialogues have a goal and participants. It is specified which kinds of moves participants can make, e.g., making claims or conceding. Participants can have specific roles, e.g., Proponent or Opponent. The actual flow of a dialogue is constrained by a protocol, consisting of rules for turn taking and termination. Effect rules determine how the commitments of participants change after a dialogue move. Outcome rules define the outcome of the dialogue, by determining, for instance, in persuasion dialogues who wins the dialogue. These elements are common to all dialogue types. By specifying or constraining the elements, one generates a system of persuasion dialogue. In particular, the dialogue goal of persuasion dialogue consists of a set of propositions that are at issue and need to be resolved. Prakken formalizes these elements and then uses his analytic model to discuss several extant persuasion systems, among them Mackenzie's

[23] For an overview of the field of multi-agent systems, see the textbook by Wooldridge (2009), which contains a chapter entitled "Arguing."

[24] The 2000 Symposium on Argument and Computation at Bonskeid House, Perthshire, Scotland, organized by Reed and Norman, has been a causal factor. See Reed and Norman (2004b).

(1979) proposals and Walton and Krabbe's (1995) model of what they call *Permissive persuasion dialogue* (see Sect. 6.9 of this volume).

Sycara's Persuader system (1989) is a persuasion system applied to labor negotiation. Parsons et al. (1998) also speak of negotiation as involving persuasion. Their model uses the belief-desire-intention model of agents (Rao and Georgeff 1995) and specifies logically how the beliefs, desires, and intentions of the agents influence the process of negotiation.[25] Dignum et al. (2001) have studied the role of argumentative dialogue for the forming of coalitions of agents that create collective intentions. Argumentation about what to do rather than about what is the case has been studied in a dialogue setting by Atkinson and colleagues (Atkinson et al. 2005, 2006; Atkinson and Bench-Capon 2007). It is noteworthy that Pollock's OSCAR model (1995) is an attempt to combine theoretical and practical reasoning, but in a single agent setting. Amgoud (2009) discusses the application of dialogical argumentation to decision making (see also Girle et al. 2004). Deliberation has been studied by McBurney et al. (2007).

Several attempts have been made to systematize the extensive work on argumentation dialogue. Bench-Capon et al. (2000), for instance, propose a formal method for modelling argumentation dialogue. Prakken (2005b) provides a formal framework that can be used to study argumentation dialogue models with different choices of underlying argument model and reply structures. McBurney and Parsons (2002a, b, 2009) have developed an abstract theory of argumentative dialogue in which syntactic, semantic, and pragmatic elements are considered.

11.8 Reasoning with Rules

We already saw examples showing the close connections between argumentation research in artificial intelligence and legal applications. Since argumentation is an everyday task of professional lawyers, this is not unexpected. An institutional reason, however, is that there exists an interdisciplinary research field, called *artificial intelligence and law*,[26] in which because of the nature of law, the topic of argumentation has been given a great deal of attention. Early work in that field (e.g., McCarty 1977; Gardner 1987) already showed the intricacies and special characteristics of legal argumentation. Thorne McCarty (1977) attempted to formalize the detailed reasoning underlying a US Supreme Court case. Anne Gardner (1987) proposed a system aimed at what she called *issue spotting*. In a legal case, there is an *issue* when no rule applies or when conflicting rules apply. In this section, we pay special attention to the work inspired by developments in

[25] A systematic overview of argumentation dialogue models of negotiation has been provided by Rahwan et al. (2003).

[26] The primary journal of the field of AI and Law is *Artificial Intelligence and Law*, with the biennial ICAIL and annual JURIX as the main conferences.

non-monotonic logic that has been carried out, mostly in the mid-1990s, regarding *reasoning with (legal) rules*.

Prakken's (1997) book *Logical Tools for Modelling Legal Argument* provides an extensive and careful treatment of the contributions of techniques from non-monotonic logic to the formal modelling of legal reasoning.[27] The formal tools presented by Prakken have gradually evolved into the ASPIC + model (Prakken 2010) (see Sect. 11.3.3). Parts of the material were developed in close collaboration with Sartor (e.g., Prakken and Sartor 1996, 1998; see also the excellent resource Sartor 2005).

The following example shows how Prakken models a case in contract law (1997, p. 171). The example concerns the defeasible rule that contracts only bind the contracting parties (d_1) and a defeasible, possibly contravening, rule specifically for contracts that concern the lease of a house, saying that such contracts also bind future owners of the house (d_2). Another exception is added by a defeasible rule saying that, even in the case of a house lease, when a tenant agrees to make such a stipulation, only the contracting parties are bound (d_3). The factual statements f_{n1} and f_{n2} say, respectively, (1) that a house lease is a special kind of contract and (2) that binding only the contracting parties and binding also future owners of a house do not go together.

d_1: x is a contract $\Rightarrow$ x only binds its parties.
d_2: x is a lease of house $y \Rightarrow x$ binds all owners of y.
d_3: x is a lease of house $y \wedge$ tenant has agreed in x that x only binds its parties $\Rightarrow x$ only binds its parties.
f_{n1}: $\forall x \forall\, y$(x is a lease of a house $y \rightarrow x$ is a contract).[28]
f_{n2}: $\forall x \forall\, y \neg$($x$ only binds its parties $\wedge$ x binds all owners of y).

When there is a contract about the lease of a house, there is an apparent conflict, since both d_1 and d_2 seem to apply. In the system, the application of d_2 blocks the application of d_1, using a mechanism of specificity defeat (see Sect. 11.5). In a case where also the condition of d_3 is fulfilled, namely, when the tenant has agreed that the lease contract only binds the contracting parties, the application of rule d_3 blocks the application of rule d_2, which in that case does no longer block the application of d_1.

Prakken uses elements from classical logic (for instance, classical connectives and quantifiers) and non-monotonic logic (defeasible rules and their names) and shows how they can be used to model rules with exceptions, as they occur prominently in the law. He treats, for instance, the handling of explicit exceptions, preferring the most specific argument, reasoning with inconsistent information, and reasoning about priority relations.

[27] The book is based on Prakken's (1993) doctoral dissertation.

[28] "∀x . . ." stands for "for every entity x it holds that . . .." Similarly, for "∀y . . .." See also Sect. 6.2 of this volume.

In the same period, Hage developed *reason-based logic* (Hage 1997; see also Hage 2005).[29] Hage presents reason-based logic as an extension of first-order predicate logic in which reasons play a central role. Reasons are the result of the application of rules.[30] Treating them as individuals allows the expression of properties of rules. Whether a rule applies depends on the rule's conditions being satisfied but also on possible other reasons for or against applying the rule. Consider for instance the rule that thieves are punishable:

punishable: thief(x) ⇒ punishable(x)

Here "punishable" before the colon is the rule's name. When John is a thief (expressed as thief(john)), the rule's applicability can follow:

Applicable(thief(john) ⇒ punishable(john))

This gives a reason that the rule ought to be applied. If there are no reasons against the rule's application, this leads to the obligation to apply the rule. From this it will follow that John is punishable.

A characteristic aspect of reason-based logic is that it models the weighing of reasons. In this system, there is no numerical mechanism for weighing; rather it can be explicitly represented that certain reasons for a conclusion outweigh the reasons against the conclusion. When there is no weighing information, the conflict remains unresolved and no conclusion follows.

Like Prakken, Hage uses elements from classical logic and non-monotonic logic. Because of the emphasis on philosophical and legal considerations, reason-based logic is less like a formal logical system and more like a semiformal system for the representation of the ways of reasoning in the domain of law. Where Prakken's book remains closer to the field of AI, Hage's book reads more like a theoretical essay in philosophy or law.

Reason-based logic has been applied, for instance, to a well-known distinction made by the legal theorist Dworkin (1978): whereas legal *rules* seem to lead directly to their conclusion when they are applied, legal *principles* are not direct and merely give rise to a reason for their conclusion. Only a subsequent weighing of possibly competing reasons leads to a conclusion. Different models of the distinction between rules and principles in reason-based logic have been proposed. Hage (1997) follows Dworkin and makes a strict formal distinction, whereas Verheij et al. (1998) show how the distinction can be softened by presenting a model in which rules and principles are the extremes of a spectrum.

Loui and Norman (1995) have argued that there is a calculus associated with what they call the *compression of rationales*, i.e., the combination and adaptation of the rules underlying arguments which are akin to Toulmin's warrants. They give the

[29] Reason-based logic exists in a series of versions, some introduced in collaboration with Verheij (e.g., Verheij 1996a).

[30] We shall simplify Hage's formalism a bit by omitting the explicit distinction between rules and principles.

following example of a compression of rules (rationales). When there is a rule "vehicles used for private transportation are not allowed in the park" and also a rule "vehicles are normally for private transportation," then a two-step argument based on these two rules can be shortened when the so-called compression rationale "no vehicles in the park," based on these two rules, is used.

11.9 Case-Based Reasoning

Reasoning with rules (Sect. 11.8) is often contrasted with *case-based reasoning*. Whereas the former is about following rules that describe existing conditional patterns, the latter is about finding relevantly similar examples that, by analogy, can suggest possible conclusions in new situations. In the domain of law, rule-based reasoning is associated with the application of legal statutes, and case-based reasoning with the following of precedents. The contrast can be appreciated by looking at the following two examples:

> *Art. 300 of the Dutch Criminal Code*
> 1. Inflicting bodily harm is punishable with up to two years of imprisonment or a fine of the fourth category.
> 2. When the fact causes grievous bodily harm, the accused is punished with up to four years of imprisonment or a fine of the fourth category.
> 3. [...]

> *Dutch Supreme Court July 9, 2002, NJ 2002, 499*
> Theft requires the taking away of a good. Can one steal an already stolen car? The Supreme Court's answer is: yes.

The first example is an excerpt from a statutory article expressing a material rule of Dutch criminal law, stating the kinds of punishment associated with inflicting bodily harm. The levels of punishment depend on specific conditions, with more severe bodily harm being punishable with longer imprisonment. The second example is a (very) brief summary of a Supreme Court decision. In this case, an already stolen car was stolen from the thief. One of the statutory requirements of the crime theft is that a good is taken away, and here the car was already taken away from the original owner of the car. The new legal question was addressed whether stealing from the original thief can count as theft from the car's owner. In other words, can an already stolen car still be taken away from the original owner? Here the Supreme Court decided that stealing a stolen car can count as theft since the original ownership is the deciding criterion; it does not matter whether a good is actually in the control of the owner at the time of theft.

In case-based reasoning, the *stare decisis* doctrine is leading: when deciding a new case, one should not depart from an earlier, relevantly similar decision, but decide analogously. In the field of AI and law, Kevin Ashley's HYPO system (1990) counts as a milestone in the study of case-based reasoning.[31] In HYPO, cases

[31] See also Rissland and Ashley (1987), Ashley (1989), and Rissland and Ashley (2002).

