Historical overview of formal argumentation

Prakken, Henry

1
Historical Overview of Formal Argumentation

Henry Prakken

ABSTRACT. This chapter gives an overview of the history of formal argumentation in terms of a distinction between argumentation-based *inference* and argumentation-based *dialogue*. Systems for argumentation-based *inference* are about which conclusions can be drawn from a given body of possibly incomplete, inconsistent of uncertain information. They ultimately define a nonmonotonic notion of logical consequence, in terms of the intermediate notions of argument construction, argument attack and argument evaluation, where arguments are seen as constellations of premises, conclusions and inferences. Systems for argumentation-based *dialogue* model argumentation as a kind of verbal interaction aimed at resolving conflicts of opinion. They define argumentation protocols, that is, the rules of the argumentation game, and address matters of strategy, that is, how to play the game well. For both aspects of argumentation the main formal and computational models are reviewed and their main historical influences are sketched. Then some main applications areas are briefly discussed.

1 Introduction

This chapter gives an overview of the history of formal argumentation. There are two ways to write such an overview. One is to describe all significant research that has been done, while another is to give insight into the historical developments underlying the current state of the art. In this chapter I will do the latter. This will inevitably lead to a stronger focus on the early developments and a less detailed description of later research. Those who want more detail about the later research can consult the other chapters of this handbook.

The historical overview is given in terms of a distinction between argumentation-based *inference* and argumentation-based *dialogue*. Systems for argumentation-based *inference* are about which conclusions can be drawn from a given body of possibly incomplete, inconsistent of uncertain information. They ultimately define a nonmonotonic notion of logical consequence, in terms of the intermediate notions of argument construction, argument attack and argument evaluation, where arguments are seen as constellations of premises, conclusions and inferences. Systems for argumentation-based *dialogue* model argumentation as a kind of verbal interaction aimed at resolving conflicts of opinion. They define argumentation protocols (the rules of the argumentation game) and address matters of strategy (how to play the game well). While accounts of argumentation as inference assume a single static and global body of information from which the arguments and attacks are constructed, in studies of argumentation as dialogue this information is dynamic (it can change during a dialogue) and distributed over the dialogue's participants. Models of argumentation as inference can be embedded in models of argumentation as dialogue in two complementary ways: at each stage of a dialogue they can be 'globally' applied to the 'current' body of information; and within each dialogue participant they can be 'locally' applied as the participant's internal reasoning model.

Like all informal distinctions, the distinction between argumentation as inference and argumentation as dialogue breaks down at some point, and therefore I will also discuss work that cannot easily be classified as belonging to either inference or dialogue, especially work on argumentation dynamics that abstracts from agent-related and dialogical aspects. Another way in which a strict distinction between inference and dialogue causes problems for a historical overview is that some historical influences cannot clearly be described as influencing just models of inference or just models of dialogue. Some work has instead more generally promoted the idea of dialectics as constructing, criticising and comparing arguments, whether in an inferential or in a dialogical setting. One such historical influence was the development of dialogue logic [Lorenzen and Lorenz, 1978], which gives a game-theoretic formulation of the semantics of logical constants in terms of a dispute between a proponent and an opponent of a claim, plus a game-theoretic notion of logical consequence as the existence of a winning strategy for the proponent. This predates modern argument games for argumentation-based inference and also influenced the development of formal dialogue systems for argumentation. Having said so, in dialogue logic these ideas were only used to reformulate existing monotonic notions of logical consequence, so dialogue logic cannot be said to model genuine argumentation.

Another historical influence that is not confined to either inference or dialogue is early AI & Law work on the computational modelling of legal argument. Among the earliest work in AI and law on legal argument was the TAXMAN II project [McCarty, 1977; McCarty, 1995]). According to McCarty [1995], p. 285 "The task for a lawyer or a judge in a "hard case" is to construct a theory of the disputed rules that produces the desired legal result, and then to persuade the relevant audience that this theory is preferable to any theories offered by an opponent". Other influential early systems were the HYPO system [Rissland and Ashley, 1987; Ashley, 1990] and its successor the CATO system [Aleven and Ashley, 1991; Aleven, 2003]. These systems were meant to model how lawyers in common-law jurisdictions make use of past decisions when arguing a case. They did not compute an 'outcome' or 'winner' of a dispute; instead they were meant to generate debates as they could take place between 'good' common-law lawyers. Several researchers who later contributed to the general formal study of argumentation originate from AI & Law, such as Trevor Bench-Capon, Tom Gordon, Giovanni Sartor, Bart Verheij and myself.

The remainder of this chapter is divided into two main sections on, respectively, argumentation-based inference (Section 2) and dialogue (Section 3). Then some main applications areas are briefly discussed in Section 4 and some concluding remarks are made in Section 5.

2 Formal and computational models of argumentation-based inference

Nowadays, many systematic introductions to argumentation start with [Dung, 1995]'s theory of abstract argumentation frameworks, which takes the notions of argument and attack as primitive, i.e., nothing is assumed about about the structure of arguments or the nature of attack. Yet there had been quite some formal work on argumentation-based inference before Dung's landmark 1995 paper, and all this early work specified the structure of arguments and the nature of attack. The seminal paper in this respect was [Pollock, 1987]. Many ideas developed in this early body of work are still important today. The focus in this early work on structured argumentation agrees with the usual approaches in informal argumentation, which do not have arguments as the primitive notion but concepts like claims, reasons and grounds. For example, Walton [2006a], p. 285 defines the term 'argument' as "the giving of reasons to support or criticize a claim that is questionable, or open to doubt".

In this section first the three main historical sources of influence are sketched, namely, philosophy, nonmonotonic logic & logic programming, and informal logic & argumentation theory. Then the two seminal bodies of work are discussed in more more detail, John Pollock's argumentation-based system for defeasible reasoning and Phan Minh Dung's theory of abstract argumentation frameworks. Their works have inspired much research on, respectively, structured and abstract approaches to argumentation-based inference, which will subsequently be discussed.

2.1 Main historical influences

The formal and computational study of argumentation-based inference is generally regarded as a subfield of AI, originating from the study of nonmonotonic logic. However, there are two main other historical influences.

2.1.1 Philosophy

Arguably, the first mature formal system for argumentation-based inference was proposed by Pollock [1987][1]. John Pollock (1940-2009) was an influential American philosopher who made important contributions to various fields, including epistemology and cognitive science. In the last 25 years of his life he also contributed to artificial intelligence, starting with his classic 1987 paper on defeasible reasoning. Many important topics in the formal study of argumentation-based inference were first studied by Pollock, or first studied in detail, such as argument structure, the nature of defeasible reasons, the

[1] Several paragraphs in this subsection are, some with minor modifications, taken from Prakken and Horty [2012].

interplay between deductive and defeasible reasons, rebutting versus undercutting defeat, argument strength, argument labellings, self-defeat, and resource-bounded argumentation.

Pollock's work on formal argumentation was heavily influenced by the idea of defeasible reasons as developed in moral philosophy by Ross [1930] in his notion of *prima facie* moral rules, in epistemology by Chisholm [1957], Rescher [1977] and Pollock himself [1970, 1974], and as applied to practical reasoning by Raz [1975]. The term 'defeasibility' originates from legal philosophy, in particular from Hart [1949] (see the historical discussion in Loui [1995]). Hart observed that legal concepts are defeasible in that the conditions for when a fact situation classifies as an instance of a legal concept (such as 'contract'), are only ordinarily, or presumptively, sufficient. If a party in a law suit succeeds in proving these conditions, this does not have the effect that the case is settled; instead, legal procedure is such that the burden of proof shifts to the opponent, whose turn it then is to prove exceptional facts which, despite the facts proven by the proponent, nevertheless prevent the claim from being granted. For instance, insanity of one of the contracting parties is an exception to the legal rule that an offer and an acceptance constitute a binding contract. The notion of burden of proof was also studied by [Rescher, 1977], in the context of epistemology. Among other things, Rescher claimed that a dialectical model of scientific reasoning can explain the rational force of inductive arguments: they must be accepted if they cannot be successfully challenged in a properly conducted scientific dispute.

Pollock's work on formal argumentation originated as an attempt to make formal sense of the intuitive notion of defeasible reasoning that seemed to be at work in these papers and books. In fact, the task had been attempted before. There is an early paper by Chisholm [1974], a heroic effort whose failure is no surprise given the limited tools available at the time. Still, in spite of the blossoming of philosophical logic in the 1960's and 1970's, the logical study of defeasible reasoning had received almost no attention at all. It is fair to say that Pollock, working in isolation, was the first philosopher working in the field of philosophy, as opposed to computer science, to outline an adequate framework for defeasible reasoning.

2.1.2 Nonmonotonic logic and logic programming

The first AI systems for argumentation-based inference were not influenced by the above-discussed philosophical developments. Instead, they were presented as new ways to do nonmonotonic logic. Nonmonotonic logic had become fashionable around 1980 and a variety of approaches was being pursued. By the late 1980's, the field of nonmonotonic logic had been recognized as an important subfield of artificial intelligence. The field was motivated by the fact that commonsense reasoning often involves incomplete or inconsistent information, in which cases logical deduction is not a useful reasoning model. If information is incomplete, then nothing useful can be deductively derived, while if it is inconsistent, then anything is deductively implied. Nonmonotonic logics allow 'jumping to conclusions' in the absence of information to the contrary. The canonical example is 'birds typically fly, Tweety is bird, therefore (presumably) Tweety can fly'. This inference holds as long as no information is available that Tweety is not a typical bird with respect to flying, such as a penguin. Nonmonotonic logic can also model the derivation of useful conclusions from inconsistent information, namely, by focusing on consistent subsets of the inconsistent information. Several years after the first nonmonotonic logics were proposed in the now famous special issue on nonmonotonic logic of the *Artificial Intelligence* journal [Bobrow, 1980], the idea arose in this field that nonmonotonic inference can be modelled as the competition between arguments.

The earliest nonmonotonic reasoning systems with an argumentation flavour include the work of Touretzky [1984; 1986] on inheritance systems, later developed along with several collaborators [Horty *et al.*, 1990]. Inheritance systems model reasoning about how objects inherit properties from the classes to which they belong. They are nonmonotonic since the inheritance of properties of classes by subclasses can be blocked by exceptions. For example, penguins do not inherit from birds the property of being able to fly. Although the work on inheritance systems did not use argumentation terms, such systems still have all the characteristics of argumentation systems. To start with, inheritance paths effectively are arguments. For example, the conclusion that Tweety the penguin can fly can be drawn via the path 'Penguins are birds and birds can fly' while the conclusion that Tweety the Penguin cannot fly can be drawn via the inheritance path 'Penguins cannot fly'. Inheritance systems also have various notions of conflict between inheritance plus definitions of whether a path is 'permitted' given its conflict relations with other paths. While the technical solutions devised in this work are now somewhat outdated, the work on inheritance paths has clearly influenced the development of the first AI argumentation systems. Among other things, the publications in inheritance are great sources of relevant examples.

An influential figure in the early days was Ron Loui. His [1987] paper was, although technically still preliminary, influential in promoting the idea of formulating nonmonotonic logic as argumentation. With Guillermo Simari he developed a technically mature version of his ideas [Simari and Loui, 1992]. Several other of his papers more generally promoted the idea of computational dialectics and were thus also relevant for dialogue models of argumentation. The fullest exposé of these ideas is [Loui, 1998], which circulated among researchers for several years until it was finally published in 1998.

Other relevant early work was the work of Nute [1988], later developed into so-called Defeasible Logic [Nute, 1994]. This approach is in spirit very close to argumentation but while in argumentation approaches conflict and defeat happen between arguments, in Defeasible Logic they happen between rules. For this reason the work on Defeasible Logic has diverged somewhat from the field of computational argument, although some work on the former has studied the formal relation with argumentation approaches. In particular, [Governatori *et al.*, 2004] studied to which extent defeasible logics can be reformulated in terms of Dung's theory of abstract argumentation frameworks.

Finally, the field of logic programming was influential since the idea arose to give semantics to negation as failure in argumentation-theoretic terms. If *not* P is assumed to hold because of the failure to derive P, then a derivation of P can be regarded as an attack on any derivation using *not* P. In other words, a logic-programming derivation can be regarded as a competition between arguments and counterarguments. Work on this idea of e.g. Geffner [1991] and Kakas *et al.* [1992] was a main source of inspiration of Dung's landmark [1995] paper on abstract argumentation frameworks.

2.1.3 Informal logic and informal argumentation theory

One would expect that the fields of informal logic and argumentation theory (which are often regarded as a single field) were also important historical influences on argumentation-based models of inference. However, in fact their influence has been relatively modest. In particular, the work of Toulmin [1958] and the resulting work on argumentation schemes was until around 2000 hardly linked to computational argument. An important event here was the 2000 Bonskeid Symposium on Argument and Computation in the Scottish mountains, organised by Tim Norman and Chris Reed, at which researchers from various formal and informal fields met in an informal setting. Various interdisciplinary collaborations resulted from this event, partly reported in [Reed and Norman, 2003].

Yet these fields originated from similar concerns about deductive logic as those that gave rise to the field of nonmonotonic logic in AI, namely, the inadequacy of deductive logic as a model of 'ordinary' reasoning. Stephen Toulmin, whose 1958 book *The Uses of Argument* is generally regarded as the origin of informal logic and argumentation theory, criticised the logicians of his days for neglecting many features of ordinary reasoning. In his well-known pictorial scheme for arguments (see Figure 1) he left room for "rebuttals" of an argument on the basis of exceptions to the "warrant" connecting the arguments "data" to its "claim". The idea of rebuttals is clearly related to Hart's [1949] ideas on exceptional circumstances that can defeat the application of a legal concept.

Toulmin's notion of a warrant was in informal logic and argumentation theory generalised into rich classifications of argument schemes for presumptive forms of reasoning, while his notion of a rebuttal was generalised into lists of critical questions attached to argument schemes [Walton, 1996]. The idea of argumentation schemes with critical questions has since the above-mentioned Bonskeid 2000 event often been used in formal and computational models of argumentation-based inference and dialogue.

Toulmin also argued that outside mathematics the validity of an argument does not depend on its syntactic form but on whether it can be defended in a properly conducted dispute, and that the task of logicians is to study the criteria for properly conducted disputes. This became an important and very

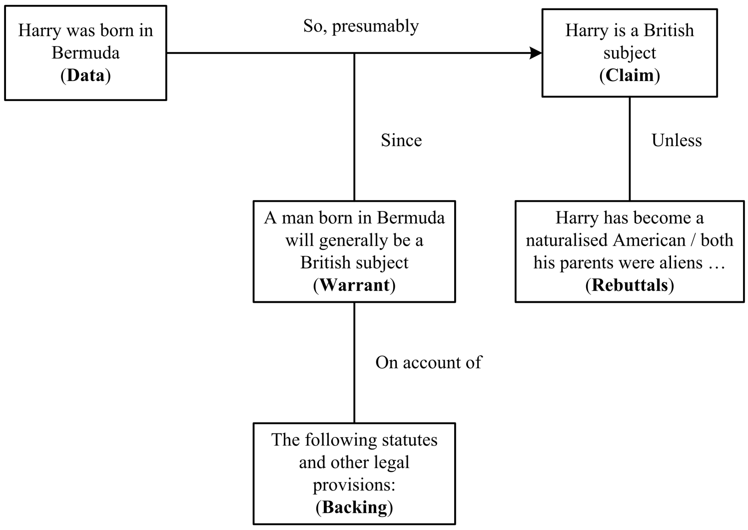

Figure 1. Toulmin argument scheme and an instance

influential idea, as further discussed below in Section 3 on argumentation-based dialogue. However, it also had an unfortunate effect. For decades, informal logic and argumentation theory rejected any use of formal methods in the study of ordinary reasoning, based on a mistaken equation of formal methods with deductive logic. As we now know after more than 35 years of research on nonmonotonic logic, belief revision and computational argument, many features of non-mathematical reasoning that Toulmin and his successors analysed can be formalised. For example, the AI work on argumentation schemes since 2000 has shown that reasoning with such schemes can to a large extent be formalised in modern argumentation logics.

2.2 Seminal work

I now discuss the two seminal contributions in the field, the ones of Pollock [1987] and Dung [1995]. These two papers successively introduced the two key ideas of the formal study of argumentation-based inference. Pollock introduced the notion of a defeasible reason, while Dung showed that argument evaluation can be formalised by assuming just two primitive notions of argument and attack. Neither of these ideas on their own define the field; it is their combination that makes the argumentation way of doing nonmonotonic logic so powerful.