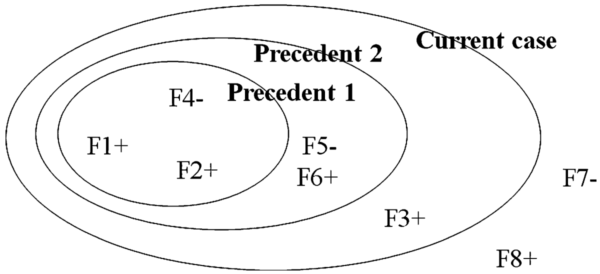


Fig. 11.12 Factors in two precedent cases and the current case

are treated as sets of factors, where factors are generalized facts pleading for or against a case. Consider the following example about an employee who has been dismissed by his employer and aims to void (i.e., cancel) the dismissal.[32]

Issue:
Can a dismissal be voided?
Precedent case:
+ The employee's behavior was always good.
- There was a serious act of violence.
Outcome:
+ (voided)
Current case:
+ The employee's behavior was always good.
- There was a serious act of violence.
+ The working atmosphere was not affected.
Outcome: ?

There is a precedent case with one factor pleading for voidance (the good behavior) and one pleading against voidance (the violence). In this precedent case, it was decided that voidance was in place. In the current case, the same factors apply, but there is also one additional factor pleading for voidance, namely, that the working atmosphere was not affected. One could say that the decision taken in the precedent case is even more strongly supported in the current case. As a result, in HYPO and similar systems, the suggested conclusion is that also in the current case, voidance of the dismissal would be called for.

The example in Fig. 11.12 shows that factors can be handled formally without knowing what they are about. There is a first precedent with pro-factors F1 and F2 and a con-factor F4. The second precedent has as additional factors a con-factor F5 and a pro-factor F6. The current case has all these factors and one more pro-factor F3. The domain also contains con-factor F7 and pro-factor F8 which do not apply to these cases.

Assume now that the first precedent was decided negatively and the second positively. The second precedent is more on point, in the sense that it shares more

[32] The example is inspired by the case material used by Roth (2003).

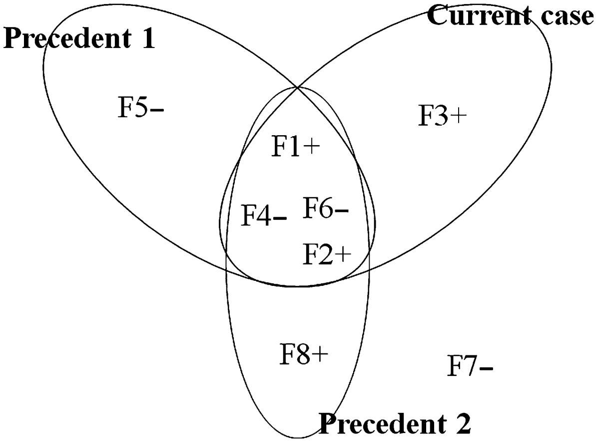


Fig. 11.13 A different constellation of precedents

factors with the current case than the first precedent. Since the current case even has an additional pro-factor, it is suggested that the current case should be decided positively, in analogy with precedent 2. Precedents do not always determine the outcome of the current case. For instance, if the second precedent had been decided negatively, there would be no suggested outcome for the current case, since pro-factor F3 may be or may not be strong enough to turn the case.

Another formal example is shown in Fig. 11.13. When both precedents have been decided positively, the suggested outcome for the current case is also positive. Precedent 1 can be followed because its support for a positive decision is weaker than that of the current case: the precedent has an additional con-factor, and the current case an additional pro-factor. Precedent 2 cannot be followed since F8 may be or may not be a stronger pro-factor than F3.

HYPO's aim is to form arguments about the current case, without determining a decision. This is made explicit in its model of 3-ply arguments. In HYPO's 3-ply model, the first argument move ("ply"), by the Proponent, is the citing of a precedent case in analogy with the current case. The analogy is based on the shared factors. The second argument move, by the Opponent, responds to the analogy, e.g., by distinguishing between the cited precedent case and the current case, pointing out differences in relevant factors, or by citing counterexamples. The third argument move, again by the Proponent, responds to the counterexamples, e.g., by making further distinctions.

HYPO's factors not only have a side (pro or con) associated with them, but can also come with a dimension pertaining in some way to the strength of the factor. This allows the citation of cases that share a certain factor, but have this factor with a different strength. For instance, by the use of dimensions, the good behavior of the employee (of the first informal example) can come in gradations, say from good via very good to excellent.

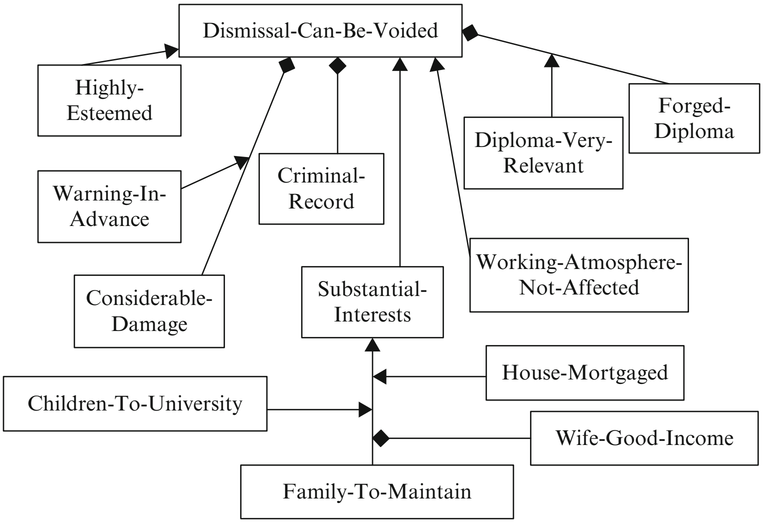


Fig. 11.14 An entangled factor hierarchy (Roth 2003)

Vincent Aleven extended the HYPO model by the use of a factor hierarchy that allowed modelling of factors with hierarchical dependencies (Aleven 1997; Aleven and Ashley 1997a, b). For instance, the factor that one has a family to maintain is a special case of the factor that one has a substantial interest in keeping one's job. Inspired by Verheij's DefLog model (2003a), which allowed for reasoning about support and attack (Sect. 11.5.5), Roth (2003) developed case-based reasoning based on an entangled factor hierarchy (Fig. 11.14). For instance, the relevance of the factor that one has a family to maintain is strengthened by one's having children that go to university and weakened by one's having a wife with a good income. A factor hierarchy allows new kinds of argument moves by making it possible to downplay or emphasize a distinction. For instance, the factor of having a family to maintain can be downplayed by pointing out that one has a partner with a good income, or emphasized by mentioning that one has children going to university.

Proposals have been made to combine case-based and rule-based reasoning. For instance, Branting's GREBE model (1991, 2000) aims to generate explanations of decisions in terms of rules and cases. Both rules and cases can serve as warrants for a decision. Branting extends Toulmin's approach to warrants by using a so-called warrant reduction graph, in which warrants can be special cases of other warrants. Prakken and Sartor (1998) have applied their model of rule-based reasoning (Prakken and Sartor 1996; see also Sect. 11.8) to the setting of case-based reasoning. Analogizing and distinguishing are connected to the deletion and addition of rule conditions that describe past decisions.

11.10 Values and Audiences

Trevor Bench-Capon (2003) has developed a model of the values underlying arguments.[33] In this endeavor he refers to Perelman and Olbrechts-Tyteca's new rhetoric:

> If men oppose each other concerning a decision to be taken, it is not because they commit some error of logic or calculation. They discuss apropos the applicable rule, the ends to be considered, the meaning to be given to values, the interpretation and characterisation of facts. (Perelman & Olbrechts-Tyteca, 1969, p. 150)

Because of the character of real-life argumentation, it is not to be expected that cases will be conclusively decided. Bench-Capon therefore aims to extend formal argumentation models by the inclusion of the values of the audiences addressed. This allows him to model the persuasion of an audience by means of argument.

Bench-Capon (2003) uses Dung's (1995) abstract argumentation frameworks (Sect. 11.4) as a starting point. He defines a value-based argumentation framework as a framework in which each argument has an associated (abstract) value. The idea is that values associated with an argument are promoted by accepting the argument. For instance, in a parliamentary debate about a tax raise, it can be argued that accepting the raise will promote the value of social equality, while the value of enterprise is demoted. In an audience-specific argumentation framework, the preference ordering of the values can depend on an audience. For instance, the Labour Party may prefer the value of social equality, and the Conservative Party that of enterprise.

Bench-Capon continues to model *defeat for an audience*: an argument A defeats an argument B for audience *a* if A attacks B and the value associated with B is not preferred to the value associated with A for audience *a*. In his model, an attack succeeds, for instance, when the arguments promote the same value, or when there is no preference between the values. Dung's notions of argument acceptability, admissibility, and preferred extension are then redefined relative to audience attack.

Bench-Capon uses a value-based argumentation framework with two values "red" and "blue" as an example (Fig. 11.15). The underlying abstract argumentation framework is the same as that in Fig. 11.9. In its unique preferred extension (which is also grounded and stable), A and C are accepted and B is rejected. For an audience preferring "red," defeat for the audience coincides with the underlying attack relation. In the preferred extension for an audience preferring "red," therefore, A and C are accepted and B is rejected. However, for an audience preferring "blue," A does not defeat B. But, for such an audience, B still defeats C. For a "blue"-preferring audience, A and B are accepted and C is not.

Bench-Capon illustrates value-based argumentation by considering the case of a diabetic who almost collapses into a coma by lack of insulin and therefore takes

[33] In AI and law, the importance of the modelling of the values and goals underlying legal decisions was already acknowledged by Berman and Hafner (1993).

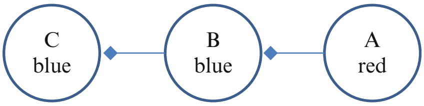


Fig. 11.15 A value-based argumentation framework with two values (Adapted from Bench-Capon 2003)

another diabetic's insulin after entering her house. He analyzes the case by discussing the roles of the value of property right infringement as opposed to that of saving one's life.

Bench-Capon and Sartor (2003) have used the value-based perspective in a treatment of legal reasoning that combines rule-based and case-based reasoning (see Sects. 11.8 and 11.9). Legal reasoning takes the form of constructing and using a theory that explains a decision in terms of the values promoted and demoted by the decision. Precedent decisions have the role of revealing preferences holding between factors. This is similar to the role of precedents in HYPO that reveal how the factors in a precedent case are weighed. In Bench-Capon and Sartor's approach, the factor preferences in turn reveal preferences between values. The resulting preferences can then be used to decide new cases.

11.11 Argumentation Support Software

When studying argumentation from an artificial intelligence perspective, it can be investigated how software tools can perform or support argumentative tasks. Some researchers in the field of argumentation in AI have openly addressed themselves to building an artificial arguer. The most prominent among them is John Pollock (see also Sect. 11.3.2), who titled one of his books about his OSCAR project ambitiously *How to Build a Person* (Pollock 1989).[34] Most researchers, however, have not aimed at realizing the grand task of addressing the so-called "strong AI" problem of building an intelligent artifact that can perform any intellectual task a human being can. Instead of building software mimicking human argumentative behavior, the more modest aim of supporting humans performing argumentative tasks was chosen. A great deal of research has been aimed at the construction of argumentation support software. Here we discuss three recurring themes: argument diagramming in software, the integration of rules and argument schemes, and argument evaluation.[35]

[34] The book's subtitle adds modestly: A Prolegomenon.

[35] The reviews by Kirschner et al. (2003), Verheij (2005b), and Scheuer et al. (2010) provide further detail about argumentation support software.

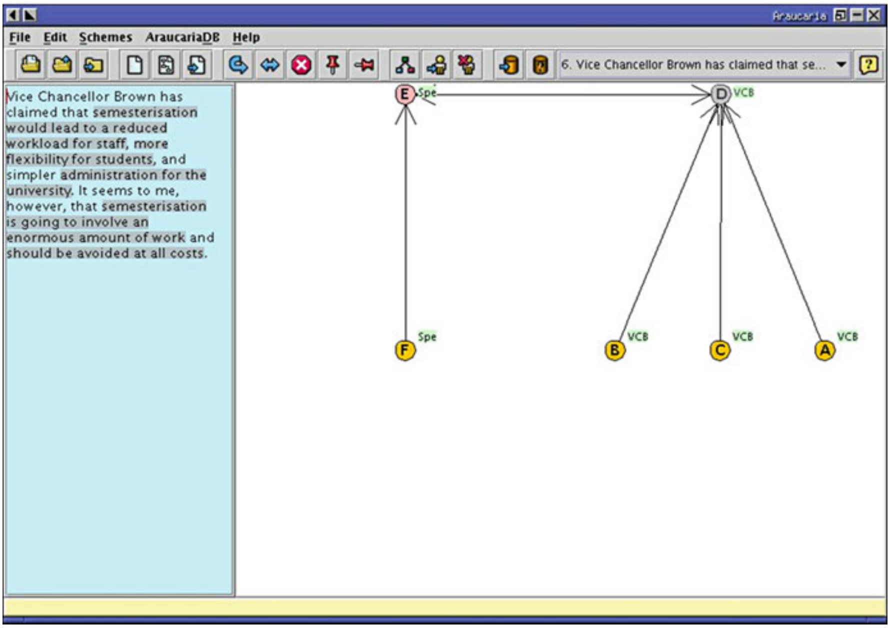


Fig. 11.16 *Boxes* and *arrows* diagramming: the Araucaria system (Source: http://araucaria.computing.dundee.ac.uk/, 25 July 2012)

11.11.1 Argument Diagramming in Software

In the literature on argumentation support software, much attention has been paid to argument diagramming. Different kinds of argument diagramming styles have been proposed, many inspired by non-computational research on argument diagrams. We shall discuss three styles: boxes and arrows, boxes and lines, and nested boxes.

The first style of argument diagramming uses boxes and arrows. Argumentative statements are enclosed in boxes, and their relations are indicated by arrows. A common use of arrows is to indicate the support relation between a reason and a conclusion. An example of a software tool that uses boxes and arrows diagrams is the Araucaria tool by Chris Reed and Glenn Rowe (2004) (Fig. 11.16). The Araucaria tool has been designed for the analysis of written arguments. Vertical arrows indicate reasons and their conclusions, and horizontal bidirectional arrows indicate conflicts between statements. The Araucaria software was one step in the development by the Dundee Argumentation Research Group, led by Reed, of open source argumentation software. For this purpose, a representation format, called the *argument markup language* (AML), has been developed that allows for the exchange of arguments and their analyses using contemporary Internet technology. The format also allows for the exchange of sets of argument schemes (see Sect. 11.6) that can be used for argument analysis. Other developments concerning machine-readable argument representation formats are the *argument* *interchange format* (Chesñevar et al. 2006) and *ArgDF*, a proposal for a language allowing for a World Wide Argument Web (Rahwan et al. 2007). One aim of the latter work is to develop classification systems for arguments, using ontology development techniques in Artificial Intelligence. In AI, an "ontology" is a systematic conceptualization of a domain, often taking the form of a hierarchical system of concepts and their relations.

Another example of a system using boxes and arrows is the Hermes system (Karacapilidis and Papadias 2001), an extension of the Zeno system (Gordon and Karacapilidis 1997). Both Hermes and Zeno have been inspired by the IBIS approach. In IBIS, an abbreviation of Issue-Based Information Systems (Kunz and Rittel 1970), problems are analyzed in terms of issues, questions of fact, positions, and arguments. The focus is on what Rittel and Webber (1973) call *wicked problems*: problems with no definitive formulation and no definitive solutions. Hence a goal of IBIS and systems such as Hermes and Zeno is to support the identification, structuring, and settling of issues.

The second style of argument diagramming uses boxes and lines. In a boxes and lines style of argument diagramming, argumentative statements are depicted in boxes and their relations are indicated by (undirected) lines between them. This diagramming style abstracts from the directionality between statements, e.g., from a reason to a conclusion, or from a cause to an event. An example of a tool using the boxes and lines style is the Belvedere system (Suthers et al. 1995; Suthers 1999). A goal of the system was to stimulate the critical discussion of science and public policy issues by middle school and high school students, taking the cognitive limitations of the intended users into account. Such limitations include difficulty in focusing attention, lack of domain knowledge, and lack of motivation. In early versions, the diagrams were richly structured: there were links for support, explanation, causation, conjunction, conflict, justification, and undercutting. Link types could be distinguished graphically and by label. To prevent unproductive discussions about which structure to use, the graphical representation was significantly simplified in later versions (Suthers 1999). Two types of statements were distinguished, data and hypotheses, and two link types, expressing a consistency and an inconsistency relation between statements. Figure 11.17 shows an example of a Belvedere screen using an even further simplified format with one statement type and one link type.

The third style of argument diagramming uses nested boxes. In this style, too, the argumentative statements are enclosed in boxes, but their relationships are indicated by the use of nesting. An example of the use of nested boxes is the Room 5 tool designed by Loui, Norman, and a group of students (Loui et al. 1997). The Room 5 system aimed at the collaborative public discussion of pending Supreme Court cases. It was Web based, which is noteworthy as the proposal predates Google and Wikipedia. In its argument diagramming format, a box inside a box expresses support, and a box next to a box indicates attack. In the argument depicted in the Room 5 screen shown in Fig. 11.18, for instance, the punishability of John is supported by the reason that he has stolen a CD and attacked by the reason that he is a minor first offender.

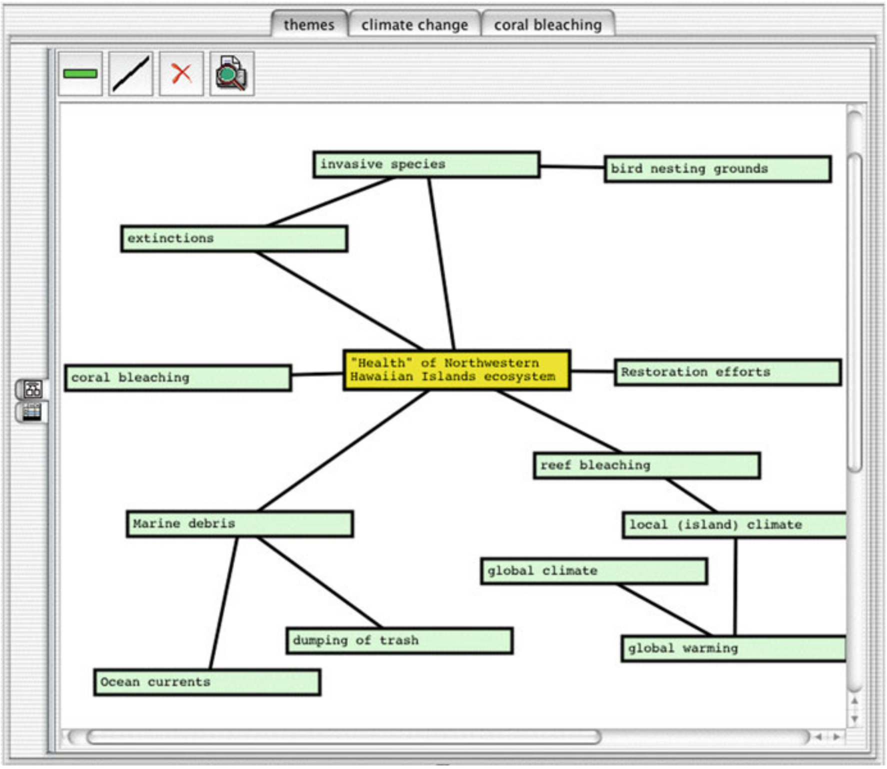


Fig. 11.17 *Boxes* and *lines* diagramming: the Belvedere 4.1 system (Source: http://belvedere.sourceforge.net/, 25 July 2012)

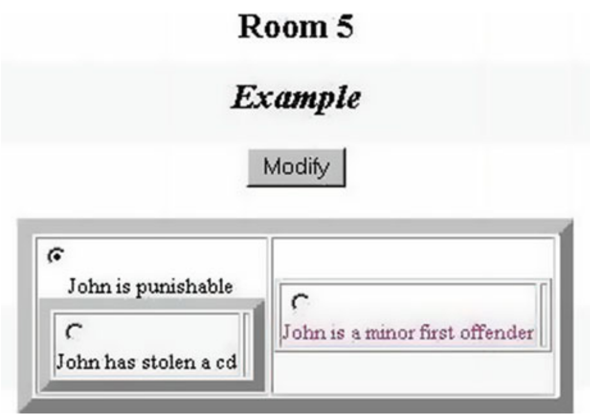


Fig. 11.18 *Nested boxes* diagramming: the Room 5 system (Screenshot of Room 5, as shown in Verheij (2005b). See also Bench-Capon et al. (2012))

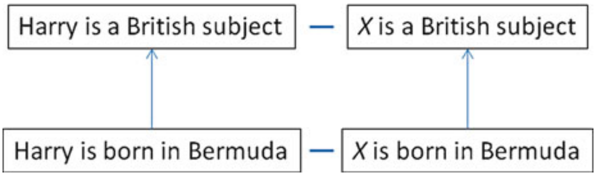


Fig. 11.19 An elementary argument step as an instance of a schematic argument

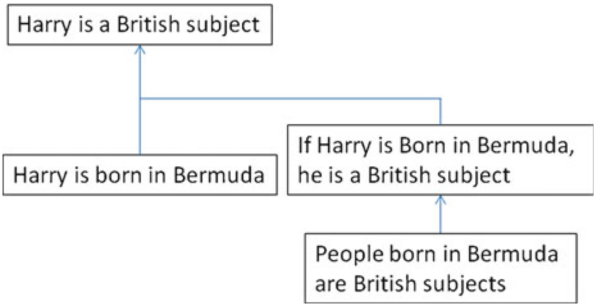


Fig. 11.20 Using a conditional sentence

11.11.2 Integration of Rules and Argument Schemes

The integration of rules and argument schemes in argument diagramming software has been addressed in different ways: by the use of schematic arguments, conditional sentences, nested arrows, and rule nodes. Consider, for instance, the elementary argument that Harry is a British subject because he is born in Bermuda (borrowed from Toulmin) and its underlying rule (or "warrant" in Toulmin's terminology) that people born in Bermuda are British subjects.