2.2.1 Pollock's work

As said above, arguably, the first mature formal system for argumentation-based inference was proposed by Pollock [1987][2]. In fact, this work became

[2]Several parts of this subsection are reused or adapted from Prakken and Horty [2012].

close to being one of the first nonmonotonic logics at all. Concerning his 1987 paper, Pollock later wrote that he first developed the idea in 1979, but that he did not initially publish it because, as he says, “being ignorant of AI, I did not think anyone would be interested.” [Pollock, 2007b, p. 469]. If Pollock had published this idea when it first occurred to him, the result would have been not only the first argument-based theory of defeasible reasoning, but one of the first systems of any kind for nonmonotonic reasoning.

I now discuss Pollock’s system in some more detail, to illustrate that it introduced several fundamental ideas into our field. As usual in logic, arguments in Pollock’s approach are inference graphs, in which a final conclusion is inferred from the premises via intermediate conclusions. Note that when an argument uses no premise more than once, the graph is a tree. What is unusual is Pollock’s ideas on how conclusions can be supported by premises. The ‘classic’ logicians’ view attacked by Toulmin [1958] had been that all arguments should be deductively valid, that is, the truth of their premises should guarantee the truth of their conclusion, and that the only source of fallibility of good arguments is their premises. Influenced by Toulmin, the fields of informal logic and argumentation theory had already questioned this view and argued that arguments that fail to meet this standard of inferential perfection can still be good, as long as they withstand critical scrutiny. Pollock [1987] gave us the tools to formalise this new account, with his notion of a defeasible reason.

In Pollock’s approach, the inference rules (in his terminology “reasons”) used to construct arguments come in two kinds: *deductive* and *defeasible* reasons (in his early work called “conclusive’ and “prima facie” reasons). An argument can be defeated on its applications of defeasible reasons, which can happen in two ways. *Rebutting* defeaters attack the conclusion of a defeasible inference by supporting a conflicting conclusion. For example, ‘Tweety can fly since it is a bird and birds typically fly’ can be attacked by ‘Tweety cannot fly since Tweety is a penguin and penguins cannot fly’. *Undercutting* defeaters instead attack the defeasible inference itself, without supporting a conflicting conclusion. For example: if the object looks red, this is a reason for concluding, defeasibly, that the object is red; but the presence of red illumination interrupts the reason relation without suggesting any conflicting conclusion. Pollock formalized several defeasible reasons that he found important in human cognition, such as reasons for perception, memory, induction, the statistical syllogism and temporal persistence, as well as undercutting defeaters for these reasons.

Pollock’s notion of a defeasible reason is clearly related to argumentation theory’s notion of an argumentation scheme: such schemes are defeasible reasons while many of their critical questions can be regarded as pointers to undercutting defeaters and other questions as pointers to rebutting defeaters or premise attacks.

Consider by way of example of Pollock’s notions of reason, argument and conflict the following version of the Tweety example. Figure 2 contains two rebutting arguments for the conclusions that Tweety flies, respectively, does not fly, and an undercutting argument defeating the argument that Tweety flies. In this figure, deductive, respectively defeasible inferences are visualized with,

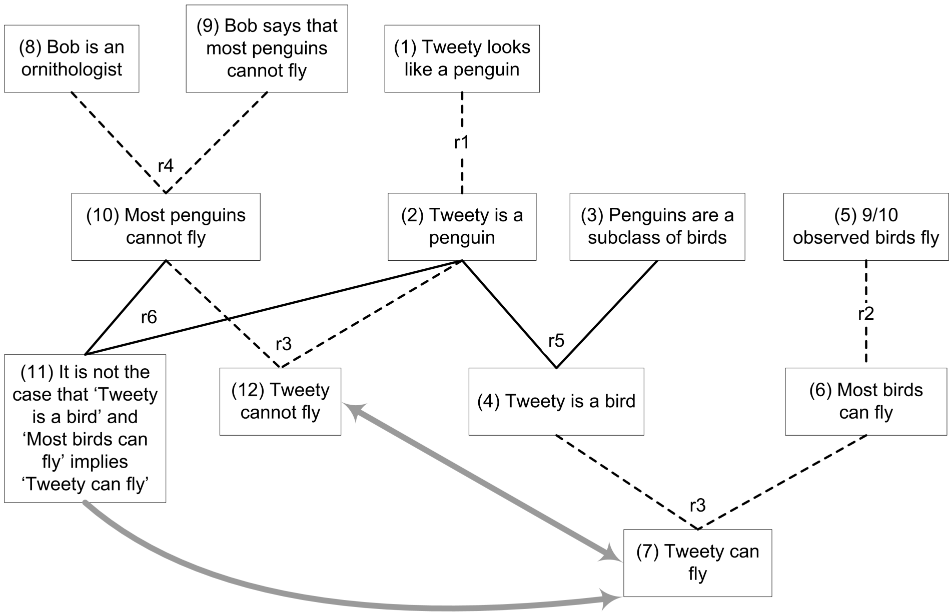

Figure 2. An example

respectively, solid and dotted lines without arrow heads, while defeat relations are displayed with arrows. The figure assumes four defeasible inference rules, informally paraphrased as follows:

- r_1: That an object looks like having property P is a defeasible reason for believing that the object has property P
- r_2: That n/m observed P's are Q's (where $n/m > 0,5$) is a defeasible reason for believing that most P's are Q's
- r_3: That most P's are Q's and x is a P is a defeasible reason for believing that x is a Q
- r_4: That an ornithologist says φ about birds is a defeasible reason for believing φ

Rule r_1 expresses that perceptions yield a defeasible reason for believing that what is perceived to be the case is indeed the case, rule r_2 captures enumerative induction, while r_3 expresses the statistical syllogism. Rule r_4 can be seen as a special case of the argumentation scheme from expert testimony; cf. [Walton, 1996].

Moreover, the figure assumes an obvious strict inference rule plus an undercutting defeater for r_3:

r_5: That P's are a subclass of Q's and a is a P is a deductive reason for believing that a is a Q

r_6: That x is an R, most R's are not Q's and R's are a subclass of P's is a deductive reason for believing $\neg r_3$

Rule r_6 is a special case of Pollock's "subproperty defeater" of the statistical syllogism, which says that conflicting statistical information about a subclass undercuts the statistical syllogism for the superclass.

Defeasible reasons should not be confused with nonmonotonic consequence notions. It is possible to design argumentation logics with nonmonotonic consequence notions in which nevertheless all arguments have to be deductively valid. For example, in classical argumentation arguments are classical implication relations from consistent subsets of a possibly inconsistent body of information and the only source of fallibility of arguments is their premises. Recent portrayals of Pollock's approach as 'deductive' [Hunter and Woltran, 2013] do no justice to his approach, given that Pollock strongly emphasised that "It is logically impossible to reason successfully about the world around us using only deductive reasoning. All interesting reasoning outside mathematics involves defeasible steps." [Pollock, 1995, p.41]. Pollock thus clearly rejected the conventional view that all arguments have to be deductively valid.

Defeasible reasons should also not be confused with deductive inference rules with assumption-type premises. Thinking otherwise would have the odd consequence that even the classically valid rules of inference become defeasible when applied to assumptions.

Once arguments can employ defeasible reasons, the support relation between their premises and conclusion can have varying strength. Pollock's 1987 system did not yet include a notion of strength but Pollock later took the notion of strength of arguments very seriously. Since his systems were meant for epistemic reasoning, he always formulated strength of reasons in terms of numerical degrees of belief. In his 1994 system, rebutting and undercutting arguments only succeed in defeating their target if the degree of belief of their conclusions is not lower than that of the attacked argument.

Finally, Pollock was well aware that just defining notions of argument and defeat are not enough and he spent much effort in designing well-behaved notions of argument acceptability. His two earliest definitions predate much current work on argumentation-based semantics. His 1987 proposal was by Dung [1995] proven to be an instance of Dung's grounded semantics, while his 1994 labelling definition predates the currently popular labeling approach to abstract argumentation and was by Jakobovits [2000] proven to be an instance of Dung's preferred semantics.

2.2.2 Dung's abstract argumentation frameworks

Dung's landmark 1995 paper is the origin of the second main idea of our field, namely, that argument evaluation can be formalised by assuming just two primitive notions of argument and attack. With just these two notions, Dung was able to develop an extremely rich and elegant abstract theory of argument evaluation. As apparent from this historic overview, Dung was not the first to study argument evaluation nor the first to provide well-behaved definitions. His great contribution was twofold: he showed that particular definitions of argument evaluation conformed to simple abstract patterns, and he showed that the same patterns are also implicit in other nonmonotonic logics, in logic programming and even in cooperative game theory. Exaggerating a little, one could say that while Pollock arguably was the father of argumentation in AI, Dung was the midwife, who smoothened its delivery into mainstream AI. His 1995 AI Journal paper was not the first work on argumentation-based inference, but its influence has been enormous, now being the de facto standard in the field. It is fair to say that Dung [1995] has made argumentation respectable in mainstream AI.

Nevertheless, the historic roots of Dung's 1995 paper should not be forgotten. As mentioned in the introduction to Section 2, all early work on argumentation-based inference specified the structure of arguments and the nature of attack (often called 'defeat'). Even Dung in his landmark 1995 paper stood in this tradition. Dung did two things: he developed the new idea of abstract argumentation frameworks, and he used this idea to reconstruct and compare a number of then mainstream nonmonotonic logics and logic-programming formalisms, namely, default logic [Reiter, 1980], Pollock's [1987] argumentation system and several logic-programming semantics. However, these days the second part of his paper, and also the third part on relations with cooperative game theory, is largely forgotten and his paper is almost exclusively cited for its general theory of abstract argumentation frameworks.

A historic overview of work on argumentation-based inference would not be complete without listing Dung's simple and elegant basic notions. An *abstract argumentation framework* (AF) is a pair $\langle AR, \mathit{attacks} \rangle$, where AR is a set arguments and $\mathit{attacks} \subseteq AR \times AR$ is a binary relation. The theory of AFs then addresses how sets of arguments (called *extensions*) can be identified which are internally coherent and defend themselves against attack. A key notion here is that of an argument being *acceptable with respect to* a set of arguments: $A \in AR$ is acceptable with respect to $S \subseteq AR$ if for all $A \in S$: if $B \in AR$ attacks A, then some $C \in S$ attacks B (nowadays it is more usual to say that $A \in AR$ is defended by $S \subseteq AR$). Then relative to a given AF various types of extensions can be defined as follows (here E is *conflict-free* if no argument in E attacks an argument in E):

- E is *admissible* if E is conflict-free and each argument in E is acceptable with respect to E;

- E is a *complete extension* if E is admissible and each argument that is acceptable with respect to E belongs to E;
- E is a *preferred extension* if E is a maximal (with respect to set inclusion) admissible set;
- E is a *stable extension* if E is conflict-free and attacks all arguments outside it;
- E is a *grounded extension* if E is the least fixpoint of operator F, where $F(S)$ returns all arguments acceptable to S.

Dung showed that the grounded extension is always unique but that there can be multiple extensions of the other types. Dung also showed that every stable extension is preferred but not vice versa, that the grounded extension is contained in every other extension, and that all extensions of any type are complete.

To illustrate how abstract argumentation frameworks can be instantiated, consider again Figure 2. There are three arguments. In fact, there are more arguments, since each of the three arguments we consider has several subarguments. However, none of these is attacked, so they can be ignored for simplicity. The two rebutting arguments for the conclusions that Tweety can fly, respectively, cannot fly attack each other, while the undercutting argument attacks the argument that Tweety flies. The resulting argumentation framework is shown in Figure 3. In this case the four semantics coincide: the set with the undercutting argument and the argument that Tweety cannot fly is the grounded extension, while it is also the unique complete, stable and preferred extension (the grey colourings indicate extension membership). To see why it is preferred, observe that the undercutting argument defends the argument that Tweety cannot fly against its rebutting attacker that Tweety can fly.

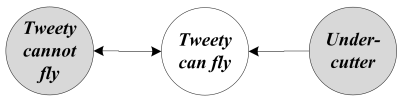

Figure 3. An abstract argumentation framework

To illustrate that argumentation frameworks can have multiple extensions, consider the simpler example in Figure 4 where the undercutting argument has been deleted from the AF of Figure 3. In grounded semantics the extension is empty (case a) but in preferred and stable semantics there are two extensions, depending on whether the argument that Tweety can (case b) or cannot fly (case c) is accepted. Finally, all three extensions are complete.

These examples point at a minor source of terminological confusion, since they use Dung's term 'attack' while Pollock always used 'defeat'. When Dung's 1995 paper appeared, 'defeat' was the standard term, not just in Pollock's work

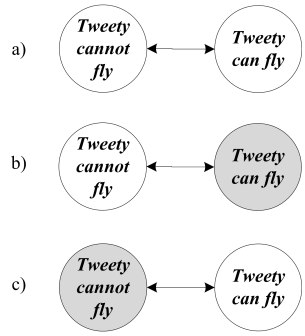

Figure 4. A simpler abstract argumentation framework and three extensions

but essentially in all early work on argumentation-based inference. Current work on the $ASPIC^+$ framework [Prakken, 2010; Modgil and Prakken, 2013; Modgil and Prakken, 2014] also uses 'defeat' and reserves the term 'attack' for more basic, purely syntactical forms of conflicts between arguments. Defeat is then successful attack according to some notion of argument strength or preference, an idea present in much early work on argumentation-based inference, although usually not employing the term 'attack'. Thus it is not $ASPIC^+$'s attack relation but its defeat relation which instantiates Dung's notion of attack.

2.3 Other early work

Initial ideas In the same year in which Pollock published his seminal paper, Loui [1987] appeared as arguably the first AI paper that explicitly proposed to design nonmonotonic logics in the argumentation way. In 1992, Simari and Loui fully formalized Loui's [1987]'s initial ideas, which work in turn led to the development of Defeasible Logic Programming [Garcia *et al.*, 1998; Garcia and Simari, 2004]. One year later, Konolige [1988] proposed an argumentation approach as a solution to the famous Yale Shooting problem in logic-based specifications of dynamic systems [Hanks and McDermott, 1986]. Although his formalism was still rather rudimentary, Konolige's discussion anticipates many issues and distinctions of later work, so that his paper can be regarded as one of the forerunners of the study of argumentation-based inference.

Argumentation as a proof theory for preferential entailment Around 1990, some papers proposed argumentation as a proof theory for model-theoretic notions of nonmonotonic consequence (preferential entailment). Baker and Ginsberg [1989] did this for a minimal-model semantics of prioritised circumscription, while Geffner [1992] and Geffner and Pearl [1992] did the same for their 'conditional entailment' semantics for default reasoning. The basic idea is twofold. First, given a propositional or first-order theory, an argument is a set or conjunction of assumptions consistent with the theory and that combined with the theory yields conclusions; and second, arguments can be attacked by arguments for the negation of the attacked argument or one of its assumptions. This idea later became the basis for assumption-based argumentation [Bondarenko *et al.*, 1997], to be discussed in Section 2.4. Although the idea to found argumentation-based inference on preferential entailment is very interesting, it has since then not been further pursued.

Abstract argumentation systems Lin and Shoham [1989] were the first to propose the idea of abstraction in structured argumentation. They developed the notion of abstract argumentation structures with strict and defeasible rules and they showed how a number of existing nonmonotonic logics could be reconstructed as such structures. Gerard Vreeswijk further developed these ideas into his abstract argumentation systems [Vreeswijk, 1991; Vreeswijk, 1993b; Vreeswijk, 1997]. Since several of Vreeswijk's ideas are included in today's *ASPIC*$^+$ framework, it is worthwhile summarising some of his definitions. Like Lin & Shoham, Vreeswijk defined arguments in terms of an unspecified logical language $\mathcal{L}$, only assumed to contain the symbol $\bot$, denoting 'falsum' or 'contradiction,' and two unspecified sets of strict ($\rightarrow$) and defeasible ($\Rightarrow$) inference rules defined over $\mathcal{L}$. In addition, he defined the main elements that are missing in Lin & Shoham's system, namely, notions of conflict and defeat between arguments. Vreeswijk defined arguments as follows:

Definition 2.1 *An* argument σ *is:*

1. φ *if* $\varphi \in \mathcal{L}$*; in that case:* $\texttt{Prem}(\sigma) = \{\varphi\}$, $\texttt{Conc}(\sigma) = \varphi$, $\texttt{Sent}(\sigma) = \{\varphi\}$*;*

2. $\sigma_1, \ldots \sigma_n \rightarrow \varphi$ *where* $\sigma_1, \ldots, \sigma_n$ *is a finite, possibly empty sequence of arguments such that* $\texttt{Conc}(\sigma_1) = \varphi_1, \ldots, \texttt{Conc}(\sigma_n) = \varphi_n$ *for some strict rule* $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$*, and* $\varphi \notin \texttt{Sent}(\sigma_1) \cup \ldots \cup \texttt{Sent}(\sigma_n)$*; in this case:* $\texttt{Prem}(\sigma) = \texttt{Prem}(\sigma_1) \cup \ldots \cup \texttt{Prem}(\sigma_n)$, $\texttt{Conc}(\sigma) = \psi$, $\texttt{Sent}(\sigma) = \texttt{Sent}(\sigma_1) \cup \ldots \cup \texttt{Sent}(\sigma_n) \cup \{\varphi\}$*;*

3. $\sigma_1, \ldots \sigma_n \Rightarrow \varphi$ *where* $\sigma_1, \ldots, \sigma_n$ *is a finite, possibly empty sequence of arguments such that* $\texttt{Conc}(\sigma_1) = \varphi_1, \ldots, \texttt{Conc}(\sigma_n) = \varphi_n$ *for some defeasible rule* $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$*, and* $\varphi \notin \texttt{Sent}(\sigma_1) \cup \ldots \cup \texttt{Sent}(\sigma_n)$*; with the further attributes defined as in (2).*

Note that this definition, unlike most other definitions of arguments in the formal literature, excludes circular arguments.