A first approach is to consider such an argument as an instance of a scheme that abstracts from the person Harry in the argument. In Fig. 11.19, an associated schematic argument is shown to the right of the argument about Harry. In the schematic argument, *X* appears as a variable that serves as the placeholder of someone's name. In software, the schematic argument is normally not shown graphically. For instance, in Araucaria, the schematic arguments are text files and can be used to annotate argument instances. The schematic arguments appear at a level separate from the arguments themselves; hence, they constitute a kind of meta-arguments. As a result, they are not themselves the subject of debate.

A second approach uses conditional sentences. The conditional sentence that expresses the connection between reason and conclusion is made explicit as an auxiliary premise. This conditional sentence can then be supported by further arguments, such as a warrant (as in Fig. 11.20) or a backing. This approach is, for instance, proposed in the *Rationale*[36]*tool* developed by van Gelder and his collaborators (van Gelder 2007).

A third approach uses nested arrows. The arrows are treated as graphical expressions of the connection between the reason and conclusion and can hence

[36] http://rationale.austhink.com/

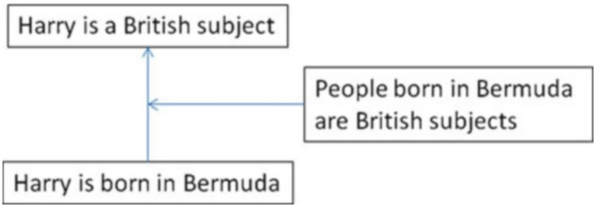


Fig. 11.21 *Nested arrows*

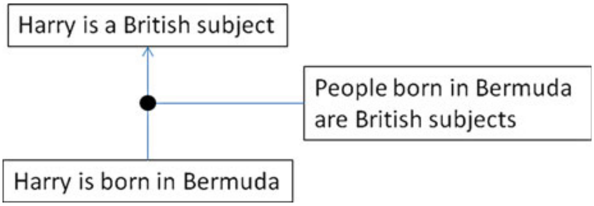


Fig. 11.22 Rule nodes

be argued about. In Fig. 11.21, for instance, the warrant has been supplied as support for the connection between reason and conclusion. This approach has a straightforward generalization when support and attack are combined (Sect. 11.5.5). The ArguMed tool developed by Verheij (2005b) uses this approach.

A variation of the nested arrows approach uses rule nodes (Fig. 11.22), instead of nested arrows. The *AVERs tool* (van den Braak et al. 2007) uses this approach.

11.11.3 Argument Evaluation

In argumentation software, different strategies for argument evaluation have been implemented. Some tools choose to leave argument evaluation as a task for the user of the system. For instance, in the Rationale system (van Gelder 2007), a user can indicate which claims follow or do not follow given the reasons in the diagram. Specific graphical elements are used to show the user's evaluative actions.

In several other systems, some form of automatic evaluation has been implemented. Automatic evaluation algorithms can be logical, or numeric.

Logical evaluation algorithms in argumentation support tools have been grounded in versions of argumentation semantics (see Sect. 11.4.1). For instance, ArguMed (Verheij 2005b) computes a version of stable semantics. Consider, for instance, Pollock's example of an undercutting defeater about red lights (see Sect. 11.3.2). ArguMed's evaluation algorithm behaves as expected: when the reason that the object looks red is assumed, the conclusion that the object is red will be justified, but that will no longer be the case when the defeater is added that the object is illuminated by a red light. A typical property of logical evaluation algorithms is reinstatement: when a defeating attacker of an initial argument is successfully attacked, the initial argument will no longer count as defeated and therefore be reinstated.

Numeric evaluation algorithms have been based on the numeric weights of the reasons supporting and attacking conclusions. A weight-based numeric evaluation algorithm has, for instance, been implemented in the Hermes system (Karacapilidis and Papadias 2001). In Hermes, positions can be assigned a numeric score by adding the weights of active pro-positions and subtracting the weights of active con-positions. A proof standard can be used to determine an activation label of a position. In the proof standard called *preponderance of evidence*, for instance, a position is active when the active pro-positions outweigh the active con-positions.

A numeric evaluation algorithm of a different kind has been implemented in the so-called "Convince me" system. It uses ECHO, which is a connectionist version of Thagard's (1992) theory of explanatory coherence. In Convince me, statements are assigned numerical values by a stepwise constraint satisfaction algorithm. In the algorithm, incremental changes of the default weights of a statement are made by considering the excitatory and inhibitory links connected to a statement. When changes become too small to be taken into account (or computation is taking too long), the algorithm stops.

11.12 Burden of Proof, Evidence, and Argument Strength

Some arguments are more successful than others. An argument can meet or not meet the burden of proof fitting the circumstances of the debate. An argument can be founded on better evidence than another. An argument can also be stronger than another. In this section, we address the topics of *burden of proof, evidence*, and *argument strength.*

11.12.1 Burden of Proof and Evidence

The topic of burden of proof is strongly connected to the dialogical setting of argumentation. A burden of proof is assigned to a party in an argumentative dialogue when the quality of the arguments produced in the dialogue depends in part on whether the arguments produced by that party during the dialogue meet certain constraints. Such constraints can be procedural, e.g., requiring that a counterargument is met by a counterattack, or material, e.g., requiring that an argument is sufficiently strong in the light of the other arguments. Constraints of the latter, material, and non-procedural type are also referred to as *proof standards*.

The topic of burden of proof is especially relevant in the law, as argumentation in court is often constrained by burden of proof constraints. As a result, in legal theory the topic has been studied extensively. The topic has also been addressed in AI approaches to argumentation, in particular by researchers connected to the field of AI and law (see also Sect. 11.7.1). In the Carneades argumentation model (Gordon et al. 2007), for instance, statements are categorized using three proof standards:

SE (scintilla of evidence). A statement meets this standard if and only if it is supported by at least one defensible pro argument.
BA (best argument). A statement meets this standard if and only if it is supported by some defensible pro argument with priority over all defensible con arguments.
DV (dialectical validity). A statement meets this standard if and only if it is supported by at least one defensible pro argument and none of its con arguments are defensible.

A theme related to proof standards is argument accrual. What happens when there are several arguments for a conclusion? See Sect. 11.3.3, where research addressing the relation between argument defeat and accrual is discussed.

AI models of argumentation have been helpful in clarifying distinctions made in legal theory. Prakken and Sartor in particular have in a series of articles (Prakken and Sartor 2007, 2009) contributed to the explication of different forms of burden of proof. They distinguish a *burden of persuasion*, a *burden of production*, and a *tactical burden*. A burden of persuasion requires that a party prove a statement to a specified degree (the standard of proof) or run the risk of losing on the issue at the end of the debate. A burden of production has been assigned to a party when the party is required by law to provide evidence for a certain claim. Burdens of persuasion and burdens of production are assigned by the applicable law. The tactical burden of proof depends on a party's own assessment of whether sufficient grounds have been adduced about a claim made by the party. Prakken and Sartor connect these different notions to a formal dialogue model of argumentation.

11.12.2 Probability and Other Quantitative Approaches to Argument Strength

Argument strength can be considered by using quantitative approaches. For instance, a conditional probability $p(H|E)$, expressing the probability of a hypothesis H given the evidence E, can be interpreted as a measure of the strength of the argument for the hypothesis based on the evidence. The idea is that higher values of $p(H|E)$ make H more strongly supported when given E. This interpretation of argument strength is associated with what is called *Bayesian epistemology* (Talbott 2011). Bayesian epistemology provides in the following way an interpretation of the relevance of additional evidence, say E': additional evidence E' strengthens the argument E for H when $p(H|E \wedge E') > p(H|E)$. In this interpretation, Bayes' theorem:

$$p(H|E) = p(E|H) \times p(H)/p(E)$$

connects the strength of the argument from E to H and that of the argument from H to E, thereby reversing the direction of the arrow. This relation is helpful when the values of $p(E|H)$, $p(H)$, and $p(E)$ are available or when they are more easily established than $p(H|E)$ itself. Bayesian epistemology also provides a perspective on the comparison of hypotheses given additional evidence. When there are two hypotheses H and H', the odds form of Bayes' theorem can be used to update the odds of the hypotheses in light of new evidence E. The following relation shows how the prior odds p(H)/p(H') is connected to the posterior odds p(H|E)/p(H'|E):

$$\mathrm{p}(H|E)/\mathrm{p}(H'|E) = (\mathrm{p}(H)/\mathrm{p}(H')) \times (\mathrm{p}(E|H)/\mathrm{p}(E|H'))$$

This formal relation is helpful when the prior odds p(H)/p(H') and the values of p(E|H) and p(E|H') are available.

Pollock has argued against a probabilistic account of argument strength (e.g., Pollock 1995, 2006, 2010), referring to this position as "generic Bayesianism" or "probabilism." Pollock argues that in a probabilistic account, we would be justified in believing a mathematical theorem even before it is proven. This is especially absurd in cases such as Fermat's last theorem that remained a conjecture for centuries before Wiles finally could complete a proof in the 1990s. Fitelson (2010) defends a probabilistic account against this and other criticisms advanced by Pollock.

Zukerman et al. (1998) have discussed the possibility of generating arguments from Bayesian networks, which are a widely studied tool for the representation of probabilistic information. Riveret et al. (2007) consider success in argument games in connection with probability. Dung and Thang (2010) have presented an approach to probabilistic argumentation in the setting of dispute resolution. Verheij (2012) has proposed a formal theory of defeasible argumentation in which logical and probabilistic properties are connected. Hunter (2013) discusses a model of deductive argumentation with uncertain premises.