Vreeswijk's notion of conflicts between arguments is unusual in that a counterargument is a *set* of arguments: a set Σ of arguments is *incompatible* with an argument τ iff the conclusions of $\Sigma \cup \{\tau\}$ give rise to a strict argument for $\bot$. While unusual, there is nothing obviously wrong with this kind of definition. The reason why currently conflict is usually defined as a relation between individual arguments is probably that such definitions better fit with Dung's theory of abstract argumentation frameworks. Vreeswijk's approach might fit better with generalisations of Dung's theory that allow attacks from sets of arguments to arguments [Bochman, 2003; Nielsen and Parsons, 2007b]. Recently, Baroni *et al.* [2015] have combined the *ASPIC*$^+$ framework with a Vreeswijk-style definition of conflict.

Conflicts can in Vreeswijk's approach be resolved with any reflexive and transitive ordering on arguments that the user likes to adopt. A set of arguments Σ is *undermined* by an argument τ if $\sigma < \tau$ for some $\sigma \in \Sigma$. Then a set of arguments Σ is a *defeater* of σ if Σ is incompatible with σ and not undermined by it.

Finally, Vreeswijk defined argument acceptability ("warrant") with a definition that is close but not equivalent to Dung's [1995] stable semantics. In light of the modern theory of abstract argumentation frameworks, Vreeswijk's definition of warrant is, unlike the rest of his approach, somewhat premature. This is understandable, since Vreeswijk developed his approach before 1995.

Logic-programming approaches The work on argumentation semantics for logic-programming's negation as failure did not only inspire Dung to develop his theory of abstract argumentation frameworks but also gave rise to logic-programming systems for argumentation with explicit negation. Two early papers here were Dung [1993] and Dimopoulos and Kakas [1995]. The first of these papers was in turn a source of inspiration for Prakken and Sartor's [Prakken and Sartor, 1997] argument-based logic programming system with defeasible priorities. Theirs was arguably the first system that was explicitly designed as an instance of Dung's [1995] approach. Strictly speaking, it was technically based not on Dung [1995] but on Dung [1993], but a reformulation in terms of abstract argumentation is trivial. Like all other work reviewed so far, it distinguished between strict and defeasible inference rules. Unlike Dimopoulos and Kakas [1995] but like Dung [1995], its language had both explicit negation and negation as failure, with corresponding "rebutting" attacks on defeasibly derived conclusions and "undercutting" attacks on negation-as-failure premises. One innovative feature was that it allowed argumentation about preferences inside the argumentation system, while another innovative feature was that the system had the first published argument game meant as a proof theory for the semantics of abstract argumentation frameworks (for more on argument games see Section 2.5.2 below).

Defeasible vs. plausible reasoning As apparent from the overview so far, until 1993 almost all accounts of argumentation-based inference made a distinction between deductive (or 'strict') and defeasible inference rules, introduced in philosophy by Pollock [1970; 1974] and in AI by Pollock [1987] and Touretzky [1984]. This approach is still being pursued today, notably in Defeasible Logic Programming, Defeasible Logic and the *ASPIC*$^+$ framework. In this approach a special definition of arguments is needed that regulates the interplay between strict and defeasible reasons (such as the above one of Vreeswijk [1993b; 1997]), since with two kinds of inference rules one cannot rely on a single given logical consequence notion to specify how conclusions are supported by premises. Around 1993 an alternative approach to structured argumentation emerged, according to which arguments are constructed in a single given deductive logic, obviating the need of a separate definition of an argument beyond being a premises-conclusion pair. In understanding and relating the two approaches, the philosophical distinction between *plausible* and *defeasible* reasoning is relevant; cf. Rescher [1976; 1977] and Vreeswijk [1993b], Ch. 8. Following Rescher, Vreeswijk described plausible reasoning as sound (i.e., deductive) reasoning on an uncertain basis and defeasible reasoning as unsound (but still rational) reasoning on a solid basis. In other words, argumentation models of plausible reasoning locate all fallibility of an argument in its premises, while argumentation models of defeasible reasoning locate all fallibility in its defeasible inferences. Thus plausible-reasoning approaches effectively view argumentation as a kind of inconsistency handling, since in these approaches conflicts between arguments can only arise if the knowledge base is inconsistent. By contrast, in defeasible-reasoning approaches conflicts can arise from consistent knowledge bases, since in those approaches it is the application of defeasible rules that makes an argument fallible.

Two groups in particular initiated the plausible-reasoning approach to argumentation, respectively at Queen Mary's University in London and at INRIA in Toulouse. Elvang-Göransson *et al.* [1993] conceived of arguments as premise-conclusion pairs (δ, p) where δ is a subset of a possibly inconsistent database Δ and there exists a natural-deduction proof of p from δ. Arguments can be attacked in two ways: an argument (δ', q) *rebuts* (δ, p) if q is logically equivalent to $\neg p$ and it *undercuts* it if q is logically equivalent to $\neg r$ for some $r \in \delta$. Note that Elvang-Göransson *et al.* thus introduced a terminological confusion into the literature that exists until today. While they fully adopted Pollock's [1974; 1987]'s terminology, they only partly adopted its meaning, since Pollock used the term 'undercutter' not for premise attack but for attack on the application of a defeasible inference rule. Today, Pollock's meaning of the term 'undercutter' is adopted in the $ASPIC^+$ framework and Dung's recent work on structured argumentation frameworks, while Elvang-Göransson *et al.*'s meaning is fashionable in work on classical and Tarskian argumentation.

Elvang-Göransson *et al.* classified arguments into five classes of increasing degrees of acceptability: arguments, consistent arguments (i.e., arguments with consistent premises), non-rebutted consistent arguments, non-rebutted and non-undercut consistent arguments, and "tautological" arguments (i.e., arguments with an empty set of premises). In light of modern work this definition of argument acceptability seems somewhat ad-hoc. Among other things, it does not model the notions of defense and admissibility that are so beautifully modelled by Dung [1995]. The ideas of Elvang-Göransson *et al.* were further developed by Krause *et al.* [1995], replacing classical logic by intuitionistic logic as the underlying logic and adding notions of argument structure and argument strength.

Around the same time as Elvang-Göransson *et al.*, Benferhat *et al.* [1993] proposed a similar system, containing what now is the standard definition of an argument in this approach, adding to Elvang-Göransson *et al.*'s definition the requirements that the set of premises is consistent and subset-minimal:

Definition 2.2 *Given a database* Σ, *a set* $\Sigma_i \subseteq \Sigma$ *is an argument for a formula* φ *iff:*

1. $\Sigma_i \not\vdash \bot$; *and*
2. $\Sigma_i \vdash \varphi$; *and*
3. *for all* $\psi \in \Sigma_i$: $\Sigma_i \setminus \{\psi\} \not\vdash \varphi$

Here, $\vdash$ denotes classical propositional consequence. Benferhat *et al.* did not define explicit notions of attack. Instead they defined φ to be an *argumentative consequence* of Σ if given Σ there exists an argument for φ but not for $\neg\varphi$. They also studied alternative consequence notions and their relations, and refined their system with a preference relation on the database. Their approach was related to abstract argumentation by Cayrol [1995], who among other things proved that with Elvang-Göransson *et al.*'s undercutting relation as the attack relation, the stable extensions given a database are in a one-to-one correspondence with the database's maximal consistent subsets. This result was later generalised by Amgoud and Besnard [2013] for any abstract Tarskian logic and by Modgil and Prakken [2013] in the context of the $ASPIC^+$ framework.

The ideas of Elvang-Göransson *et al.* and Benferhat *et al.* were picked up by e.g. Amgoud and Cayrol [1998] and Besnard and Hunter [2001] and evolved into classical, or classical-logic argumentation [Besnard and Hunter, 2008; Gorogiannis and Hunter, 2011, e.g.] and its generalisations to deductive [Besnard and Hunter, 2014] and abstract Tarskian argumentation [Amgoud and Besnard, 2013], to be further discussed below.

2.4 Structured argumentation: developments until now

While until 1995 work on structured argumentation had specific and sometimes ad-hoc definitions of argument evaluation, since 1995 most work on structured argumentation adopts Dung's approach or at least explicates the relation with it. Work that adopts Dung's approach does so by giving definitions of the structure of arguments and the nature of attack. Thus abstract argumentation frameworks are generated, so that arguments can be evaluated according to one of the abstract argumentation semantics and their acceptability status can be used to define nonmonotonic consequence notions for their statements. However, there is also work that deviates from Dung's approach. In this section I will give an overview of these research strands.

2.4.1 Argumentation models of plausible reasoning

Current argumentation models of plausible reasoning are essentially of two kinds.

Assumption-based argumentation Around the same time as argumentation was proposed as a way of inconsistency handling in classical logic, assumption-based argumentation (ABA) emerged from attempts to give an argumentation-theoretic semantics to logic-programming's negation as failure [Bondarenko *et al.*, 1993; Bondarenko *et al.*, 1997]. Like the classical-logic approaches, ABA also assumes a unique 'base logic', which in ABA is called a "deductive system", consisting of set of inference rules defined over some logical language. Given a set of so-called 'assumptions' formulated in the logical language, arguments are then deductions of claims using rules and supported by sets of assumptions. Contrary to in classical and abstract argumentation, the premises of ABA arguments, i.e., its assumptions, do not have to be consistent. ABA leaves both the logical language and set of inference rules unspecified in general, so it is like Vreeswijk's [1993b; 1997] approach and the later $ASPIC^+$ framework, an abstract framework for structured argumentation. However, unlike these approaches, ABA only allows attacks on an argument's assumptions, so that ABA's rules are effectively equivalent to Vreeswijk's and $ASPIC^+$'s strict inference rules (as formally confirmed in [Prakken, 2010]).

In order to express conflicts between arguments, ABA makes like Vreeswijk a minimum assumption on the logical language, which in ABA is that each assumption in the logical language has a *contrary.* That b is a contrary of a, written as $b = \overline{a}$, informally means that b contradicts a. An argument using an assumption a is then attacked by any argument for conclusion $\overline{a}$. Contrary relations do not have to be symmetric. This feature allows an argumentation-theoretic semantics for negation as failure (*not*) by for every formula *not p* letting $p = \overline{not\ p}$ but not vice versa. However, ABA's application is not limited to logic programming; in the landmark ABA paper [Bondarenko *et al.*, 1997], it is instantiated with various nonmonotonic logics, including default logic, circumscription and Poole's [1989] Theorist system.

Although ABA and Dung's approach clearly have commonalities, ABA as originally formulated by Bondarenko *et al.* [1997] does not generate abstract argumentation frameworks. Instead, its extensions are (in some sense maximal) sets of assumptions, induced by transforming attack relations between arguments to attack relations between sets of assumptions. Only ten years later was ABA given an explicit Dungean formulation by Dung *et al.* [2007]. Currently, there is some controversy about whether the correspondence holds for all current abstract argumentation semantics or not; cf. Gabbay [2015] and Caminada [2015].

ABA was originally used theoretically as a framework for nonmonotonic logic. Over the years, the focus has shifted somewhat to developing algorithms and implementations and to applying these to a wide range of reasoning and decision problems. For more details the reader is referred to the other chapters in this handbook.

An interesting variant of assumption-based argumentation is Verheij's [2003] DefLog system. Verheij assumes a logical language with just two connectives, a unary connective $\times$ which informally stands for 'it is defeated that' and a binary connective $\rightsquigarrow$ for expressing defeasible conditionals. Verheij then assumes a single inference scheme for this language, namely, modus ponens for $\rightsquigarrow$. A set of sentences T is said to *support* a sentence φ if φ is in T or follows from T by repeated application of $\rightsquigarrow$-modus ponens. Moreover, T is said to *attack* φ if T supports $\times\varphi$. Verheij then considers partitions (J, D) of sets of sentences Δ such that J (the "justified" sentences) is conflict-free and attacks every sentence in D (the "defeated" sentences). As observed by Verheij, DefLog can be encoded as an ABA instance with stable semantics by setting ABA's assumptions to Δ, defining the ABA ABF contrary mapping as $\times\varphi = \overline{\varphi}$ for any φ and letting ABA's set of rules be generated by the modus scheme for $\rightsquigarrow$.

Classical, deductive and Tarskian argumentation The initial work of Elvang-Göransson *et al.* [1993] and Benferhat *et al.* [1993] led to a family of approaches usually called 'classical' or 'deductive' argumentation [Amgoud and Cayrol, 2002; Besnard and Hunter, 2001; Kaci *et al.*, 2007; Besnard and Hunter, 2008; Amgoud and Vesic, 2010; Kaci, 2010]. The first name refers to instances with as base logic classical propositional or first-order logic, while the term 'deductive argumentation' is used for approaches that abstract from particular base logics, as long as they are "deductive". Often the term 'deductive' is here used in an informal sense. For example, Besnard and Hunter [2014] describe a deductive inference as an inference that is "infallible in the sense that it does not introduce uncertainty". This agrees with Pollock's notion of a deductive reason. Recently Amgoud and Besnard [2010; 2013] gave a precise interpretation by assuming that the base logic satisfies the properties of a so-called Tarskian abstract logic.

In all these approaches arguments are, as in Benferhat *et al.* [1993] for the special case of classical propositional logic, premises-conclusion pairs such that the premises are, according to the base logic, consistent and subset-minimal sets logically implying their conclusion. Unlike in many other approaches, these approaches do not commit to specific definitions of argument attack but explore the consequences of various definitions, all exhibiting some form of premise- and/or conclusion attack. Given that these approaches locate all fallibility of arguments in their premises, one might expect that definitions that only allow premise attack are the best-behaved. This was formally confirmed by Gorogiannis and Hunter [2011] and Amgoud and Besnard [2013] who, for respectively classical and Tarskian argumentation, showed that when abstract argumentation frameworks are generated, only particular forms of premise attack fully guarantee the consistency of the conclusion sets of extensions of abstract argumentation frameworks.

Until these investigations, research in this strand was not much concerned with argument evaluation. Instead, other properties were studied, such as relations between kinds of attack, and the formalisms were used as a tool for investigating dialogue-related questions, such as enthymemes [Black and Hunter, 2012] and persuasive force of arguments [Hunter, 2004]. See for further details Besnard and Hunter [2008] and other chapters in this handbook.

2.4.2 Argumentation models of defeasible reasoning

Defeasible Logic Programming Defeasible Logic Programming, or DeLP [Garcia *et al.*, 1998; Garcia and Simari, 2004] is a further development of Simari and Loui's [1992] argumentation system with strict and defeasible rules. While Simari and Loui only allowed specificity as a source of preferences, DeLP allows any preference ordering. DeLP's logic-programming rules can contain both explicit negation and negation as failure. It is noteworthy that while the consequence notion of Simari and Loui's system is equivalent to Dung's [1995] grounded semantics, DeLP as described by Garcia *et al.* [1998] and Garcia and Simari [2004] does not conform to any of Dung's semantics. Instead, it is based on the notion of a dialectical tree, which essentially captures all ways in which a proponent and an opponent of a claim can have a debate about the claim by defeating each other's arguments. This notion is very similar to the notion of an argument game as a proof theory for the semantics of abstract argumentation frameworks (see further Section 2.5.2). However, while the constraints on argument games are based on the semantics for abstract argumentation frameworks, DeLP's constraints on dialectical trees are based on intuitions concerning concrete examples.

A unifying approach: the $ASPIC^+$ framework The $ASPIC^+$ framework [Prakken, 2010; Modgil and Prakken, 2013; Modgil and Prakken, 2014] unifies plausible and defeasible reasoning. Its main sources of inspiration are the systems of Pollock [1987; 1994; 1995] and Vreeswijk [1993b; 1997], which model defeasible reasoning. However, $ASPIC^+$ adds to these systems the possibility to attack an argument's premises, which makes it also suitable for modelling plausible reasoning. Apart from this, $ASPIC^+$ adopts Pollock's distinction between deductive (strict) and defeasible inference rules, Vreeswijk's definition of an argument and Pollock's notions of rebutting and undercutting attack, with the exception that in $ASPIC^+$, unlike in Pollock's systems, undercutting attack succeeds as defeat irrespective of preferences. Also, like Vreeswijk, $ASPIC^+$ abstracts from particular logical languages, sets of inference rules and argument orderings. Unlike Vreeswijk's particular method of argument evaluation, $ASPIC^+$ generates abstract argumentation frameworks, so that any semantics for such frameworks can be used to evaluate arguments.

A preliminary version of $ASPIC^+$ was developed during the EC-sponsored ASPIC project, which ran from 2004 to 2007. This version was used by Caminada and Amgoud [2007] as a vehicle for proposing the idea of rationality postulates for structured argumentation. The first publication focusing on $ASPIC^+$ as a framework for structured argumentation was Prakken [2010]. Modgil and Prakken [2013] proposed some small modifications and variations and proved further results on the framework and its relation with other work. Recently, several other variations of the $ASPIC^+$ framework have been studied, which are further described in this handbook's chapter on rule-based argumentation.

Its abstract nature makes that $ASPIC^+$ can be instantiated in many different ways and captures a number of other approaches as special cases. For example, Prakken [2010] proves that Dung *et al.*'s [2007]'s version of assumption-based argumentation can be reconstructed as a special case of $ASPIC^+$ with only strict inference rules, no unattackable premises and no preferences. And Modgil and Prakken [2013] reconstruct two forms of classical argumentation as studied by Gorogiannis and Hunter [2011] as the special case with only strict rules, being all valid classical inferences from finite sets, no unattackable premises, no preferences and the constraint that an argument's premises are classically consistent and subset-minimal. They then generalise this reconstruction with a preference relation on the knowledge base and prove that the resulting stable extensions are in a one-to-one correspondence with Brewka's [1989] preferred subtheories. Thus they also extend Cayrol's [1995] similar result without preferences for maximal consistent subsets.

Not only $ASPIC^+$ but also assumption-based argumentation is an abstract model of structured argumentation. Compared to ABA, $ASPIC^+$ is more complex, with its two kinds of inference rules, its three kinds of attack and its explicit preferences to distinguish between attack and defeat. As stated by Toni [2014], the philosophy behind ABA is instead to translate preferences and defeasible rules into ABA rules plus ABA assumptions, so that rebutting and undercutting attack and the application of preferences all reduce to premise attack. This approach has its merits but it is an open question whether $ASPIC^+$ can in its full generality be translated into ABA. Currently there are only partial answers to this question. Dung and Thang [2014] prove for the case without preferences that defeasible $ASPIC^+$ rules can be translated to ABA rules with assumption premises. Moreover, in an early paper, Kowalski and Toni [1996] give a partial method for encoding rule preferences with explicit assumption premises. However, it remains to be seen whether this can be done for any argument ordering. Moreover, $ASPIC^+$ representations of examples are often arguably closer to natural-language than ABA presentations, in which every conflict has to be translated to premise attack and every preference statement to explicit exceptions. If the aim is to formalise modes of reasoning in a way that corresponds with human modes of reasoning and debate, then there is some merit in having a theory with explicit notions of rebutting and undercutting attack and preference application.

2.4.3 The study of rationality postulates

An important recent development is the introduction by Caminada and Amgoud [2005; 2007] of the idea of rationality postulates for structured argumentation. According to Caminada and Amgoud, all systems of structured argumentation that have notions of negation, strict rules and subarguments should satisfy the following properties:

> **Sub-argument Closure**: For any argument A in E, all sub-arguments of A are in E.
>
> **Closure under Strict Rules**: If E contains arguments with conclusions

$\alpha_1, \ldots .\alpha_n$, then any arguments obtained by applying only strict inference rules to these conclusions, are in E.

Direct Consistency: The set of conclusions of all arguments in E are directly consistent, i.e., it contains no pair of formulas φ and $\neg\varphi$.

Indirect Consistency: The set of conclusions of all arguments in E are indirectly consistent, i.e., its closure under strict rules is directly consistent.

$ASPIC^+$ unconditionally satisfies closure under subarguments. Whether $ASPIC^+$ satisfies closure under strict rules and the consistency postulates depends on whether the non-attackable premises are consistent, on structural properties of the strict rules and on properties of the argument ordering [Caminada and Amgoud, 2007; Prakken, 2010; Modgil and Prakken, 2013]. These results on $ASPIC^+$ directly generalise to systems that can be reconstructed within $ASPIC^+$, such as assumption-based argumentation and several forms of classical and deductive argumentation with preferences. Recently, Dung and Thang [2014] identified alternative and partly weaker sufficient conditions for satisfying strict closure and consistency.

Three further rationality postulates were proposed by Caminada *et al.* [2012] and are about the extent to which contradictions can trivialise the set of conclusions. These postulates have been further studied by Wu and Podlaszewski [2015].

Although Caminada and Amgoud defined their postulates for rule-based systems, they can be straightforwardly adapted to systems that define argument structure in terms of consequence notions instead of inference rules, such as classical and deductive argumentation. In particular the consistency postulates have been studied for these approaches [Gorogiannis and Hunter, 2011; Amgoud and Besnard, 2013]. One insight here (of which the core is already in Caminada and Amgoud [2007]) is that satisfaction of the consistency postulates partly depends on the definitions of attack and defeat. Building on this idea, Dung [2014; 2016] proposes several desirable properties for defeat relations (which in line with his 1995 paper he calls 'attack' relations) and studies their effect on satisfaction of the consistency postulates.

Finally, the recent research on rationality postulates is reminiscent of work in other areas of nonmonotonic logic on general properties of nonmonotonic consequence notions [Gabbay, 1985; Kraus *et al.*, 1990; Makinson, 1994]. One much discussed property in that body of work is *cautious monotony*. Informally, this property is that if φ and ψ are implied by a knowledge base and φ is added to the knowledge base, then ψ is still implied by the new knowledge base. Recently, Dung [2014; 2016] has argued that this property should hold for credulous argumentation-based inference, i.e., for membership of at least one extension. By contrast, Prakken and Vreeswijk [2002], Section 4.4 argue that satisfaction of this property is not desirable in general, since strengthening a nonmonotonic conclusion to an indisputable fact can give arguments using the fact the power to defeat other arguments that they did not have before; and this may well result in the loss of the extension from which the conclusion was promoted to an indisputable fact.

2.4.4 Preferences and argument strength

An important element in many argumentation systems is the use of some notion of preference or strength to resolve conflicts between arguments. In Dungean terms, this boils down to defining his attack relation in terms of a more basic, non-evaluative notion of conflict between arguments and some binary preference relation on arguments. As noted above, most work before Dung [1995] used the term 'defeat' instead of 'attack' while much work after 1995 explicitly renamed Dung's attack relation to 'defeat' in order be able to call the more basic, non-evaluative notion of conflict 'attack'. This is what I will also do in this section. The use of preferences then amounts to checking which attacks succeed as defeats.

Arguably the first systems embodying some form of argument preference were the inheritance systems of Touretzky [1984] and Horty *et al.* [1990], which used syntactic specificity checks on inheritance paths to let inheritance paths from more specific classes defeat conflicting inheritance paths form more general classes. Loui [1987] and Simari and Loui [1992] also used specificity for conflict resolution.

Although Pollock's earliest system, from 1987, did not yet include a notion of strength, Pollock later took the notion of strength of arguments very seriously. Since his systems were meant for epistemic reasoning, he always formulated strength in terms of numerical degrees of belief. His approach here was non-standard. Against Bayesian approaches, he argued that degrees of belief and justification do not conform to the laws of probability theory. In his [1994, 1995], Pollock used a weakest-link approach to compute the strength of arguments: given numerical strengths of reasons (where deductive reasons have infinite strength), the strength of an argument's conclusion is the minimum of the strengths of the reason with which the conclusion is derived and the strengths of the intermediate conclusions to which this reason is applied. While thus arguments can have various strengths, defeat is still an all-or-nothing matter in that defeaters that are weaker than their target cannot affect the status of their target at all. This allows a reconstruction of Pollock's [1994, 1995] approach in terms of Dung's theory of abstract argumentation frameworks. Later, in his [2002, 2007a, 2010] Pollock explored the idea that weaker defeaters can still weaken the justification status of their stronger targets. To formalize this, he now made the justification status of statements a matter of numerical degree, being a function of the strengths of both supporting and defeating arguments. Thus in his latest work he deviated from a Dungean approach.

Similar to Pollock's [1994; 1995] way to use degrees of belief is Chesñevar *et al.*'s [2004] use of possibilistic logic in the context of Defeasible Logic Programming. In this paper, possibilistic strengths are added to rules, which are propagated through arguments according to possibilistic logic. Then the propagated strengths are used to resolve attacks into defeats.

Other early work resolved attacks with qualitative preference relations on premises or inference rules. One of the first argumentation models of defeasible reasoning with rule preferences from arbitrary sources was Prakken [1993], developed into Prakken and Sartor [1997]. One of the first argumentation models of plausible reasoning with prioritized knowledge bases was Benferhat *et al.* [1993]. Amgoud and Cayrol [1998; 2002] combined Benferhat *et al.*'s idea of prioritised knowledge bases and Cayrol's [1995] Dungean modelling of classical argumentation with Prakken and Sartor's way to distinguish between attack and defeat in Dung's grounded semantics and their argument game for it. Later papers included preferences in classical argumentation in other ways; e.g. Amgoud and Vesic [2010] and Kaci [2010].

Vreeswijk [1993a; 1997] was the first to include a binary argument ordering as primitive in his approach. The $ASPIC^+$ framework adopts this idea and several papers on $ASPIC^+$ study instantiations with qualitative preference relations on defeasible rules and attackable premises, building on the work of Benferhat *et al.* [1993], Prakken and Sartor [1997] and their successors. Recently, Dung [2014; 2016] has also contributed to this study.

Since there is not a unique kind of content of arguments, there is also not a unique kind of argument preference. In epistemic reasoning, argument preferences are often based on probabilistic considerations, degrees of belief, or on credibility estimates of information sources. In argumentation as decision making they have been based on preferences for decision outcomes. In normative (legal or moral reasoning) they have been derived from hierarchical relations between elements of normative systems. In addition, some have modelled argumentation *about* preference relations within argumentation logics. One of the first proposals of this kind was made by Prakken and Sartor [1997]. Modgil [2009] extended abstract argumentation frameworks with the possibility to attack attacks. Modgil then, among other things, showed that Prakken and Sartor's proposal can be reconstructed as an instance of his 'extended argumentation frameworks'.

One question here is whether preference relations logically behave the same regardless of their source. Dung [2016] seems to answer this question affirmatively, while Modgil and Prakken [2014] suggest that the right way to use preferences may depend on the kind of content of arguments, for example, on whether the reasoning is epistemic, normative or about decision making.

2.5 Abstract argumentation: developments into now

In the first years after publication of Dung's landmark paper it gave rise to two kinds of follow-up work. Some continued to use AFs as Dung did in his paper, namely, to reconstruct and compare existing systems for structured argumentation as instances of AFs. In line with this was work on developing new systems for structured argumentation as instances of AFs. Others further developed the theory of abstract argumentation frameworks in the form of proof of properties (such as complexity results), reformulations (e.g. in terms of labellings), argument games as a proof theory, and algorithms. Somewhat later a third kind of follow-up work emerged, namely, extending AFs with new elements without specifying the structure of arguments. I now briefly review these three bodies of work.

2.5.1 Instantiating abstract argumentation frameworks

Some continued Dung's work on reconstructing and comparing existing systems for structured argumentation as instances of AFs. For example, Jakobovits [2000] and Jakobovits and Vermeir [1999b] showed that Pollock's [1994; 1995] system for defeasible reasoning has preferred semantics and Cayrol [1995] related various forms of classical argumentation to Dung's stable semantics and (with Amgoud in [Amgoud and Cayrol, 2002]) to Dung's grounded semantics for AFs. More recent work in this vein is Gorogiannis and Hunter [2011] and Amgoud and Besnard [2013].

Others developed new systems for structured argumentation as an instantiation of abstract argumentation frameworks. As described above, possibly the first system developed in this way was Prakken and Sartor's [1997] system for argumentation-based logic programming. More recently, the $ASPIC^+$ framework was designed in this way.

2.5.2 Developing the theory of abstract argumentation frameworks

Labellings A few years after Dung introduced his extension-based approach to abstract argumentation, an alternative labelling-based approach became popular, based on the following definition:

> A *labelling* of an $AF = \langle AR, attacks \rangle$ assigns to zero or more members of AR either the status *in* or *out* (but not both) such that:
>
> 1. an argument is *in* iff all arguments attacking it are *out*.
> 2. an argument is *out* iff it is attacked by an argument that is *in*.
>
> Let $In = \{A \in AR \mid A \text{ is } in\}$ and $Out = \{A \in AR \mid A \text{ is } out\}$ and $Undecided = AR \setminus (In \cup Out)$. Then
>
> 1. A labelling is *stable* if $Undecided = \emptyset$.
> 2. A labelling is *preferred* if *Undecided* is minimal (wrt set inclusion)
> 3. A labelling is *grounded* if *Undecided* is maximal (wrt set inclusion)
> 4. Any labelling is *complete*.

These notions coincide with Dung's extension-based definitions as follows. Let $S \in \{\text{stable, preferred, grounded, complete}\}$. Then (In, Out) is an S-labelling iff In is an S-extension.

To illustrate the labelling definition, in Figure 3 the grey-white colourings correspond to the *in*-*out* labels in the unique stable/preferred/grounded/complete labelling. In Figure 4(b,c) the grey-white colourings correspond to the *in-out* labels of the two stable-and-preferred labellings but in Figure 4(b,c) both arguments are undecided.

Actually, Pollock was a source of inspiration here too, since he used a labelling definition in his [1994; 1995] system. Pollock was possibly in turn inspired by Doyle's [1979] justification-based truth maintenance systems. Pollock's 1994 system was, as just noted, by Jakobovits [2000] proved to be an instance of Dung's preferred semantics. Jakobovits' PhD thesis contains an in-depth investigation of the labelling approach, summarised by Jakobovits and Vermeir [1999b]. Other early work on labellings was done by Verheij [1996] and the labelling approach was finally popularised by Caminada [2006].

Argument games Both the extension- and the labelling-based approach can be regarded as a semantics of argumentation-based inference in that the main focus is on characterising properties of *sets* of arguments, without specifying procedures for determining whether a given argument is a member of the set. The proof theory of argumentation-based inference amounts to specifying such procedures. An elegant form of such a proof theory is that of an *argument game* between a proponent and an opponent of an argument. The precise rules of the game depend on the semantics the game is meant to capture. The rules should be chosen such that the existence of a winning strategy (in the usual game-theoretic sense) for the proponent of an argument corresponds to the investigated semantic status of the argument, for example, 'being in the grounded' or 'being in at least one (or in all) preferred extensions'.

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

The idea of argument games had been around since the beginning of the formal study of argumentation (see e.g. Vreeswijk [1993a]) but they were not formally linked to argumentation-based semantics until the mid 1990s. Dung [1995] refers to a technical report [Dung, 1992] that was never formally published and in which he proposed argument games for two logic-programming semantics. Prakken and Sartor [1997] proposed an argument game for their logic-programming instantiation of Dung's grounded semantics. Arguably the first publication on argument games for abstract argumentation semantics was Prakken [1999], who proposed the above game for grounded semantics as an abstraction of the game of Prakken and Sartor. Vreeswijk and Prakken [2000] proposed argument games for preferred semantics, which were further developed and studied by Dunne and Bench-Capon [2003].