11.12.3 Evidence and Inference to the Best Explanation

When an argument is aimed at establishing the truth, empirical evidence can be used to support alleged facts. For instance, a witness's testimony can provide evidence for the claim that the suspect was at the scene of a crime, a clinical test can provide evidence against a medical diagnosis, and the outcome of a laboratory experiment can be evidence confirming (or falsifying) a psychological phenomenon. The conclusions based on the available evidence can be regarded as hypothetical explanations for the occurrence of the evidence. As a result, reasoning on the basis of evidence is a specimen of what Peirce referred to as *abductive reasoning*, or *inference to the best explanation*: reasoning that goes from data describing something to a hypothesis that best explains or accounts for the data (Josephson and Josephson 1996, p. 5). Josephson and Josephson conceive of inference to the best explanation as a kind of argument scheme (see Sect. 11.6):

D is a collection of data (facts, observations, givens).
H explains D (would, if true, explain D).
No other hypothesis can explain D as well as H does.
Therefore, H is probably true.
(Josephson and Josephson 1996, p. 5)

The explanatory connection between D and H is often regarded as going against the causal direction. For instance, a causal and expectation-evoking rule “If there is a fire, then there is smoke” can be used to infer, or argue for, the effect “there is smoke” after observing the cause “there is fire.” The causal rule has an evidential, explanation-evoking counterpart, “If there is smoke, then there is a fire,” that can be used to infer (argue for) the explanation “there is a fire” after observing “there is smoke.” Arguments based on causal or evidential rules are typically defeasible: not all fires generate smoke, and not all smoke stems from a fire.

In artificial intelligence, the distinction between causal and evidential rules has been emphasized by Pearl (1988, p. 499f.). He argues that special care is needed when mixing causal and evidential reasoning. To make his point, Pearl uses the following examples:

> Bill showed slight difficulties standing up, so I believed he was injured.
> Harry seemed injured, so I believed he would be unable to stand up.

The former uses the evidential pathway from the observation of Bill’s difficulties in standing up to the explanation that he is injured and the latter the reverse causal pathway from the observation of Harry’s injuries to the effect that he is unable to stand up. The question is then addressed whether it is likely that Bill and Harry are drunk, drunkenness being a second cause for difficulties in standing up, independent from injury. Both Bill’s and Harry’s intoxicated state could be argued for using the evidential rule “If someone has difficulties standing up, then he may be drunk.” However, for Bill the conclusion that he may be drunk seems more likely than for Harry, since for Bill both explanations for his difficulties in standing up, namely, injury or being drunk, seem to be reasonable, whereas for Harry drunkenness is a less likely hypothesis now that an injury has been observed. The distinction between causal and evidential rules has played a central role in Pearl’s thinking about causality (Pearl 2000/2009), in relation with the probabilistic modelling tool of Bayesian Networks (see Jensen and Nielsen 2007; Kjaerulff and Madsen 2008). Bayesian Networks have been connected to the modelling of argumentation with legal evidence by Hepler et al. (2007) and by Fenton et al. (2012) (see also Taroni et al. 2006).

The distinction between causal and evidential rules has been used in the formalized hybrid argumentative-narrative model of reasoning with evidence developed by Bex and his colleagues (Bex et al. 2010; Bex 2011). In this model, the elements of a scenario, or narrative, describing how a crime may have been committed, can be supported by arguments grounded in the available evidence. Causal connections between the elements of a scenario contribute to its coherence. It is possible that more than one scenario is available, each scenario with different evidential support and a different kind of coherence. Bex and Verheij (2012) have presented the argumentative-narrative model in terms of argument schemes and their associated critical questions (see Sect. 11.6).

11.13 Applications and Case Studies

A first reason for the popularity of argumentation research in the field of artificial intelligence is that it has led to theoretical advances. A second reason is that the theoretical advances have been corroborated by a variety of interesting applications and case studies, including advances in natural language processing. We give some examples.

Fox and Das (2000) provided a book-length study of AI technology in medical diagnosis and decision making, with much emphasis on the argumentative aspects (see also Fox and Modgil 2006, where argumentation-based decision making is used to extend the Toulmin model). Aleven and Ashley (1997a, b) developed a case-based argumentation tool that was empirically tested for its effects on learning. Buckingham Shum and Hammond (1994) approached the design of artifacts such as software as an argumentation problem. Grasso et al. (2000) worked on argumentative conflict resolution in the context of health promotion. Teufel (1999) has worked on the problem of automatically estimating a sentence's role in argumentation, using a model of seven text categories called *argumentative zones*. Mochales Palau and Moens (2009) developed software for the mining of argumentative elements in legal texts. Hunter and Williams (2010) investigated the aggregation of evidence in a healthcare setting. Grasso (2002) and Crosswhite et al. (2004) have worked on the computational modelling of rhetorical aspects of argument. Reed and Grasso (2007) have collected argumentation-oriented research using natural language techniques. They discuss, for instance, the generation of argumentative texts as studied by Elhadad (1995), Reed (1999), Zukerman et al. (1998), and Green (2007).

Rahwan and McBurney (2007) edited a special issue on argumentation technology of the journal *IEEE Intelligent Systems*. Application areas addressed in the issue are medical decision making, emotional strategies to persuade people to follow a healthy diet, ontology engineering, discussion mediation, and Web services. In the 2012 edition of the COMMA conference proceedings series on the computational modelling of argument, a separate section was devoted to innovative applications. The topics included automatic mining of arguments in opinions, a learning environment for scientific argumentation, semiautomatic analysis of online product reviews, argumentation with preferences in the setting of eco-efficient biodegradable packaging, hypothesis generation from cancer databases, sense making in policy deliberation, music recommendation, and argumentation about firewall policy. For applications focusing on argumentation support and facilitation, the reader is referred to Sect. 11.11.

In the domain of AI and law, theories and systems were developed and tested by the use of case studies. For instance, McCarty (1977, 1995) analyzed a seminal case in US tax law (Eisner v. Macomber, 252 U.S. 189 (1920)). In that case, the US Supreme Court decided that a federal rule of tax law was invalid. McCarty's aims were set high, namely, to build a software implementation that could handle a number of elusive, argumentative aspects of legal reasoning, illustrated in the majority opinion and dissenting opinions concerning the issues in this case. Quoting McCarty (1995):

1. Legal concepts cannot be adequately represented by definitions that state necessary and sufficient conditions. Instead, legal concepts are incurably "open-textured."
2. Legal rules are not static, but dynamic. As they are applied to new situations, they are constantly modified to "fit" the new "facts." Thus the important process in legal reasoning is not theory application, but theory construction.
3. In this process of theory construction, there is no single "right answer." However, there are plausible arguments, of varying degrees of persuasiveness, for each alternative version of the rule in each new factual situation.

Berman and Hafner (1993) studied the 1805 Pierson v. Post case concerning the ownership of a dead fox chased by Post, but killed and taken by Pierson. They emphasize the teleological aspects of legal argumentation, in which the goals of legal rules and decisions are taken into account. Bex (2011) used the Anjum case, a Dutch high-media-profile murder case, to test his proposal for a hybrid argumentative-narrative model of reasoning with evidence. Atkinson (2012) edited an issue of the journal *Artificial Intelligence and Law* on the modelling of a 2002 case about the ownership of a baseball, representing a value in the order of a million dollars, being the one that Barry Bonds hit when he broke the record of homeruns in one season (Popov v Hayashi).

11.14 The Need for Continued Collaboration

It has become clear that there are a great many issues that can be fruitfully researched if argumentation and artificial intelligence scholars cooperate (Reed & Norman, Eds., 2004b). One could think that arguments between humans have to be the area of argumentation theory, and arguments between machines (programs) have to be the area of artificial intelligence theory, and that never the twain need to meet, but one has only to think of discourse between humans and machines to see the inanity of such a conception. To achieve such a thing as an argumentation machine, both disciplines need to work side by side. And it is not just the traditionally logic-related or formalized part of argumentation that is involved: one also has to take in Toulmin's model and other theories of argumentation structures and argument schemes, the role of emotion in argument, and the rhetorical dimension of argumentation. All these issues are being studied in the interdisciplinary field of argumentation and computation.

References

Aleven, V. (1997). *Teaching case-based reasoning through a model and examples*. Doctoral dissertation, University of Pittsburgh.

Aleven, V., & Ashley, K. D. (1997a). Evaluating a learning environment for case-based argumentation skills. In *Proceedings of the sixth international conference on artificial intelligence and law* (pp. 170–179). New York: ACM Press.

Aleven, V., & Ashley, K. D. (1997b). Teaching case-based argumentation through a model and examples. Empirical evaluation of an intelligent learning environment. In B. du Boulay & R. Mizoguchi (Eds.), *Artificial intelligence in education. Proceedings of AI-ED 97 world conference* (pp. 87–94). Amsterdam: IOS Press.

Alexy, R. (1978). *Theorie der juristischen Argumentation* [Theory of legal argumentation]. Frankfurt am Main: Suhrkamp Verlag.

Amgoud, L. (2009). Argumentation for decision making. In I. Rahwan & G. R. Simari (Eds.), *Argumentation in artificial intelligence* (pp. 301–320). Dordrecht: Springer.

Amgoud, L., Cayrol, C., Lagasquie-Schiex, M. C., & Livet, P. (2008). On bipolarity in argumentation frameworks. *International Journal of Intelligent Systems, 23*(10), 1062–1093.

Antonelli, G. A. (2010). Non-monotonic logic. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Summer 2010 ed. http://plato.stanford.edu/archives/sum2010/entries/logic-non-monotonic/

Antoniou, G., Billington, D., Governatori, G., & Maher, M. (2001). Representation results for defeasible logic. *ACM Transactions on Computational Logic, 2*(2), 255–287.

Ashley, K. D. (1989). Toward a computational theory of arguing with precedents. Accommodating multiple interpretations of cases. In *Proceedings of the second international conference on artificial intelligence and law* (pp. 93–102). New York: ACM Press.

Ashley, K. D. (1990). *Modeling legal argument. Reasoning with cases and hypotheticals*. Cambridge, MA: The MIT Press.

Atkinson, K. (2012). Introduction to special issue on modelling Popov v. Hayashi. *Artificial Intelligence and Law, 20*, 1–14.

Atkinson, K., & Bench-Capon, T. J. M. (2007). Practical reasoning as presumptive argumentation using action based alternating transition systems. *Artificial Intelligence, 171*, 855–874.

Atkinson, K., Bench-Capon, T. J. M., & McBurney, P. (2005). A dialogue game protocol for multi-agent argument over proposals for action. *Autonomous Agents and Multi-Agent Systems, 11*, 153–171.