New semantics and general study of semantics While Dung [1995] originally proposed four semantics for abstract argumentation frameworks, in later years several alternative semantics were proposed; cf. Baroni *et al.* [2011a]. A related development is the study of general characterisations of types of semantics and their properties and relations, initiated by Baroni and Giacomin [2007] and further pursued by e.g. Dvorak and Woltran [2011] and Baroni *et al.* [2014]. Baroni and Giacomin [2007] also had a normative aim, namely, to propose a set of principles for the evaluation of semantics for abstract argumentation frameworks. Thus their work can be seen as an abstract counterpart of Caminada and Amgoud's [2007] introduction of rationality postulates for structured argumentation formalisms (see Section 2.4.3 above). Part E of Volume 1 of this handbook reviews this line of research in detail.

Complexity results and algorithms The graph-based format of abstract argumentation frameworks naturally lends itself to studies of computational complexity. A leading figure here has been Paul Dunne [Dunne and Bench-Capon, 2002; Dunne and Bench-Capon, 2003; Dunne, 2007].

Algorithms for proof theories for abstract argumentation frameworks were proposed by e.g. Cayrol *et al.* [2003], Vreeswijk [2006] and Verheij [2007]. Early work on algorithms for enumerating extensions or labellings is reviewed by Modgil and Caminada [2009]. An interesting strategy for developing algorithms is encoding argumentation frameworks in some other formalism and to utilise algorithms for the other formalism. For example, Besnard and Doutre [2004] encoded abstract argumentation frameworks in propositional logic in order to apply model-checking and SAT solver techniques. They also proposed an equation checking approach, which was later further developed by Gabbay [2011]. Some other examples of this approach are Grossi's [2010] encoding of abstract argumentation frameworks in modal logic and Egli *et al.*'s [2010] encoding in answer set programming.

2.5.3 Adding new elements to abstract argumentation frameworks

A third research strand in the abstract approach to argumentation is to extend AFs with new elements without specifying the structure of arguments. In this subsection I briefly discuss various ways in which this has been done.

Adding preferences or values Amgoud and Cayrol [1998] added to abstract argumentation frameworks a a preference relation on AR, resulting in *preference-based argumentation frameworks* ($PAFs$), which are a triple $\langle AR, attacks, \preceq \rangle$. An argument A then *defeats* an argument B if A attacks B and $A \not\prec B$. Thus each PAF generates an AF of the form $\langle AR, defeats \rangle$, to which Dung's theory of abstract argumentation frameworks can be applied.

Bench-Capon [2003] proposed a variant of idea called *value-based argumentation frameworks* ($VAFs$), in which each argument is said to promote some value. The notion of value should be taken here not in a numerical sense but in the sense of, for example, legal, moral or societal values, such as welfare, equality, fairness, certainty of the law, freedom of speech, privacy, and so on. Attacks are in $VAFs$ resolved in terms of one or more orderings on the values.

These value orderings are assumed to be provided by an audience evaluating the arguments.

Adding abstract support relations There have been several recent proposals to extend Dung's [1995] well-known abstract argumentation frameworks (AFs) with abstract support relations, such as Cayrol and Lagasquie-Schiex's [2005b; 2009; 2013] Bipolar Argumentation Frameworks (BAFs), the work of Martinez *et al.* [2006] and Oren and Norman's [2008] Evidential Argumentation Systems (EASs). Various semantics for such frameworks have been defined, claimed to capture different notions of support. For example, Martinez *et al.* want to abstract from subargument relations in systems for structured argumentation. Boella *et al.* [2010a] study semantics of what they call "deductive" support, which satisfies the constraint that if A is acceptable and A is a deductive support of B, then B is acceptable. Nouioua and Risch [2011] consider "necessary support", which satisfies the constraint that if B is acceptable and A is a necessary support of B, then A is acceptable.

Other additions Both Bochman [2003] and Nielsen and Parsons [2007b] generalised Dung's attack relation to a relation from *sets* of arguments to arguments. As noted above, Modgil [2009] extended abstract argumentation frameworks with attacks on attacks, as an abstraction of earlier proposals to model reasoning about priorities in nonmonotonic logics. Coste-Marquis *et al.* [2006] added constraints to argumentation frameworks in the form of propositional encodings of properties of extensions. Finally, Dunne *et al.* [2011] added weights to attacks, the idea being that attacks that are of insufficient weight (modelled by a "weight budget") can be ignored.

A word of caution Although it is tempting to extend abstract argumentation frameworks with additional elements, a word of caution is in order. One should resist the temptation to think that for any given argumentation phenomenon the most principled analysis is at the level of abstract argumentation frameworks. In fact, it often is the other way around, since at the abstract level crucial notions like claims, reasons and grounds are abstracted away.

An example where this leads to problems is the way preferences are used in $PAFs$ and $VAFs$ to resolve attacks. As shown in work on structured argumentation with preferences (e.g. Pollock's or Vreeswijk's system, $ASPIC^+$ or DeLP), the structure of arguments is crucial in determining how preferences must be applied to attacks. Consider the following semi-formal example adapted from Prakken [2012] and Modgil and Prakken [2013], which can easily be formalised in any of the above-mentioned systems for structured argumentation.

$A = p$
$B_1 = \neg p$
$B_2 = \neg p$, therefore, *presumably*, q

Here p and $\neg p$ are default assumptions. Note that B_1 is a subargument of B_2, so B_2 includes B_1 as part of itself. The arguments with their internal structure and their direct attack relations are displayed in Figure 5. In any

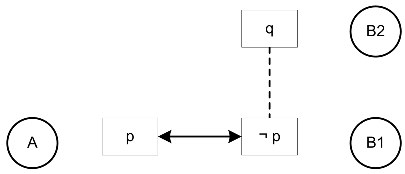

Figure 5. Argument structure and direct attack

of the above systems for structured argumentation we then have that A and B_1 directly attack each other while, moreover, A indirectly attacks B_2, since it directly attacks B_2's subargument B_1. So we have the abstract argumentation framework displayed in Figure 6(a).

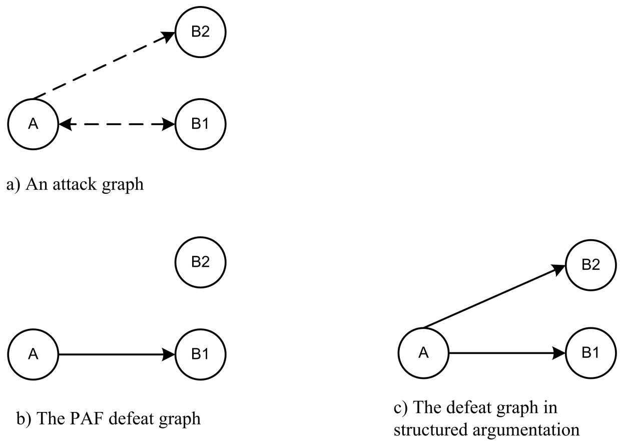

Figure 6. The abstract attack and defeat graphs

Assume next that A is preferred over B_1 and B_2 is preferred over A. Such an ordering could, for instance, be the result of comparing arguments according to their last fallible elements. A *PAF* modelling then generates the following single defeat relation: A defeats B_1; see Figure 6(b). Then we have a single extension (in whatever semantics), namely, $\{A, B_2\}$. So not only A but also B_2 is justified.

However, this violates Caminada and Amgoud's [2007] rationality postulate of subargument closure of extensions, since B_2 is in the extension while its subargument B_1 is not. (Prakken [2012] also discusses examples in which the postulate of indirect consistency is violated.) The cause of the problem is that the *PAF* modelling of this example cannot recognise that the reason why A

attacks B_2 is that A directly attacks B_1, which is a subargument of B_2. So the PAF modelling fails to capture that in order to check whether A's attack on B_2 succeeds, we should compare A not with B_2 but with B_1. Now since $B_2 \prec A$ we also have that A defeats B_2; see Figure 6(c). So the single extension (in whatever semantics) is $\{A\}$, so closure under subarguments is respected.

This shows that *PAFs* (and also *VAFs*) only behave correctly under the assumption that all attacks are direct. We can conclude that for a principled analysis of the use of preferences to resolve attacks, the structure of arguments must be made explicit.

More generally, this analysis shows that in proposing an abstract model of argumentation, it is important to be aware what is abstracted from. Yet in the study of abstract support relations there is, unlike with Dung's original abstract frameworks, hardly any formal study of the relation between the abstract and the structured level. In consequence, it remains unclear what exactly is being modelled. One of the few studies in this vein is my [Prakken, 2014], in which I studied to what extent bipolar and evidential abstract frameworks can be interpreted as abstracting from the inferential relations in structured argumentation, as captured in $ASPIC^+$ by its subargument relations. I obtained mixed results. A form of BAFs that by Boella *et al.* [2010a] was claimed to be suitable for "deductive support" turned out to have no relation with classical-logic approaches to structured argumentation but Oren and Norman's [2008] evidential frameworks turned (for preferred semantics) out to be a suitable abstraction of $ASPIC^+$'s subargument relation. The same holds (for all four of Dung's [1995] semantics) for Dung and Thang's [2014] proposal. They add a binary support relation to abstract argumentation frameworks with the sole additional constraint that if B supports C and A attacks B then A also attacks C, and they then evaluate arguments as in Dung [1995] by only taking the thus constrained attack relation into account. The resulting system conforms to Nouioua and Risch's [2011] notion of "necessary support". Apart from these results it is still an open question what abstract models of argumentation with support relations abstract from.

These discussions lead me to propose a (to some readers possibly controversial) methodological guideline that every new proposal for extending abstract argumentation frameworks should in the same paper be accompanied by at least one non-trivial instantiation in order to demonstrate the significance of the new extension. Work that respects this guideline, respects the historic origins of the abstract study of argumentation, since the prime example of how this guideline can be applied is Dung [1995], who instantiated his frameworks with four nonmonotonic logics. A more recent example is Modgil [2009], who showed that his 'extended argumentation frameworks', which extend abstract argumentation frameworks with attacks on attacks, can be instantiated with Prakken and Sartor's [1997] modelling of reasoning about preferences.

2.6 Further developments

I now briefly sketch some important further developments in the formal study of argumentation as inference.

2.6.1 More recent graph-based approaches

Since 2007 several graph-based approaches have been proposed, in which not arguments and their relations but statements and their relations are the main focus of attention. This idea also goes back to the work of Pollock, since the system of Pollock [1994] is strictly speaking not formalised in terms of arguments but in terms of so-called 'inference graphs', in which nodes are connected either by inference links (applications of inference rules) or by defeat links. The nodes are 'lines of argument', which are propositions plus an encoding of the argument lines from which they are derived. Nodes are evaluated in terms of the recursive structure of the graph. As noted above, Jakobovits [2000] proved that Pollock's system can be given an equivalent formulation as an instance of Dung's abstract argumentation frameworks with preferred semantics.

Gordon *et al.* [2007] proposed the Carneades framework 'of argument and burden of proof'. Carneades' main structure is that of an argument graph, which, despite its name, is similar to Pollock's inference graphs. Statement nodes are linked to each other via argument nodes, which record the inferences from one or more nodes to another. This notion of an argument does not have the usual recursive structure in systems for structured argumentation but instead stands for a single inference step. Unlike Pollock, Carneades does not express conflicts as a special type of link between statement nodes. Instead, inferences (i.e., arguments) can be either pro or con a statement. The evaluation of statements in an argument graph is, as with Pollock's inference graphs, defined in terms of the recursive structure of the graph. Statements are acceptable if they satisfy their 'proof standard'. The general framework abstracts from their nature but Gordon *et al.* give several examples of proof standards.

Inspired by Carneades, Brewka and Woltran [2010] proposed their *Abstract Dialectical Frameworks*, which are directed graphs in which nodes are arguments, statements or positions which can be accepted or not and the links represent dependencies between arguments. The dialectical status (accepted or rejected) of a node depends on the status of its parents as specified in an acceptance condition for the node. Brewka and Woltranpresent ADFs as generalisations of abstract argumentation frameworks. In a purely technical sense they are, but so are assumption-based argumentation, Deflog and $ASPIC^+$, which can all represent AFs as a special case. For example, in assumption-based argumentation arguments from the AF can be made assumptions and an assumption can be said to be a contrary of another assumption if it attacks it in the AF. So far, applications of ADFs have instead interpreted the nodes as statements, e.g. Strass [2013], thus making ADFs more similar to Pollock's inference graphs. Future research should shed more light on the potential of ADFs as generalisations of abstract argumentation frameworks in a conceptual sense also.

2.6.2 Decision making as argumentation

While most early work on argumentation-based inference was on epistemic reasoning (what is the case?), in recent years there has been much attention for practical reasoning (what should we do?). Among the first papers on this topic was Fox and Parsons [1997], motivated by medical decision making.

One strand of work was initiated by Grasso *et al.*'s [2001] design for a nutrition advice system and Bench-Capon's [2003] formal work on value-based argumentation frameworks. Both works were influenced by Perelman and Olbrechts-Tyteca's [1969] idea (further discussed in Section 3.1) that whether an argument in ordinary discourse is good does not depend on its logical form but on whether it is capable of persuading the addressed audience, which in turn depends on the extent to which it takes the audience's "values" into account. The work on value-based argumentation frameworks was further developed by e.g. Atkinson [2005] and Atkinson and Bench-Capon [2007], who instantiated value-based argumentation frameworks with an argumentation scheme approach inspired by Walton's [2006b] schemes for practical reasoning. This work has among other things been applied to legal interpretation [Atkinson *et al.*, 2005a], seen as a decision problem in which the various interpretation options promote or demote various legal or societal values.

Another strand of work is the work of Amgoud and others, e.g. Amgoud *et al.* [2005; 2009], Amgoud and Prade [2009], which combines argumentation models of the inferential aspects of argumentation with models of qualitative decision theory for the choice aspects of decision making.

To compare and contrast the various bodies of work, note that decision making has various aspects: identifying possible decision options in the form of possible actions, identifying the decision criteria (preferences, desires, goals, values), determining the consequences of actions with respect to these criteria and choosing between the decision options. There is consensus that all these aspects up to the choice problem can be modelled as argumentation as inference, but there is no consensus whether they can be modelled as instantiating the above-discussed general models of structured argumentation or whether special argumentation formats should be developed. Examples of the former approach are Kakas and Moraitis [2003], who model decision-making arguments in Dimopoulos and Kakas's [1995] logic-programming system for argumentation, van der Weide *et al.* [2011; 2011], who model practical-reasoning arguments in a combination of $ASPIC^+$ with Wooldridge *et al.*'s [2006] formal model of meta-argumentation, and Fan and Toni [2013], who model decision making in assumption-based argumentation. Examples of the latter approach are Amgoud [2009], who proposes specific formats for decision-making arguments, and Atkinson [2005] and Atkinson and Bench-Capon [2007], who define special argument-schemes for practical-reasoning.

Another issue is whether argumentation-based decision-making can be fully modelled as argumentation-based inference. Amgoud [2009], p. 318 claims that decision making goes beyond inference when a choice has to be made between the decision options; all argumentation can do according to her is generating the decision options that have justified epistemic subarguments or assumptions. Accordingly, Amgoud and Prade [2009] model the choice between epistemically justified decision options outside their argumentation model in terms of models of qualitative decision theory. By contrast, the above-mentioned work in *ASPIC*$^+$ and assumption-based argumentation tries to model choice through the general conflict-resolution mechanisms of argumentation-based inference, such as *ASPIC*$^+$'s argument ordering.

2.6.3 Argumentation combined with probability theory

A recent trend is the combination of argumentation-based inference with probability theory. This is not surprising, since argumentation has from the early days on been proposed as a model for reasoning under uncertainty. Yet systematic studies of the combination of argumentation with probability were sparse until recently.

Argumentation has been combined with probability theory for three different kinds of purposes. First, there has been some work in which probabilistic models are the object of argumentative discourse, such as Nielsen and Parsons [2007a], who model how Bayesian networks can be jointly constructed in an argumentation process. In all other work the uncertainty does not concern the probabilistic but the argumentation model. Two approaches can be distinguished, depending on whether the uncertainty is *in* or *about* the arguments. When the uncertainty is in the arguments, probabilities are *intrinsic* to an argument in that they are used for capturing the strength of an argument given uncertainty concerning the truth of its premises or the reliability of its inferences. An example is default reasoning with probabilistic generalisations, as in *The large majority of Belgian people speak French, Mathieu is Belgian, therefore (presumably) Mathieu speaks French.* Clearly, if all premises of an argument are certain and it only makes deductive inferences, the argument should be given maximum probabilistic strength. Hunter [2013] calls this use of probability the *epistemic* approach.

When the uncertainty is about the arguments, probabilities are *extrinsic* to an argument in that they are used for expressing uncertainty about whether arguments are accepted as existing by some arguing agent. Hunter [2014] gives the example of a dialogue participant who utters an enthymeme and where the listener can imagine two reasonable premises that the speaker had in mind: the listener can then assign probabilities to these options, which translate into probabilities on which argument the speaker meant to construct. This uncertainty has nothing to do with the intrinsic strengths of the two candidate completed arguments: one might be stronger than the other while yet the other is more likely the argument that the speaker had in mind. Hunter [2013] calls this use of probability the *constellations* approach. Note that in this approach even deductive arguments from certain premises can have less than maximal strength.

The intrinsic, or epistemic approach can be applied in two ways: by simply computing probability values of conclusions or by using such probabilities to resolve attacks into defeats. Computing probability values of conclusions is done in early work by Haenni *et al.* [2000]. Their argumentation model is a rather specific one for diagnosis and has no clear relations with more general structured and abstract models of argumentation. More recent work in this vein is Dung and Thang [2010], who within assumption-based argumentation allow rules to be labelled with probabilities. As noted above in Section 2.4.4, Pollock [2002; 2007a; 2010] (using a non-standard account of probability) made the justification status of statements a matter of numerical degree, being a function of the strengths of both supporting and defeating arguments.

Other examples of the intrinsic/epistemic approach are methods for extracting arguments from (qualitative or quantitative) Bayesian networks. Older work in this vein is Parsons [1998a; 1998b], using a logic similar to the one of Krause *et al.* [1995] and Williams and Williamson [2006], using the logic of Prakken and Sartor [1997]. In this work no probability values of conclusions are computed: Parsons just generates arguments while Williams & Williamson generate abstract argumentation frameworks using rebutting attack without preferences. Recent work is of Timmer *et al.* [2017], who generates arguments in $ASPIC^+$ and resolves their rebutting attacks with probabilistic strengths of arguments. The latter is also done in two pieces of work using alternatives for standard probability theory, viz. Pollock's [1994; 1995] use of degrees of belief and Chesñevar *et al.*'s [2004] use of possibilistic logic (both discussed above in Section 2.4.4).

The extrinsic/constellations approach has been largely applied to abstract argumentation frameworks, as in Li *et al.* [2012] and Hunter [2014] (but see the early work of Riveret *et al.* [2007; 2008] using the logic of Prakken and Sartor [1997]). For a recent overview see Hunter and Thimm [2016]. In this approach, probabilities can unlike in the intrinsic approach, also be attached to, for instance, legal rules or moral value judgements. Another difference is that the extrinsic use of probability is defined on top of a separate model of argumentation existing independently of the probabilistic model, while in the second use probability is part of the argumentation model itself.

Assigning probabilities to arguments in the abstract is problematic, since in probability theory probabilities are assigned to the truth of statements or to outcomes of events, and an argument is neither a statement nor an event. What is required here is a precise specification of what the probability of an argument means. If it corresponds to the degree of justification of the argument's justification, then this should arguably be specified at the level of structured argumentation. For a preliminary attempt to do so in the context of classical-logic argumentation see Hunter [2013]. If the probability of an argument corresponds to the probability of a statement about the argument, then the nature of that statement should be made clear. More generally, here too the need arises to be explicit about what is abstracted from, in this case in abstract models of probabilistic argumentation.

2.6.4 Argumentation dynamics

A development that is in the border area of inference and dialogue is the logical study of the dynamics of argumentation, insofar as it abstracts from agent aspects and the dialogical setting. For example, Coste-Marquis *et al.* [2007] study the merging of abstract argumentation frameworks, with attention for the resolution of conflicts between the merged frameworks. Also, much work has recently been done on the nature and effects of change operations on a given argumentation state [Modgil, 2006; Rotstein *et al.*, 2008; Baumann and Brewka, 2010; Baroni *et al.*, 2011b]. Among other things, enforcing and preservation properties are studied. Enforcement concerns the extent to which desirable outcomes can or will be obtained by changing an argumentation state, while preservation is about the extent to which the current status of arguments is preserved under change. Quite recently, the revision of argumentation frameworks has been studied analogously to revising belief sets or bases in belief revision, i.e. as incorporating new or deleting old elements while keeping the changes minimal [Coste-Marquis *et al.*, 2014].

Almost all current work on argumentation dynamics concerns abstract argumentation frameworks. In particular the following operations have been studied: addition or deletion of (sets of) arguments (e.g. [Baumann and Brewka, 2010; Cayrol *et al.*, 2010; Baumann, 2012b; Baumann, 2012a]) and addition or deletion of (sets of) attack relations (e.g. [Modgil, 2006; Boella *et al.*, 2010b; Baroni *et al.*, 2011b; Bisquert *et al.*, 2013]). Deleting attacks can be seen as an abstraction from the use of preferences to resolve attacks into defeats.

This current abstract work on argumentation dynamics abstracts from the structure of arguments and the nature of their conflicts, which is a significant limitation. See e.g. Modgil and Prakken [2012], who for this reason propose a model of preference dynamics in $ASPIC^+$. For example, abstract models of argumentation dynamics do not recognise that some arguments are not attackable (such as deductive arguments with certain premises) or that some attacks cannot be deleted (for example between arguments that were determined to be equally strong), or that the deletion of one argument implies the deletion of other arguments (when the deleted argument is a subargument of another, as in Figure 6 above), or that the deletion or addition of one attack implies the deletion or addition of other attacks (for example, attacking an argument implies that all arguments of which the attacked argument is a subargument are also attacked; in Figure 6 above attacking B_1 implies attacking B_2). All this means that formal results about the abstract model may only be relevant for specific cases and may fail to cover many realistic situations in argumentation. To give a very simple example, in models that allow the addition of arguments and attacks, any non-selfattacking argument can be made a member of every extension by simply adding non-attacked attackers of all its attackers. However, this result at the abstract level does not carry over to instantiations in which not all arguments are attackable. Here, too, the importance shows of being aware what the model abstracts from.

2.6.5 Other work

I end this section on argumentation-based inference with a very brief review of some other relevant work (without any hope of being complete).

Wooldridge *et al.* [2006] proposed a formalism for meta-argumentation, supporting a hierarchical formalisation of logic-based arguments. At each level of the hierarchy, arguments, statements and positions can refer to arguments, statements and positions at lower levels. This is achieved by using a hierarchical first-order meta-logic, a type of first-order logic in which individual terms in the logic can refer to terms in another language. One application of this formalism is van der Weide's [2011] model of reasoning about preferences in argumentation about decision making.

Finally, a recent trend is to develop gradual notions of argument acceptability in terms of structural properties of abstract argumentation frameworks [Cayrol and Lagasquie-Schiex, 2005a; Grossi and Modgil, 2015].

3 Formal and computational models of argumentation-based dialogue

So far we have discussed argumentation as a form of (nonmonotonic) inference. However, argumentation can also be seen as a form of dialogue, in which two or more agents aim to resolve a conflict of opinion by verbal means. When argumentation is viewed as a kind of dialogue between 'real' agents (whether human or artificial), new issues arise, namely, the *distributed* nature of information (over the agents), the *dynamic* nature of information, since agents do not reveal everything they believe from the start and since they can learn from each other, and *strategic* issues, since agents will have their internal preferences, desires and goals. At first sight, it might be thought that the argument games for argumentation-based semantics discussed above in Section 2.5.2 are dialogical models of argumentation. However, this is not the case, since they are not meant for discussions between real agents but as a proof theory. There is no dynamics, no distributed information and the notions of a proponent and an opponent are just proof-theoretic metaphors, not real agents with preferences, desires and goals.

Research on argumentation-based dialogue divides into research on communication languages and protocols (their formal definition and study of their properties) and research on agent behaviour in argumentation dialogues (strategies, tactics, heuristics). Some work studies the combination of a protocol and agent behaviour within that protocol. The main idea of work on argumentation protocols is that such protocols should promote fair and effective resolution of conflicts of opinion. In work on argumentative agent design the agents are assumed to adhere to this purpose of the dialogue but within the rules of the protocol they can pursue their own interests and objectives.

Research on argumentation-based dialogue is often done against the background of Walton's [1984] classification of dialogues into six types according to their goal (see also e.g. [Walton and Krabbe, 1995]). *Persuasion* aims to resolve a difference of opinion, *negotiation* tries to resolve a conflict of interest by reaching a deal, *information seeking* aims at transferring information, *deliberation* wants to reach a decision on a course of action, *inquiry* is aimed at "growth of knowledge and agreement" and *quarrel* is the verbal substitute of a fight. This classification is not meant to be exhaustive and leaves room for dialogues of mixed type. Persuasion can, given its purpose, be seen as 'pure' argumentation and is often embedded in other dialogue types in that dialogues of other types may shift to persuasion if a conflict of opinion arises. For example, in information-seeking a conflict of opinion could arise on the credibility of a source of information, in deliberation the participants may disagree about likely effects of plans or actions and in negotiation they may disagree about the reasons why a proposal is in one's interest; also, in all three cases the participants may disagree about relevant factual matters.

The formal study of argumentation-based dialogue is less substantial and less advanced than the formal study of argumentation-based inference. Unlike with inference, it largely consists of a variety of different approaches and individual systems, with few unifying accounts or general frameworks. For these reasons this section is shorter than Section 2 on argumentation-based inference.

3.1 Main historical influences

As noted in Section 2.1.3, Toulmin [1958] claimed that outside mathematics the validity of an argument does not depend on its syntactic form but on whether it can be defended in a properly conducted dispute. It might be argued that Toulmin thus anticipated argumentation-based inference, especially in argument-game form. However, more importantly, he thus planted the seed of an idea that later became prominent in informal logic and argumentation theory, namely, that arguments can only be evaluated in the context of a dialogue. Toulmin's call to logicians of his days to study the criteria for properly conducted disputes can be regarded as a call to study dialogue protocols for argumentation.

The Belgian philosopher Chaïm Perelman also emphasised the dialogical nature of argument evaluation. However, he did not address protocol but strategy, in arguing that arguments in ordinary discourse should not be evaluated in terms of their syntactic form but on their rhetorical potential to persuade an audience [Perelman and Olbrechts-Tyteca, 1969]. In particular, an argument is more persuasive the more it takes the audience's "values" into account. For example, an argument that governments should not tap internet communications of their citizens since this infringes on their privacy is not very persuasive to an audience that values security over privacy. While initially Perelman's work was only influential in informal logic and argumentation theory, it was around 2000 taken up by AI researchers in both inferential and dialogical models of argumentation about action selection, starting with Grasso *et al.*'s [2001] design of a nutrition advice system and Bench-Capon's [2003] formal work on value-based argumentation frameworks (the latter was discussed above in Section 2.6.2). Other work more generally aimed at characterising the persuasive force of arguments in terms of the similarity of an argument with the beliefs of a typical audience [Hunter, 2004].

While argumentation logics define notions of consequence from a given body of information, dialogue systems for argumentation regulate disputes between real agents. Systems for persuasion dialogues were already studied in medieval times [Angelelli, 1970]. The modern study of formal dialogue systems for persuasion probably started with two publications by Charles Hamblin [1970, 1971], who coined the term 'formal dialectic', and was also inspired by speech act theory in philosophy [Searle, 1969] and dialogue logic [Lorenzen and Lorenz, 1978]. It should be noted that formal systems for persuasion dialogue differ from dialogue logics in one crucial respect. Dialogue logic aims to define the semantics of logical operators in terms of rules of attack and defence. Accordingly the purpose of a dialogue is to determine whether a proposition is implied by a given set of propositions and the roles of proponent and opponent are just logical metaphors, just as in the logical argument games discussed above in Section 2.5.2. By contrast, the purpose of a persuasion dialogue is to resolve a conflict of opinion between real agents, who can ask for and provide substantive reasons for their claims.

Initially, formal systems for argumentation as dialogue were studied only within philosophical logic and argumentation theory; see, for example, [Mackenzie, 1979; Mackenzie, 1990; Woods and Walton, 1978; Walton and Krabbe, 1995]. From the early nineties the study of argumentation dialogues was taken up in several fields of computer science. In Artificial Intelligence logical models of commonsense reasoning have been extended with formal models of persuasion dialogue as a way to deal with resource-bounded reasoning [Loui, 1998; Brewka, 2001]. Persuasion dialogues have also been used in the design of intelligent tutoring systems [Moore, 1993; Yuan, 2004] and were proposed as an element of computer-supported collaborative argumentation [Maudet and Moore, 1999]. In AI & law formal dialogue systems for persuasion were developed as a model of procedural justice in the sense of e.g. Alexy [1978]. See, for example, [Gordon, 1994; Hage *et al.*, 1993; Bench-Capon, 1998; Bench-Capon *et al.*, 2000; Lodder, 1999; Prakken, 2001a; Prakken, 2008]. Finally, in the field of multi-agent systems dialogue systems have been incorporated into models of rational agent interaction based on the observation that many kinds of agent interaction (such as negotiation and group decision making) involve argumentation. Accordingly, interaction protocols for various dialogue types involving argumentation have been designed [Parsons and Jennings, 1996; Kraus *et al.*, 1998; Parsons *et al.*, 1998; Amgoud *et al.*, 2000a; McBurney and Parsons, 2002; Parsons *et al.*, 2002; Parsons *et al.*, 2003].

Most dialogue systems for argumentation are formulated in an informal mathematical metalanguage, but some have studied the full formalisation of protocols in logical action languages, such as Brewka [2001] in the situation calculus, Bodenstaff *et al.* [2006] in the event calculus and Artikis *et al.* [2007]

in C^{++}.

3.2 General remarks on dialogue systems for argumentation

Persuasion is usually modelled as a two-party dialogue between a proponent and an opponent of an initial claim. Essentially, dialogue systems define a *communication language* (the well-formed utterances) and a *protocol* (when a well-formed utterance may be made and when the dialogue terminates). The communication language consists of a set of locutions applied to statements or arguments expressed in a logical language according to some adopted monotonic or nonmonotonic logic. If this logic is nonmonotonic, it can but need not be an argumentation logic.

Dialogue systems define the principles of coherent dialogue. Carlson [1983] defined coherence in terms of the purpose of a dialogue. According to him, whereas logic defines the conditions under which a proposition is true, dialogue systems define the conditions under which an utterance is appropriate, and this is the case if the utterance furthers the purpose of the dialogue in which it is made. Thus according to Carlson the principles governing the meaning and use of utterances should not be defined at the level of individual speech acts but at the level of the dialogue in which the utterance is made. This justifies why most work on argumentation dialogues, like Carlson, takes a game-theoretic approach to dialogues, where speech acts are viewed as moves in a game and rules for their appropriateness are formulated as rules of the game. Loui [1998] distinguished between *effectiveness* and *fairness* of dialogue systems. Effectiveness means that the protocol furthers the purpose of the dialogue (in the case of persuasion that the conflict of opinion is resolved). Some aspects of effectiveness are efficiency (how long are dialogues and is there a guarantee of termination?) and relevance (is every move relevant to the dialogue topic?). Fairness means that the participants have a fair opportunity to argue their case. Some aspects of fairness are that the participants always have the opportunity to move relevant moves and that the outcome of a dialogue agrees with the parties' commitments.

Communication language Here are some common speech acts that can be found in the literature on persuasion dialogues, with their informal meaning and the various terms with which they have been denoted in the literature.

- *claim* φ (assert, statement, ...). The speaker asserts that φ is the case.
- *why* φ (challenge, deny, question, ...) The speaker challenges that φ is the case and asks for reasons why it would be the case.
- *concede* φ (accept, admit, ...). The speaker admits that φ is the case.
- *retract* φ (withdraw, no commitment, ..) The speaker declares that he is not committed (any more) to φ. Retractions are 'really' retractions if the speaker is committed to the retracted proposition, otherwise it is a mere declaration of non-commitment (e.g. in reply to a question).

- φ *since* S (argue, argument, ...) The speaker provides reasons why φ is the case. Some protocols do not have this move but require instead that reasons be provided by a *claim* φ or *claim* S move in reply to a *why* ψ move (where S is a set of propositions). Also, in some systems the reasons provided for φ can have structure, for example, of a proof three or a deduction.

- *question* φ (...) The speaker asks another participant's opinion on whether φ is the case.

Structural degrees of freedom Dialogue systems can vary in their structural properties in several ways [Loui, 1998]: whether players can reply just once to the other player's moves or may try alternative replies (*unique- vs. multi-reply protocols*); whether players can make just one or may make several moves before the turn shifts (*unique- vs. multi-move protocols*); and whether the turn shifts as soon as the player-to-move has made himself the winning side or may shift later (*immediate- vs. non-immediate-reply protocols*). According to Loui [1998], the desired degree of structural 'strictness' of a dialogue system depends on the context of a dialogue. In contexts with little time and resources a unique-move, unique- and immediate reply protocol may be best, to force the participants not to waste resources, while in other contexts with more time and resources it is better to allow the participants more freedom to explore alternatives and return to earlier choices.