Atkinson, K., Bench-Capon, T. J. M., & McBurney, P. (2006). Computational representation of practical argument. *Synthese, 152*, 157–206.

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

Barth, E. M., & Krabbe, E. C. W. (1982). *From axiom to dialogue. A philosophical study of logics and argumentation*. Berlin: de Gruyter.

Bench-Capon, T. J. M. (2003). Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation, 13*(3), 429–448.

Bench-Capon, T. J. M., Araszkiewicz, M., Ashley, K., Atkinson, K., Bex, F., Borges, F., Bourcier, D., Bourgine, D., Conrad, J. G., Francesconi, E., Gordon, T. F., Governatori, G., Leidner, J. L., Lewis, D. D., Loui, R. P., McCarty, L. T., Prakken, H., Schilder, F., Schweighofer, E., Thompson, P., Tyrrell, A., Verheij, B., Walton, D. N., & Wyner, A. Z. (2012). A history of AI and Law in 50 papers: 25 years of the international conference on AI and Law. *Artificial Intelligence and Law, I, 20*(3), 215–319.

Bench-Capon, T. J. M., & Dunne, P. E. (2007). Argumentation in artificial intelligence. *Artificial Intelligence, 171*, 619–641.

Bench-Capon, T. J. M., Freeman, J. B., Hohmann, H., & Prakken, H. (2004). Computational models, argumentation theories and legal practice. In C. A. Reed & T. J. Norman (Eds.), *Argumentation machines. New frontiers in argument and computation* (pp. 85–120). Dordrecht: Kluwer.

Bench-Capon, T. J. M., Geldard, T., & Leng, P. H. (2000). A method for the computational modelling of dialectical argument with dialogue games. *Artificial Intelligence and Law, 8*, 233–254.

Bench-Capon, T. J. M., Prakken, H., & Sartor, G. (2009). Argumentation in legal reasoning. In I. Rahwan & G. R. Simari (Eds.), *Argumentation in artificial intelligence* (pp. 363–382). Dordrecht: Springer.

Bench-Capon, T. J. M., & Sartor, G. (2003). A model of legal reasoning with cases incorporating theories and values. *Artificial Intelligence, 150*, 97–143.

Berman, D., & Hafner, C. (1993). Representing teleological structure in case-based legal reasoning. The missing link. In *Proceedings of the fourth international conference on artificial intelligence and law* (pp. 50–59). New York: ACM Press.

Besnard, P., & Hunter, A. (2008). *Elements of argumentation*. Cambridge, MA: The MIT Press.

Bex, F. J. (2011). *Arguments, stories and criminal evidence. A formal hybrid theory*. Dordrecht: Springer.

Bex, F. J., van Koppen, P., Prakken, H., & Verheij, B. (2010). A hybrid formal theory of arguments, stories and criminal evidence. *Artificial Intelligence and Law, 18*(2), 123–152.

Bex, F. J., Prakken, H., Reed, C., & Walton, D. N. (2003). Towards a formal account of reasoning about evidence. Argumentation schemes and generalisations. *Artificial Intelligence and Law, 11*, 125–165.

Bex, F. J., & Verheij, B. (2012). Solving a murder case by asking critical questions. An approach to fact-finding in terms of argumentation and story schemes. *Argumentation, 26*(3), 325–353.

Bondarenko, A., Dung, P. M., Kowalski, R. A., & Toni, F. (1997). An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence, 93*, 63–101.

van den Braak, S. W., Vreeswijk, G., & Prakken, H. (2007). AVERs. An argument visualization tool for representing stories about evidence. In *Proceedings of the 11th international conference on artificial intelligence and law* (pp. 11–15). New York: ACM Press.

Branting, L. K. (1991). Building explanations from rules and structured cases. *International Journal of Man–Machine Studies, 34*, 797–837.

Branting, L. K. (2000). *Reasoning with rules and precedents. A computational model of legal analysis*. Dordrecht: Kluwer.

Bratko, I. (2001). *PROLOG. Programming for artificial intelligence* (3rd ed.). Harlow: Pearson (1st ed. 1986).

Brewka, G. (2001). Dynamic argument systems. A formal model of argumentation processes based on situation calculus. *Journal of Logic and Computation, 11*, 257–282.

Buckingham Shum, S., & Hammond, N. (1994). Argumentation-based design rationale. What use at what cost? *International Journal of Human-Computer Studies, 40*(4), 603–652.

Caminada, M. (2006). Semi-stable semantics. In P. E. Dunne & T. J. M. Bench-Capon (Eds.), *Computational models of argument. Proceedings of COMMA 2006, September 11–12, 2006, Liverpool, UK* (Frontiers in artificial intelligence and applications, Vol. 144). Amsterdam: IOS Press.

Cayrol, C., & Lagasquie-Schiex, M. C. (2005). On the acceptability of arguments in bipolar argumentation frameworks. In L. Godo (Ed.), *Symbolic and quantitative approaches to reasoning with uncertainty. 8th European conference, ECSQARU 2005* (pp. 378–389). Berlin: Springer.

Chesñevar, C., McGinnis, J., Modgil, S., Rahwan, I., Reed, C., Simari, G., South, M., Vreeswijk, G., & Willmott, S. (2006). Towards an argument interchange format. *Knowledge Engineering Review, 21*(4), 293–316.

Chesñevar, C. I., Simari, G. R., Alsinet, T., & Godo, L. (2004). A logic programming framework for possibilistic argumentation with vague knowledge. In *Proceedings of the 20th conference on uncertainty in artificial intelligence* (pp. 76–84). Arlington, VA: AUAI Press.

Crosswhite, J., Fox, J., Reed, C. A., Scaltsas, T., & Stumpf, S. (2004). Computational models of rhetorical argument. In C. A. Reed & T. J. Norman (Eds.), *Argumentation machines. New frontiers in argument and computation* (pp. 175–209). Dordrecht: Kluwer.

d'Avila Garcez, A. S., Lamb, L. C., & Gabbay, D. M. (2009). *Neural-symbolic cognitive reasoning*. Berlin: Springer.

Dignum, F., Dunin-Kęplicz, B., & Verbrugge, R. (2001). Creating collective intention through dialogue. *Logic Journal of the IGPL, 9*(2), 305–319.

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

Dung, P. M., & Thang, P. M. (2010). Towards (probabilistic) argumentation for jury-based dispute resolution. In P. Baroni, F. Cerutti, M. Giacomin, & G. R. Simari (Eds.), *Computational models of argument – Proceedings of COMMA 2010* (pp. 171–182). Amsterdam: Ios Press.

Dunne, P. E. (2007). Computational properties of argument systems satisfying graph-theoretic constraints. *Artificial Intelligence, 171*(10), 701–729.

Dunne, P. E., & Bench-Capon, T. J. M. (2003). Two party immediate response disputes. Properties and efficiency. *Artificial Intelligence, 149*(2), 221–250.

Dworkin, R. (1978). *Taking rights seriously. New impression with a reply to critics*. London: Duckworth.

van Eemeren, F. H., & Grootendorst, R. (1992a). *Argumentation, communication, and fallacies. A pragma-dialectical perspective*. Hillsdale: Lawrence Erlbaum (transl. into Bulgarian (2009), Chinese (1991b), French (1996), Romanian (2010), Russian (1992b), Spanish (2007)).

van Eemeren, F. H., Grootendorst, R., & Kruiger, T. (1978). *Argumentatietheorie* [Argumentation theory]. Utrecht: Het Spectrum. (2nd extended ed. 1981; 3rd ed. 1986; English trans. 1984, 1987).

van Eemeren, F. H., Grootendorst, R., & Kruiger, T. (1984). *The study of argumentation*. New York: Irvington. Engl. transl. by H. Lake of F. H. van Eemeren, R. Grootendorst & T. Kruiger (1981). *Argumentatietheorie*. 2nd ed. Utrecht: Het Spectrum. (1 st ed. 1978). (Reprinted as Eemeren, F. H. van, Grootendorst, R., & Kruiger, T. (1987). *Handbook of argumentation theory. A critical survey of classical backgrounds and modern studies*. Dordrecht/Providence: Foris).

van Eemeren, F. H., & Kruiger, T. (1987). Identifying argumentation schemes. In F. H. van Eemeren, R. Grootendorst, J. A. Blair, & C. Willard (Eds.), *Argumentation. Perspectives and approaches* (pp. 70–81). Dordrecht: Foris.

Egly, U., Gaggl, S. A., & Woltran, S. (2010). Answer-set programming encodings for argumentation frameworks. *Argument and Computation, 1*(2), 147–177.

Elhadad, M. (1995). Using argumentation in text generation. *Journal of Pragmatics, 24*, 189–220.

Falappa, M. A., Kern-Isberner, G., & Simari, G. R. (2002). Explanations, belief revision and defeasible reasoning. *Artificial Intelligence, 141*(1–2), 1–28.

Fenton, N. E., Neil, M., & Lagnado, D. A. (2012). A general structure for legal arguments using Bayesian networks. *Cognitive Science*, advance access. http://dx.doi.org/10.1111/cogs.12004

Fitelson, B. (2010). Pollock on probability in epistemology. *Philosophical Studies, 148*, 455–465.

Fox, J., & Das, S. (2000). *Safe and sound. Artificial intelligence in hazardous applications*. Cambridge, MA: The MIT Press.

Fox, J., & Modgil, S. (2006). From arguments to decisions. Extending the Toulmin view. In D. Hitchcock & B. Verheij (Eds.), *Arguing on the Toulmin model. New essays in argument analysis and evaluation* (pp. 273–287). Dordrecht: Springer.

Gabbay, D. M., Hogger, C. J., & Robinson, J. A. (Eds.). (1994). *Handbook of logic in artificial intelligence and logic programming, 3. Non-monotonic reasoning and uncertain reasoning*. Oxford: Clarendon.

García, A. J., & Simari, G. R. (2004). Defeasible logic programming. An argumentative approach. *Theory and Practice of Logic Programming, 4*(2), 95–138.

Gardner, A. (1987). *An artificial intelligence approach to legal reasoning*. Cambridge, MA: The MIT Press.

Garssen, B. (2001). Argument schemes. In F. H. van Eemeren (Ed.), *Crucial concepts in argumentation theory* (pp. 81–99). Amsterdam: Amsterdam University Press.

van Gelder, T. (2007). The rationale for rationale. *Law, Probability and Risk, 6*, 23–42.

Gelfond, M., & Lifschitz, V. (1988). The stable model semantics for logic programming. In R. A. Kowalski & K. A. Bowen (Eds.), *Logic programming. Proceedings of the fifth international conference and symposium* (pp. 1070–1080). Cambridge, MA: The MIT Press.

Ginsberg, M. L. (1994). AI and non-monotonic reasoning. In D. M. Gabbay, C. J. Hogger, & J. A. Robinson (Eds.), *Handbook of logic in artificial intelligence and logic programming* (Non-monotonic reasoning and uncertain reasoning, Vol. 3, pp. 1–33). Oxford: Clarendon.

Girle, R., Hitchcock, D., McBurney, P., & Verheij, B. (2004). Decision support for practical reasoning. A theoretical and computational perspective. In C. A. Reed & T. J. Norman (Eds.), *Argumentation machines. New frontiers in argument and computation* (pp. 55–83). Dordrecht: Kluwer.

Gómez Lucero, M., Chesñevar, C., & Simari, G. (2009). Modelling argument accrual in possibilistic defeasible logic programming. In *ECSQARU '09 proceedings of the 10th European conference on symbolic and quantitative approaches to reasoning with uncertainty* (pp. 131–143). Berlin: Springer.

Gómez Lucero, M., Chesñevar, C., & Simari, G. (2013). Modelling argument accrual with possibilistic uncertainty in a logic programming setting. *Information Sciences, 228*, 1–25.

Gordon, T. F. (1993). The pleadings game. *Artificial Intelligence and Law, 2*(4), 239–292.

Gordon, T. F. (1995). *The pleadings game. An artificial intelligence model of procedural justice*. Dordrecht: Kluwer.

Gordon, T. F., & Karacapilidis, N. (1997). The Zeno argumentation framework. In *Proceedings of the ICAIL 1997 conference* (pp. 10–18). New York: ACM Press.

Gordon, T. F., Prakken, H., & Walton, D. (2007). The Carneades model of argument and burden of proof. *Artificial Intelligence, 171*, 875–896.

Grasso, F. (2002). Towards computational rhetoric. *Informal Logic, 22*, 195–229.

Grasso, F., Cawsey, A., & Jones, R. (2000). Dialectical argumentation to solve conflicts in advice giving. A case study in the promotion of healthy nutrition. *International Journal of Human-Computer Studies, 53*(6), 1077–1115.

Green, N. (2007). A study of argumentation in a causal probabilistic humanistic domain. Genetic counseling. *International Journal of Intelligent Systems, 22*, 71–93.

Habermas, J. (1973). Wahrheitstheorien [Theories of truth]. In H. Fahrenbach (Ed.), *Wirklichkeit und Reflexion. Festschrift, für W. Schulz* [Reality and reflection. Festschrift for W. Schulz] (pp. 211–265). Pfullingen: Neske.

Hage, J. C. (1997). *Reasoning with rules. An essay on legal reasoning and its underlying logic*. Dordrecht: Kluwer.

Hage, J. C. (2000). Dialectical models in artificial intelligence and law. *Artificial Intelligence and Law, 8*, 137–172.

Hage, J. C. (2005). *Studies in legal logic*. Berlin: Springer.

Hage, J. C., Leenes, R., & Lodder, A. R. (1993). Hard cases: A procedural approach. *Artificial Intelligence and Law, 2*(2), 113–167.

Hart, H. L. A. (1951). The ascription of responsibility and rights. In A. Flew (Ed.), *Logic and language*. Oxford: Blackwell. (Originally Proceedings of the Aristotelian Society, 1948–1949)

Hepler, A. B., Dawid, A. P., & Leucari, V. (2007). Object-oriented graphical representations of complex patterns of evidence. *Law, Probability & Risk, 6*, 275–293.

Hitchcock, D. L. (2001). John L. Pollock's theory of rationality. In C. W. Tindale, H. V. Hansen & E, Sveda (Eds.), *Argumentation at the Century's turn*. (Proceedings of the 3rd OSSA Conference, 1999). Windsor, ON: Ontario Society for the Study of Argumentation. CD rom.

Hitchcock, D. L. (2002). Pollock on practical reasoning. *Informal Logic, 22*, 247–256.

Hofstadter, D. (1996). *Metamagical themas. Questing for the essence of mind and pattern*. New York: Basic Books.

Hunter, A. (2013). A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning, 54*(1), 47–81.

Hunter, A., & Williams, M. (2010). Qualitative evidence aggregation using argumentation. In P. Baroni, F. Cerutti, M. Giacomin, & G. R. Simari (Eds.), *Computational models of argument – Proceedings of COMMA 2010* (pp. 287–298). Amsterdam: Ios Press.

Jakobovits, H., & Vermeir, D. (1999). Robust semantics for argumentation frameworks. *Journal of Logic and Computation, 9*(2), 215–261.

Jensen, F. V., & Nielsen, T. D. (2007). *Bayesian networks and decision graphs*. New York: Springer.

Josephson, J. R., & Josephson, S. G. (Eds.). (1996). *Abductive inference. Computation, philosophy, technology*. Cambridge: Cambridge University Press.

Karacapilidis, N., & Papadias, D. (2001). Computer supported argumentation and collaborative decision making. The HERMES system. *Information Systems, 26*, 259–277.

Kienpointner, M. (1992). *Alltagslogik. Struktur and Funktion von Argumentationsmustern* [Everyday logic. Structure and functions of specimens of argumentation]. Stuttgart: Fromman-Holzboog.

Kirschner, P. A., Buckingham Shum, S. J., & Carr, C. S. (Eds.). (2003). *Visualizing argumentation. Software tools for collaborative and educational sense-making*. London: Springer.

Kjaerulff, U. B., & Madsen, A. L. (2008). *Bayesian networks and influence diagrams*. New York: Springer.

Kowalski, R. A. (2011). *Computational logic and human thinking. How to be artificially intelligent*. Cambridge: Cambridge University Press.

Kunz, W., & Rittel, H. (1970). *Issues as elements of information systems* (Technical Report 0131). Universität Stuttgart, Institut für Grundlagen der Planung.

Kyburg, H. E. (1994). Uncertainty logics. In D. M. Gabbay, C. J. Hogger, & J. A. Robinson (Eds.), *Handbook of logic in artificial intelligence and logic programming, 3. Non-monotonic reasoning and uncertain reasoning* (pp. 397–438). Oxford: Clarendon.

Lodder, A. R. (1999). *DiaLaw. On legal justification and dialogical models of argumentation*. Dordrecht: Kluwer.

Lorenzen, P., & Lorenz, K. (1978). *Dialogische Logik* [Dialogical logic]. Darmstadt: Wissenschaftliche Buchgesellschaft.

Loui, R. P. (1987). Defeat among arguments. A system of defeasible inference. *Computational Intelligence, 2*, 100–106.

Loui, R. P. (1995). Hart's critics on defeasible concepts and ascriptivism. In *The fifth international conference on artificial intelligence and law. Proceedings of the conference* (pp. 21–30). New York: ACM. Extended report available at http://www1.cse.wustl.edu/~loui/ail2.pdf. Accessed 10 July 2012.

Loui, R. P. (1998). Process and policy. Resource-bounded nondemonstrative reasoning. *Computational Intelligence, 14*, 1–38.

Loui, R., & Norman, J. (1995). Rationales and argument moves. *Artificial Intelligence and Law, 3*, 159–189.

Loui, R., Norman, J., Altepeter, J., Pinkard, D., Craven, D., Linsday, J., & Foltz, M. A. (1997). Progress on room 5. A testbed for public interactive semi-formal legal argumentation. In *Proceedings of the sixth international conference on artificial intelligence and law* (pp. 207–214). New York: ACM Press.

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

Makinson, D. (1994). General patterns in non-monotonic reasoning. In D. M. Gabbay, C. J. Hogger, & J. A. Robinson (Eds.), *Handbook of logic in artificial intelligence and logic programming* (Non-monotonic reasoning and uncertain reasoning, Vol. 3, pp. 35–110). Oxford: Clarendon.

McBurney, P., Hitchcock, D., & Parsons, S. (2007). The eightfold way of deliberation dialogue. *International Journal of Intelligent Systems, 22*, 95–132.

McBurney, P., & Parsons, S. (2002a). Games that agents play. A formal framework for dialogues between autonomous agents. *Journal for Logic, Language and Information, 11*, 315–334.

McBurney, P., & Parsons, S. (2002b). Dialogue games in multi-agent systems. *Informal Logic, 22*, 257–274.

McBurney, P., & Parsons, S. (2009). Dialogue games for agent argumentation. In I. Rahwan & G. R. Simari (Eds.), *Argumentation in artificial intelligence* (pp. 261–280). Dordrecht: Springer.

McCarty, L. (1977). Reflections on TAXMAN. An experiment in artificial intelligence and legal reasoning. *Harvard Law Review, 90*, 89–116.

McCarty, L. (1995). An implementation of Eisner v. Macomber. In *Proceedings of the fifth international conference on artificial intelligence and law* (pp. 276–286). New York: ACM Press.

Mochales Palau, R., & Moens, S. (2009). Argumentation mining. The detection, classification and structure of arguments in text. In *Proceedings of the 12th international conference on artificial intelligence and law (ICAIL 2009)* (pp. 98–107). New York: ACM Press.

Modgil, S. (2005). Reasoning about preferences in argumentation frameworks. *Artificial Intelligence, 173*(9–10), 901–934.

Nute, D. (1994). Defeasible logic. In D. M. Gabbay, C. J. Hogger, & J. A. Robinson (Eds.), *Handbook of logic in artificial intelligence and logic programming* (Non-monotonic reasoning and uncertain reasoning, Vol. 3, pp. 353–395). Oxford: Clarendon.

Parsons, S., Sierra, C., & Jennings, N. R. (1998). Agents that reason and negotiate by arguing. *Journal of Logic and Computation, 8*, 261–292.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems. Networks of plausible inference*. San Francisco: Morgan Kaufmann Publishers.

Pearl, J. (2009). *Causality. Models, reasoning, and inference* (2nd ed.). Cambridge: Cambridge University Press (1st ed. 2000).

Perelman, C., & Olbrechts-Tyteca, L. (1969). *The new rhetoric. A treatise on argumentation*. Notre Dame: University of Notre Dame Press. [trans.: Wilkinson, J. & Weaver, P. of C. Perelman and L. Olbrechts-Tyteca (1958). *La nouvelle rhétorique. Traité de l'argumentation*. Paris: Presses Universitaires de France].

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

Pollock, J. L. (1989). *How to build a person. A prolegomenon*. Cambridge, MA: The MIT Press.

Pollock, J. L. (1994). Justification and defeat. *Artificial Intelligence, 67*, 377–407.

Pollock, J. L. (1995). *Cognitive carpentry. A blueprint for how to build a person*. Cambridge, MA: The MIT Press.

Pollock, J. L. (2006). *Thinking about acting. Logical foundations for rational decision making*. New York: Oxford University Press.

Pollock, J. L. (2010). Defeasible reasoning and degrees of justification. *Argument & Computation, 1*(1), 7–22.