Commitments An important notion in systems for argumentation dialogue is that of propositional *commitments* [Walton and Krabbe, 1995]. Commitments are an agent's publicly declared points of view about a proposition, which may or may not agree or coincide with the agent's internal beliefs. An example of where they often do not agree is criminal trial, where the accused may very well publicly defend his innocence while he knows he is guilty. Commitments are typically incurred by stating claims or arguments, while they are typically lost by retracting a claim or argument. Commitments can serve several purposes in dialogue systems. One role is in enforcing a participant's dialogical consistency, for instance, by requiring him to keep his commitments consistent at all times or to make them consistent upon demand, or to defend one's commitments when challenged or else give them up. Another role of commitments is to determine termination and outcome of a dialogue. For example, persuasion dialogues can be defined to terminate if the opponent is committed to the proponent's main claim or the proponent is not committed any more to the main claim.

3.3 Some work on systems for persuasion dialogue

Since persuasion is 'pure' argumentation, I now review some historically important work on systems for persuasion dialogue in more detail. Then I will more briefly review work that embeds argumentation in systems for other kinds of dialogues.

3.3.1 Mackenzie [1979]

Mackenzie's [1979] system has been historically influential especially for its set of locutions. His system has the *claim*, *why*, *concede* and *retract* locutions. The logical language is that of propositional logic but the logic is not full PL but instead a restricted notion of "immediate consequence", to capture resource-bounded reasoning (e.g. p, $p \rightarrow q$ and $q \rightarrow r$ immediately imply q but not r). Arguments are moved implicitly, by replying to a *why* move with a *claim*. An argument may be incomplete but its mover becomes committed to the material implication *premises* $\rightarrow$ *conclusion*. In addition, Mackenzie has a *question* speech act, which asks the hearer to declare a standpoint with respect to a proposition, and a *resolve* speech act for demanding resolution of conflicts in or logical implication by commitments. Mackenzie does not define outcomes or termination of dialogues. This makes his system underspecified as to the dialogue purpose, so that it can be extended to various types of dialogues. The protocol is unique-move and unique-reply but it nevertheless hardly enforces coherence of dialogues. Only the moves required after *why* and *question* and the use of the *resolve* move are constrained; the participants may freely exchange unrelated claims, and may freely challenge, retract or question. For instance, the following dialogue is legal:

P: *claim* p, O: *claim* q, P: *question* r, O: *claim* $\neg r$, P: *retract* s.

3.3.2 Walton & Krabbe [1995]

Walton and Krabbe [1995] developed the ideas of Mackenzie [1979] and also Woods and Walton [1978] into a full system for persuasion dialogues. To Mackenzie's locutions they added an explicit *since* locution for moving arguments. In their system, the only way to attack an argument is by challenging its premises, so the underlying logic is monotonic. The dialogues allowed by Walton and Krabbe are much more focused than Mackenzie's, since moves in a new turn must reply to a move in the previous turn of the other player. So, for instance, in the just-given example dialogue in Mackenzie's system, O's *claim* q move is not allowed and O must instead either concede or challenge p. This constraint also makes backtracking and postponement of replies impossible. Apart from this, the protocol allows that more than one move is made in one turn and alternative arguments for the same challenged proposition are moved. However, each move from the last turn must be replied-to (though other moves may be made as well).

Commitments are used by the protocol to enforce a participant's dialogical coherence. For example, if a participant's commitments logically imply an assertion of the other participant but do not contain that assertion, then the initial participant must either concede the assertion or retract one of the implying commitments.

The following example illustrates how the system deals with implicit premises:

P_1: *claim* this car is safe O_1: *why* is this car safe?; P_2: this car is safe *since* it has an airbag; P_2: safe *since* airbag.

Now the opponent must either challenge or concede both the explicit premise that the car has an airbag and the implicit premise 'if the car has an airbag, then it is safe'.

3.3.3 Gordon's Pleadings Game

Gordon's [1995] work on the *Pleadings Game*is seminal AI & Law work on the modelling of legal procedures as dialogue games. The game was intended as a normative model of civil pleading in Anglo-American legal systems, where the participants aim to identify the issues to be decided in court. The underlying logic is a nonmonotonic one, viz. conditional entailment [Geffner and Pearl, 1992], which as discussed above in Section 2.3 has a model-theoretic semantics and an argument-based proof theory. The game contains speech acts for conceding and challenging a claim, for stating and conceding arguments, and for challenging challenges of a claim. The latter has the effect of leaving the claim for trial. The Pleadings Game can be argued to have an implicit distinction between attacking and surrendering replies (as later made explicit in [Prakken, 2005]) in its distinction between three kinds of moves that have been made during a dialogue: the *open moves*, which have not yet been replied to, the *conceded moves*, which are the arguments and claims that have been conceded, and the *denied moves*, which are the claims and challenges that have been challenged and the arguments that have been attacked with counterarguments. The protocol is multi-move but unique-reply. At each turn a player must respond in some allowed way to every open move of the other player that is still 'relevant' (in a sense similar but not identical to that of Prakken [2005]), and may reply to any other open move. If no allowed move can be made, the turn shifts to the other player, except when this situation occurs at the beginning of a turn, in which case the game terminates. Move legality is further defined by specific rules for the various speech acts, which are mostly standard.

The result of a terminated game is twofold: a list of issues identified during the game (i.e., the claims on which the players disagree), and a winner, if there is one. Winning is defined relative to the background theory constructed during a game. If issues remain, there is no winner and the case must be decided by the court. If no issues remain, then the plaintiff wins iff his main claim is defeasibly implied by the final background theory, while the defendant wins otherwise.

3.3.4 Deriving locutions from argument schemes

The Toulmin Diagram Game (TDG) [Bench-Capon, 1998; Bench-Capon *et al.*, 2000] was intended to produce more natural dialogues than the "stilted" ones produced by systems such as those reviewed thus far. To this end, its speech acts are based on an adapted version of Toulmin's [1958] well-known argument scheme. In this scheme a *claim* is supported by *data*, which support is *warranted* by an inference license, which possibly has *presuppositions*, and which is *backed* by grounds for its acceptance; finally, a claim can be attacked with a *rebuttal*, which itself is a claim and thus the starting point of a counterargu-

Table 1. Attackers and surrenders in TDG

Locutions	Attacks	Surrenders
claim φ	*why* φ	*concede* φ
why φ	*supply data*$_\varphi$ ψ	*retract* φ
concede φ		
supply data$_\psi$ φ	*so*$_\psi$ φ *why* φ	*concede* φ
so$_\psi$ φ	*supply warrant* $\psi \Rightarrow \varphi$	
supply warrant w	*presupposing* w *on account of* w	*OK* w
presupposing w	*supply presupposition*$_w$ φ	*retract* w
on account of w	*supply backing*$_w$ b	*retract* w
supply backing$_w$ b		

ment. Arguments can be chained by regarding data also as claims, for which further data can be provided.

The locutions of TDG's communication language correspond to the elements of this scheme, as shown in Table 1. For ease of comparison, this table has an explicit reply structure as in [Prakken, 2005], to be discussed below, although the original TDG system leaves this structure implicit in its protocol.

The idea to generate natural dialogues by defining the communication language in terms of some argumentation scheme was later applied to practical reasoning by Atkinson *et al.* [2005b; 2006], who embedded Atkinson's [2005] argumentation scheme for practical reasoning in a dialogue system for persuasion over action.

3.4 Later formal work

All systems reviewed so far are either philosophically motivated or geared towards application domains, and none of them were formally investigated on their properties. This changed in later AI work on dialogue systems for argumentation, some of which I will now discuss.

3.4.1 Parsons, Wooldridge & Amgoud [2003]

Parsons *et al.* [2002; 2003] were among the first to undertake a systematic formal study of argumentation as dialogue. They proposed dialogue systems for various types of dialogues involving argumentation and formally investigated them on various kinds of properties. In all of them the underlying logic is nonmonotonic, namely, Amgoud and Cayrol's [2002] system for classical-logic argumentation with grounded semantics. In this section I discuss their system for persuasion dialogues. Its communication language consists of claims, challenges, concessions and questions. Arguments are moved implicitly as *claim* replies to *why* moves (where sets of propositions may be claimed). The protocol has a rigid, unique-move and unique-reply nature, except that each premise of an argument may be responded to in turn. Unlike the above work, Parsons *et al.* make several assumptions on agent behaviour. Participants have their own, possibly inconsistent belief base and they are assumed to adopt an assertion and acceptance attitude, which they must respect throughout the dialogue. Moreover, claims moved in support of other claims must be from the participant's internal belief base.

Parsons *et al.* distinguish the following assertion attitudes: a *confident* agent can assert any proposition for which he can construct an argument, a *careful* agent can do so only if he can construct such an argument and cannot construct a stronger counterargument and a *thoughtful* agent can do so only if he can construct an acceptable argument for the proposition (according to grounded semantics). The corresponding acceptance attitudes also exist: a *credulous* agent concedes a proposition if he can construct an argument for it, a *cautious* agent does so only if in addition he cannot construct a stronger counterargument and a *skeptical* agent does so only if he can construct an acceptable argument for the proposition. In verifying these attitudes, each player must reason with its own beliefs and the commitments of the other side.

Consider the following example, where the proponent P believes p and $p \rightarrow q$, the opponent believes r and $r \rightarrow \neg q$ and all formulas are of equal preference. If P starts with *claim* q, then O must, depending on its dialogical attitudes, concede q if possible, otherwise claim $\neg q$ if possible, otherwise challenge q. If O is credulous or cautious, then perhaps surprisingly she must concede, since she has to reason with P's commitment p so she can construct a trivial argument for q, namely, $\{q\} \vdash q$. In both cases the dialogue terminates with agreement. By contrast, if O is skeptical, she has to challenge q. Then P has to move *claim* $\{p, p \rightarrow q\}$. Then O, being skeptical, must challenge both p and $p \rightarrow q$. The proponent then has to reply with *claim* $\{p\}$ and *claim* $\{p \rightarrow q\}$, after which the dialogue terminates without agreement, because the players are not allowed to repeat their moves, while O's acceptance attitude tells her to repeat her last two challenges.

Parsons *et al.* [2002; 2003] were among the first to undertake a systematic formal study of argumentation as dialogue. They proposed dialogue systems for various types of dialogues involving argumentation and formally investigated them on various kinds of properties. In all of them the underlying logic is nonmonotonic, namely, Amgoud and Cayrol's [2002] system for classical-logic argumentation with grounded semantics. In this section I discuss their system for persuasion dialogues. Its communication language consists of claims, challenges, concessions and questions. Arguments are moved implicitly as *claim* replies to *why* moves (where sets of propositions may be claimed). The protocol has a rigid, unique-move and unique-reply nature, except that each premise of an argument may be responded to in turn. Unlike the above work, Parsons *et al.* make several assumptions on agent behaviour. Participants have their own, possibly inconsistent belief base and they are assumed to adopt an assertion and acceptance attitude, which they must respect throughout the dialogue. Moreover, claims moved in support of other claims must be from the

Table 2. An example communication language in Prakken's framework

Locutions	Attacks	Surrenders
claim φ	*why* φ	*concede* φ
φ *since* S	*why* $\psi (\psi \in S)$	*concede* ψ $(\psi \in S)$
	φ' *since* S' (φ' *since* S' defeats φ *since* S)	*concede* φ
why φ	φ *since* S	*retract* φ
concede φ		
retract φ		

participant's internal belief base.

Parsons *et al.* investigate various properties of the protocols and their outcomes. Some results are on whether termination of dialogues is guaranteed. Other results are on the computational complexity of the various aspects. Yet other results concern possible agent behaviours. For example, they studied the extent to which one agent can mislead the other agent by making her concede a proposition he himself does not believe. They thus were among the first to address issues of trust in argumentation dialogue.

A very interesting aspect of this work is the definition of the various dialogical attitudes. However, these notions are perhaps better seen as aspects of strategy than of protocol, since if they are referred to by the protocol, an outside observer cannot verify protocol compliance, which is often regarded as a drawback of communication protocols.

3.4.2 Prakken [2005]

In [Prakken, 2005] a framework for specifying two-party persuasion dialogues is presented, which is then instantiated with some example protocols. The aim of this work was to allow a more general study of properties of dialogue systems of argumentation than the work reviewed so far. To this end, the framework largely abstracts from the logical language, the logic and the communication language, except that the communication language has to have an explicit reply structure and that underlying logic is assumed to be a system that is much like a preliminary version of $ASPIC^+$. Moreover, different protocols were defined, all extending a partial core protocol.

A main motivation of the framework was to ensure focus of dialogues while yet allowing for freedom to move alternative replies and to postpone replies. This was achieved with two main features of the framework. Firstly, an explicit reply structure on the communication language is assumed (implicit in several other systems), where each move either *attacks* or *surrenders to* its target. An example $\mathcal{L}_c$ of this format is displayed in Table 2. Secondly, winning is defined for each dialogue, whether terminated or not, and it is defined in terms of a notion of *dialogical status* of moves. The *dialogical status* of a move is recursively defined as follows, exploiting the tree structure of dialogues. A move is *in* if it is surrendered or else if all its attacking replies are *out*. This implies that a move without replies is *in*. And a move is *out* if it has a reply that is *in*. Actually, this has to be refined to allow that some premises of an argument are conceded while others are challenged; see [Prakken, 2005] for the details. Then a dialogue is (currently) won by the proponent if its initial move is *in* while it is (currently) won by the opponent otherwise.

Together, these two features of the framework allow for a notion of relevance that ensures focus while yet leaving the desired degree of freedom (generalised from [Prakken, 2001b]): a move is *relevant* just in case making its target *out* would make the speaker the current winner. Termination is defined as the situation that a player is to move but has no legal moves. The players can also agree to terminate a dialogue.

Consider by way of example the following dialogue in a protocol that allows replies to all moves of the other player but only if the move is relevant.

P_1:	*claim p*	
O_1:	*why p*	(replying to P_1)
P_2:	*p since q*	(replying to O_1)
O_2:	*why q*	(replying to P_2)
P_3:	*p since r*	(replying to O_1)

At this point a reply to P_2 is irrelevant, since P_2 is *out*, so replying to it cannot change the status of P_1. Note that the dialogue can only terminate after either P has replied to O_1 with *retract p* or O has replied to P_1 with *concede p*. In all other cases, legal moves can always be made.

3.4.3 Argument games as dialogue systems

Argument games for abstract argumentation semantics were above in Section 2.5.2 discussed as a proof theory for abstract argumentation semantics. However, they have also been studied as genuine dialogue games for disagreeing agents, by dropping the assumption that all arguments are taken from a fixed and globally known argumentation framework [Loui, 1998; Jakobovits and Vermeir, 1999a; Jakobovits, 2000; Prakken, 2001b]. If this assumption is dropped, the properties of the game can change. A positive change is proven by Jakobovits, viz. that certain dynamic argument-game protocols prevent the construction of AFs containing odd loops (it is well known that such theories may have no extensions). A negative result is proven in [Prakken, 2001b], viz. that the dynamified game for grounded semantics loses soundness with respect to the joint framework constructed during a dialogue. However, if the game is changed by allowing any relevant reply (in the sense of [Prakken, 2005]) to any earlier move of the other side, then soundness is restored.

While the study of argument games as dialogue systems is theoretically very interesting, their very simple logic and communication language make that they cannot be a realistic model of persuasion dialogue.

3.5 Persuasion embedded in other types of dialogues

I now briefly review work that embeds argumentation in a dialogue system for other types of dialogues.

3.5.1 Negotiation

Much work on embedding argumentation into negotiation protocols is motivated by the claim that argumentation can be beneficial to negotiation. From the point of view of the negotiating agents, adding reasons for a proposal could increase the chance of acceptance. This was the idea of Sycara's [1985; 1990] early work on modelling threats and reward in labour negotiation. For example 'if you do not accept our offer, we will go on strike' (a threat) or 'if you accept that you have to work during the weekends, you will receive an increase in salary' (a reward). This idea was generalised by Parsons *et al.* [1998] and Kraus *et al.* [1998] for BDI-style agents, that is, agents that form their intentions to act according to their beliefs and, possibly prioritised, desires [Rao and Georgeff, 1991]. The general idea is that the other agent should be made to change its beliefs or preferences in such a way that it will form the intention to accept or make an offer that the initial agent wants.