Poole, D. L. (1985). On the comparison of theories. Preferring the most specific explanation. In *Proceedings of the ninth international joint conference on artificial intelligence* (pp. 144–147). San Francisco: Morgan Kaufmann.

Prakken, H. (1993). *Logical tools for modelling legal argument*. Doctoral dissertation, Free University Amsterdam.

Prakken, H. (1997). *Logical tools for modelling legal argument. A study of defeasible reasoning in law*. Dordrecht: Kluwer.

Prakken, H. (2005a). A study of accrual of arguments, with applications to evidential reasoning. In *Proceedings of the tenth international conference on artificial intelligence and law* (pp. 85–94). New York: ACM Press.

Prakken, H. (2005b). Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation, 15*, 1009–1040.

Prakken, H. (2006). Formal systems for persuasion dialogue. *The Knowledge Engineering Review, 21*(2), 163–188.

Prakken, H. (2009). Models of persuasion dialogue. In I. Rahwan & G. R. Simari (Eds.), *Argumentation in artificial intelligence* (pp. 281–300). Dordrecht: Springer.

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

Prakken, H., & Sartor, G. (1996). A dialectical model of assessing conflicting arguments in legal reasoning. *Artificial Intelligence and Law, 4*, 331–368.

Prakken, H., & Sartor, G. (1998). Modelling reasoning with precedents in a formal dialogue game. *Artificial Intelligence and Law, 6*, 231–287.

Prakken, H., & Sartor, G. (2007). Formalising arguments about the burden of persuasion. In *Proceedings of the eleventh international conference on artificial intelligence and law* (pp. 97–106). New York: ACM Press.

Prakken, H., & Sartor, G. (2009). A logical analysis of burdens of proof. In H. Kaptein, H. Prakken, & B. Verheij (Eds.), *Legal evidence and proof. Statistics, stories, logic* (pp. 223–253). Farnham: Ashgate.

Prakken, H., & Vreeswijk, G. A. W. (2002). Logics for defeasible argumentation. In D. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic* (2nd ed., Vol. 4, pp. 219–318). Dordrecht: Kluwer.

Rahwan, I., & McBurney, P. (2007). Argumentation technology. Guest editors' introduction. *IEEE Intelligent Systems, 22*(6), 21–23.

Rahwan, I., Ramchurn, S. D., Jennings, N. R., McBurney, P., Parsons, S., & Sonenberg, E. (2003). Argumentation-based negotiation. *Knowledge Engineering Review, 18*(4), 343–375.

Rahwan, I., & Simari, G. R. (Eds.). (2009). *Argumentation in artificial intelligence*. Dordrecht: Springer.

Rahwan, I., Zablith, F., & Reed, C. (2007). Laying the foundations for a world wide argument web. *Artificial Intelligence, 171*(10–15), 897–921.

Rao, A., & Georgeff, M. (1995). BDI agents. From theory to practice. In *Proceedings of the 1st international conference on multi-agent systems* (pp. 312–319). Cambridge, MA: The MIT Press.

Reed, C. A. (1999). The role of saliency in generating natural language arguments. In *Proceedings of the 16th international joint conference on AI (IJCAI'99)* (pp. 876–881). San Francisco: Morgan Kaufmann.

Reed, C. A., & Grasso, F. (2007). Recent advances in computational models of natural argument. *International Journal of Intelligent Systems, 22*, 1–15.

Reed, C. A., & Norman, T. J. (2004a). A roadmap of research in argument and computation. In C. A. Reed & T. J. Norman (Eds.), *Argumentation machines. New frontiers in argument and computation* (pp. 1–13). Dordrecht: Kluwer.

Reed, C. A., & Norman, T. J. (Eds.). (2004b). *Argumentation machines. New frontiers in argument and computation*. Dordrecht: Kluwer.

Reed, C. A., & Rowe, G. W. A. (2004). Araucaria. Software for argument analysis, diagramming and representation. *International Journal on Artificial Intelligence Tools, 13*, 961–979.

Reed, C. A., & Tindale, C. W. (Eds.). (2010). *Dialectics, dialogue and argumentation. An examination of Douglas Walton's theories of reasoning*. London: College Publications.

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence, 13*, 81–132.

Rissland, E. L., & Ashley, K. D. (1987). A case-based system for trade secrets law. In *Proceedings of the first international conference on artificial intelligence and law* (pp. 60–66). New York: ACM Press.

Rissland, E. L., & Ashley, K. D. (2002). A note on dimensions and factors. *Artificial Intelligence and Law, 10*, 65–77.

Rittel, H., & Webber, M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*, 155–169.

Riveret, R., Rotolo, A., Sartor, G., Prakken, H., & Roth, B. (2007). Success chances in argument games. A probabilistic approach to legal disputes. In A. R. Lodder & L. Mommers (Eds.), *Legal knowledge and information systems (JURIX 2007)* (pp. 99–108). Amsterdam: Ios Press.

Roth, B. (2003). *Case-based reasoning in the law. A formal theory of reasoning by case comparison*. Doctoral dissertation, University of Maastricht.

Sartor, G. (2005). *Legal reasoning. A cognitive approach to the law* (Treatise on legal philosophy and general jurisprudence, Vol. 5). Berlin: Springer.

Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B. M. (2010). Computer-supported argumentation. A review of the state of the art. *Computer-Supported Collaborative Learning, 5*, 43–102.

Schwemmer, O., & Lorenzen, P. (1973). *Konstruktive Logik, Ethik und Wissenschaftstheorie [Constructive logic, ethics and theory of science]*. Mannheim: Bibliographisches Institut.

Simari, G. R., & Loui, R. P. (1992). A mathematical treatment of defeasible reasoning and its applications. *Artificial Intelligence, 53*, 125–157.

Suthers, D. (1999). Representational support for collaborative inquiry. In *Proceedings of the 32nd Hawaii international conference on the system sciences (HICSS-32)*. Los Alamitos, CA: Institute of Electrical and Electronics Engineers (IEEE).

Suthers, D., Weiner, A., Connelly, J., & Paolucci, M. (1995). Belvedere. Engaging students in critical discussion of science and public policy issues. In *Proceedings of the 7th world conference on artificial intelligence in education (AIED '95)*. Charlottesville, VA: Association for the Advancement of Computing in Education, pp. 266–273.

Sycara, K. (1989). Argumentation. Planning other agents' plans. In *Proceedings of the eleventh international joint conference on artificial intelligence* (pp. 517–523). San Francisco: Morgan Kaufmann Publishers.

Talbott, W. (2011). Bayesian epistemology. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2011 ed.). http://plato.stanford.edu/archives/sum2011/entries/epistemology-bayesian/

Taroni, F., Aitken, C., Garbolino, P., & Biedermann, A. (2006). *Bayesian networks and probabilistic inference in forensic science*. Chichester: Wiley.

Teufel, S. (1999). *Argumentative zoning. Information extraction from scientific articles*. Doctoral dissertation, University of Edinburgh.

Thagard, P. (1992). *Conceptual revolutions*. Princeton: Princeton University Press.

Toulmin, S. E. (2003). *The uses of argument*. Cambridge: Cambridge University Press (1st ed. 1958).

Verheij, B. (1996a). *Rules, reasons, arguments. Formal studies of argumentation and defeat*. Doctoral dissertation, University of Maastricht.

Verheij, B. (1996b). Two approaches to dialectical argumentation. Admissible sets and argumentation stages. In J.-J. C. Meyer & L. C. van der Gaag (Eds.), *NAIC'96. Proceedings of the eighth Dutch conference on artificial intelligence* (pp. 357–368). Utrecht: Utrecht University.

Verheij, B. (2003a). DefLog. On the logical interpretation of prima facie justified assumptions. *Journal of Logic and Computation, 13*(3), 319–346.

Verheij, B. (2003b). Dialectical argumentation with argumentation schemes. An approach to legal logic. *Artificial Intelligence and Law, 11*(1–2), 167–195.

Verheij, B. (2005a). Evaluating arguments based on Toulmin's scheme. *Argumentation, 19*, 347–371. [Reprinted in Hitchcock, D. L., & Verheij, B. (Eds.). (2006), *Arguing on the Toulmin model. New essays in argument analysis and evaluation* (pp. 181–202). Dordrecht: Springer].

Verheij, B. (2005b). *Virtual arguments. On the design of argument assistants for lawyers and other arguers*. The Hague: T. M. C. Asser Press.

Verheij, B. (2007). A labeling approach to the computation of credulous acceptance in argumentation. In M. M. Veloso (Ed.), *IJCAI 2007, Proceedings of the 20th international joint conference on artificial intelligence* (pp. 623–628). San Francisco: Morgan Kaufmann Publishers.

Verheij, B. (2012). Jumping to conclusions. A logico-probabilistic foundation for defeasible rule-based arguments. In L. Fariñas del Cerro, A. Herzig, & J. Mengin (Eds.), *Logics in artificial intelligence. 13th European conference, JELIA 2012. Toulouse, France, September 2012. Proceedings* (LNAI, Vol. 7519, pp. 411–423). Berlin: Springer.

Verheij, B., Hage, J. C., & van den Herik, H. J. (1998). An integrated view on rules and principles. *Artificial Intelligence and Law, 6*(1), 3–26.

Vreeswijk, G. A. W. (1993). *Studies in defeasible argumentation*. Doctoral dissertation, Free University, Amsterdam.

Vreeswijk, G. A. W. (1995a). *Formalizing nomic. Working on a theory of communication with modifiable rules of procedure* (Technical Report CS 95–02). Maastricht: Vakgroep Informatica (FdAW), Rijksuniversiteit Limburg. http://arno.unimaas.nl/show.cgi?fid=126

Vreeswijk, G. A. W. (1995b). The computational value of debate in defeasible reasoning. *Argumentation, 9*, 305–342.

Vreeswijk, G. (1997). Abstract argumentation systems. *Artificial Intelligence, 90*, 225–279.

Vreeswijk, G. A. W. (2000). Representation of formal dispute with a standing order. *Artificial Intelligence and Law, 8*, 205–231.

Walton, D. N., & Krabbe, E. C. W. (1995). *Commitment in dialogue. Basic concepts of interpersonal reasoning*. Albany: State University of New York Press.

Walton, D. N., Reed, C. A., & Macagno, F. (2008). *Argumentation schemes*. Cambridge: Cambridge University Press.

Wooldridge, M. (2009). *An introduction to multiagent systems*. Chichester: Wiley.

Zukerman, I., McConachy, R., & Korb, K. (1998). Bayesian reasoning in an abductive mechanism for argument generation and analysis. In *Proceedings of the fifteenth national conference on artificial intelligence (AAAI-98, Madison)* (pp. 833–838). Menlo Park: AAAI Press.