From the perspective of protocol design the idea is that if negotiating agents exchange and discuss reasons for their proposals and rejections, the negotiation process may become more efficient and the negotiation outcome may be of higher quality. If an agent explains why he rejects a proposal, the other agent knows which of her future proposals will certainly be rejected so she will not waste effort at such proposals. Thus efficiency is promoted. In such exchanges, reasons are not only exchanged, they can also become the subject of debate. Suppose a car seller offers a Peugeot to the customer but the customer rejects the offer on the grounds that French cars are not safe enough. The car seller might then try to persuade the customer that he is mistaken about the safety of French cars. If she succeeds in persuading the customer that he was wrong, she can still offer her Peugeot. Thus the quality of the negotiation is promoted, since the buyer has revised his preferences to bring them in agreement with reality. This example illustrates that a negotiation dialogue (where the aim is to reach a deal) sometimes contains an embedded persuasion dialogue (where the aim is to resolve a conflict of opinion).

Since all this is about giving reasons for or against acting in a certain way, the kind of argumentation that is involved is, inferentially speaking, argumentation about decision options (see Section 2.6.2 above), although it can, as the car sales example shows, also shift to epistemic argumentation about the underlying facts. The early work of Sycara [1985; 1990] and Kraus *et al.* [1998] applied informal rhetorical models of argumentation. Later work incorporated formal inferential models of argumentation in negotiation protocols. For example, Parsons *et al.* [1998] embed Krause *et al.*'s [1995] logic of argumentation, Amgoud *et al.* [2000b] embed Amgoud and Cayrol's [1998] model of classical argumentation with preferences, Amgoud and Prade [2004] incorporate the model summarised by Amgoud [2009], and van Veenen and Prakken

[2006] combine Wooldridge and Parson's [2000] negotiation protocol with one of Prakken's [2005] persuasion protocols, thus also embedding its preliminary version of $ASPIC^+$.

3.5.2 Deliberation

The purpose of deliberation is to agree on a course of action. It differs from persuasion over action, as modelled by e.g. Atkinson *et al.* [2005b; 2006] in that at the start of a deliberation dialogue there typically just is a problem and no proposed solutions yet. It differs from negotiation in that deliberating agents are assumed not to be self-interested but collaborative, sharing the goals of the group or community they are part of. The group may be small, such as a few people choosing a restaurant for dinner, it may be big, such as in parliamentary debate, and it may be huge, such as in public debate about political or societal issues. Clearly, different settings require different kinds of protocols.

Embedding argumentation in deliberation has much the same benefits as embedding it in negotiation: for the agents it may increase the chance of acceptance of their proposals, and for the dialogue it may increase the quality of the outcome. Research on deliberation with argumentation started later than research on argumentation-based negotiation and is not as extensive. Here is brief overview of some work.

Tang and Parsons [2005] proposed a rather specific dialogue system for argumentation about means-end planning, not based on a formal model of argumentation.

McBurney *et al.* [2007] proposed a framework for multi-agent deliberation dialogues. The protocol is intended to allow for the open nature of deliberation, giving the agents much freedom for establishing goals, constraints, perspectives, facts, actions and evaluations. Accordingly, the dialogue cyclicly moves through various stages. After initial inform and propose stages, the agents evaluate and decide on actions in the consideration, revision, recommendation and confirmation stages. The framework does not assume a specific argumentation logic.

Black and Atkinson *et al.* [2011] proposed a much more rigid system for two-agent deliberation based on Atkinson's [2005] embedding of an argument scheme for practical reasoning in value-based argumentation frameworks. The rigidness of the system allows them to show that if the agents adhere to the dialogue protocol and construct their arguments on the basis of their own belief bases, then any agreed proposal is also acceptable to both agents individually.

Finally, Kok *et al.* [2011] proposed a dialogue system for multi-agent deliberation dialogues as part of an experimental setup for testing the usefulness of argumentation in such dialogues. The system is an instance of Prakken's [2005] framework for persuasion but adapted to deliberation. It incorporates $ASPIC^+$ as the underlying logic.

3.5.3 Inquiry

Only little work has been done on embedding argumentation in inquiry. Early work is McBurney and Parsons' [2001] model of scientific inquiry. More recently, Black and Hunter [2007; 2009] embedded Garcia and Simari's [2004] DeLP argumentation system in a protocol for inquiry dialogue. They combined the protocol with a strategy that selects exactly one of the legal moves to make. This allowed them to prove soundness and completeness properties with respect to the participants' belief bases, provided the agents construct their arguments from their own belief base.

3.6 Work on strategic aspects of argumentation

Dialogue systems for argumentation only cover the rules of the game, i.e., which moves are allowed; they do not cover principles for playing the game well, i.e., strategies, tactics and heuristics for the individual players. Above we already discussed some work that studies the combination of a protocol with strategies, such as [Black and Atkinson, 2011] for deliberation and [Black and Hunter, 2007; Black and Hunter, 2009] for inquiry. Moreover, as remarked above, the assertion and acceptance policies studied by Parsons *et al.* [2002; 2003] could be seen as heuristics for move selection (although Parsons *et al.* make them part of their protocols).

Other early work on strategic aspects of argumentation is of Amgoud and Maudet [2002], who, building on the even earlier work of Moore [1993] on argumentation dialogues for intelligent tutoring, formulated move selection strategies and tactics based on human strategies in natural dialogues. One example is that agents have to choose between a *build* or *destroy* attitude, i.e., whether they want to support their own or to attack their opponent's position. This idea was later also used by Kok [2013] in his simulation experiments on whether argumentation is beneficial to deliberating agents.

In the context of dialogue games for abstract argumentation, Paul Dunne studied issues arising from the mismatch between the purpose of persuasion dialogues and the arguing agent's own objectives. In [Dunne, 2003] he studied the use of delay tactics and in [Dunne, 2006] he studied situations where agents have a 'hidden agenda'.

More recently, there is an emerging research strand on opponent modelling for strategic purposes, for example in terms of probability distributions or expected-utility distributions over the possible actions of the opponent [Matt and Toni, 2008; Thimm and Garcia, 2010; Oren and Norman, 2010; Hadjinikolis *et al.*, 2013; Rienstra *et al.*, 2013]. Somewhat earlier, Riveret *et al.* [2008] probabilistically modelled not an opponent but an impartial adjudicator who has the power to accept or reject premises of arguments put forward by the adversaries. In this work, probabilistic game theory can be used to determine optimal strategies.

Other recent work that uses game theory is that on mechanism design for argumentation [Rahwan and Larson, 2008; Rahwan *et al.*, 2009]. The goal here is to develop protocols that make unwanted behaviour (such as lying or withholding information) suboptimal.

All this recent work on strategic aspects of argumentation is still preliminary and therefore I leave a further description to other chapters in this handbook. Until these chapters are available, the reader can consult [Thimm, 2014] for a recent overview. I confine myself to one concluding observation. On the one hand, the recent work on strategy, heuristics and tactics is a natural continuation of the earlier work on communication languages and protocols. However, in one respect it is a step backwards, since it generally assumes much simpler dialogue systems than were developed before, with, for example, much recent work assuming simple dialogue games for abstract argumentation semantics.

4 Application areas

Formal and computational models of argumentation have been applied in several areas. Although a comprehensive review is beyond the scope of this chapter, a brief overview is in order. For more detailed overviews the reader can consult [Modgil *et al.*, 2013] and some references given below. I will mainly focus on three main application areas, viz. medicine, law and debating technologies. In addition, in the literature many specific applications can be found, such as to recommender systems, trust and reputation management, robot soccer, waste management, licensing policy management, the internet of things, and so on.

Below I will only discuss applications of formal models of argumentation. In several areas there is much applied research based on informal or ad-hoc models of argumentation. For example, argumentation has been used in work on risk assessment and design rationale in software engineering for explaining why a design meets a design requirement or avoids a risk [Haley *et al.*, 2008; Franqueira *et al.*, 2011] . Moreover, there is quite some work on support tools for argument visualisation [Reed *et al.*, 2007; ter Berg *et al.*, 2009] and collaborative argumentation and decision making [Conklin *et al.*, 2001; Scheuer *et al.*, 2010; Kirschner *et al.*, 2003], sometimes in educational contexts [Pinkwart and McLaren, 2012], and with applications for the social web [Schneider *et al.*, 2013]. Finally, recently research in argument mining [Palau and Moens, 2009; Lippi and Torroni, 2016] has become popular, which aims to recognise (elements of) arguments and their relations in natural-language texts.

As for the nature of the applications mentioned below, theoretical, user-oriented and fielded applications can be distinguished. *Theoretical applications* use a non-trivial domain example to demonstrate the adequacy or motivate design features of the model. In *user-oriented applications* (which usually are of computational architectures) the usefulness of the architecture for designated types of users or tasks is an essential aspect. *Fielded applications* have actually been used by the intended user group in a realistic context, either experimentally or in actual use.

4.1 Medical applications

Medicine has been an important application field of argumentation, with John Fox as a historically influential figure. Several systems developed by him and his colleagues have been experimentally tested or are even in actual use [Fox *et al.*, 2007], so these count as fielded applications. While their underlying argumentation model is rather simple, this group also studied formal foundations of their systems, e.g. in [Elvang-Göransson *et al.*, 1993] and [Krause *et al.*, 1995]. Moreover, Fox and Parsons [1997] proposed one of the first formal argumentation-based models of decision making, using arguments for expressing and comparing the positive and negative effects of medical treatments. This idea was combined with an argument-scheme approach by Tolchinsky *et al.* [2006; 2012], who present a model for multi-agent deliberation about safety-critical medical actions, such as donor organ selection for patients. The intended system plays the role of a mediating agent whose task is to inform the participants about their valid move options, to decide whether an argument is relevant enough to be admitted into the process, and to evaluate the admitted arguments in order to assess whether the proposed action should be undertaken. Since this system was tested experimentally with medical doctors, it counts as a fielded application.

More recently, Hunter and Williams [2012] have applied argumentation in a user-oriented way to the problem of aggregating evidence-based arguments for and against treatment options from clinical trials. They use preference-based abstract argumentation frameworks instantiated with one-steps applications of domain-specific inference rules, and express argument preferences in terms of outcome indicators of the treatments. The approach was evaluated by comparison with recommendations made in published healthcare guidelines.

4.2 Legal applications

There has been much cross-pollination with the field of AI & Law [Prakken and Sartor, 2015]. This is understandable, given the inherently adversarial nature of the law and the importance of written justifications of legal decisions. Rule-based argumentation formalisms such as assumption-based argumentation and the system of Prakken and Sartor [1997] have been applied to preference-based reasoning with conflicting rules [Kowalski and Toni, 1996; Prakken and Sartor, 1996]. Prakken & Sartor [2009] also used their logic to formalise notions of burdens of proof, as was done by Gordon and Walton [2009] with their Carneades system. Work on applying dialogue systems to the formalisation of legal procedure was discussed above in Sections 3.1 and 3.3.3.

An important contribution of AI & Law to the formal study of argumentation is the study of the role of cases in argumentation; for a recent detailed overview see [Bench-Capon, 2017]. In Section 1 above the still influential HYPO system [Ashley, 1990] and its successor CATO [Aleven, 2003] were mentioned. Their underlying argumentation model is for 'factor'- or 'dimension'-based reasoning, where cases are collections of abstract fact patterns that favour or oppose a conclusion, either in an all-or nothing fashion (factors) or to varying degrees (dimensions). This work inspired subsequent formal work using the tools of formal argumentation, e.g. [Hage, 1993; Loui *et al.*, 1993; Prakken and Sartor, 1998; Bench-Capon and Sartor, 2003]. A key idea in this work is that case decisions give rise to conflicting rules (or conflicting sets of reasons) plus a preference expressing how the court resolved this conflict. In the notation of Prakken and Sartor [1998]:

r_1: *Pro-factors* $\Rightarrow$ *Decision*
r_2: *Con-factors* $\Rightarrow$ *Not D*ecision
$r_1 > r_2$

The rule preference expresses the court's decision that the pro factors in the body of rule r_1 together outweigh the con factors in the body of rule r_2. This approach allows for 'a fortiori' reasoning in that adding factors to a pro-decision rule or removing factors from a con-decision rule does not affect the rule priority. Horty [2011], using a non-argumentation-based nonmonotonic logic, formalises the conditions under which a decision is allowed or forced by body of precedents and then uses this to also formalise the concepts of following, distinguishing and overruling a precedent.

A related line of research is to compare cases not in terms of their factors but in terms of underlying legal and social values. Berman and Hafner [1993] argued that often a factor can be said to favour a decision by virtue of the purposes served or values promoted by taking that decision because of the factor. A choice in case of conflicting factors is then explained in terms of a preference ordering on the purposes, or values, promoted or demoted by the decisions suggested by the factors. Cases can then be compared in terms of the values at stake rather than on the factors they contain. Bench-Capon [2002] first computationally modelled this approach, leading to a series of papers culminating in [Prakken *et al.*, 2015] and using argument schemes for practical reasoning of the kinds also used in argumentation-based models of decision making (see Section 2.6.2 above).

All the AI & law applications mentioned so far are theoretical applications. User-oriented legal applications of argumentation are rare, with most applications in the field of e-democracy, e.g. [Cartwright and Atkinson, 2009; Gordon, 2011]. Finally, to my knowledge only one fielded application exists, namely, the CATO system, which was experimentally tested for teaching case-based argumentation skills to American law students.

4.3 Debating technologies

Most work on debating technologies is based on informal or ad-hoc models of argumentation; for overviews see the references given above. An exception is the work of the Arg-tech group at the University of Dundee, Scotland, led by Chris Reed. This group has developed various user-oriented web-based argumentation tools partly based on formal foundations [Bex *et al.*, 2013a].

For example, they have been using the so-called Argument Interchange Format [Chesñevar *et al.*, 2006], which was given a logical foundation in $ASPIC^+$ by Bex *et al.* [2013b] and they have an online implementation of an instance of $ASPIC^+$ called TOAST [Snaid and Reed, 2012]. Several tools developed by the Arg-tech group have been experimentally tested with intended users, so these count as fielded applications.

5 Conclusion

Looking back on the history of formal research on argumentation, there is a marked difference between the study of argumentation as inference and that of argumentation as dialogue. The theory of argumentation-based inference is mature, with an almost universally accepted formal foundation in Dung's theory of abstract argumentation frameworks and its extensions and with a converging study of structured argumentation, with just a small number of general frameworks and increasing knowledge about their relations. By contrast, the study of argumentation-based dialogue consists of a variety of different approaches and individual systems, all exciting work but with few unifying accounts or general frameworks. There are a few exceptions, such as a series of papers just after 2000 by Peter McBurney, Simon Parsons and others on principles for the design of dialogue systems e.g. [McBurney *et al.*, 2002; McBurney and Parsons, 2002], and my own formal framework for persuasion dialogue in [Prakken, 2005]. However, this work is still far from being foundational.

In my own personal opinion, the following are the four main main theoretical contributions of the field.

1. The idea that dialectical evaluation of arguments can be formalised. While logic textbooks routinely write that a valid argument does not dictate the acceptance of its conclusion since it can always be attacked on its premises, formal argumentation has shown that attack relations between arguments conform to patterns that can be formally studied. In its purest form this is captured in Dung's [1995] theory of abstract argumentation frameworks.

2. The idea of defeasible rules. Dogma has it that all arguments should be deductively valid, that is, the truth of their premises should guarantee the truth of their conclusion. The fields of informal logic, argumentation theory and epistemology have questioned this dogma and argued that arguments that fail to meet this standard of perfection can still be good, as long as they withstand critical scrutiny. The field of formal argumentation has shown that this idea can be formalised.

3. The idea that the principles for evaluating arguments in the context of a dialogue can be formalised. Toulmin [1958] first proposed that arguments should be evaluated not on their syntactic form but on whether they can be defended in a properly conducted dispute. He urged logicians of his day

to study the principles of proper dispute. The formal study of dialogue studies has met this challenge and thus also opened the prospects for precise formal studies of strategy and tactics for persuasion.

4. The idea that reasoning under uncertainty can be formalised in a qualitative way. There is an increasing trend of advocating quantitative (especially Bayesian) models of uncertainty as the only way to reason about uncertainty. Likewise with quantitative models of decision making. However, for humans such quantitative theories are often hard to grasp, while they largely ignore the dialogical and procedural aspects of reasoning. This is especially a problem for applications with humans in the loop, such as support tools for human argumentation and decision making. Our field has shown that a natural qualitative theory of reasoning under uncertainty can be formalised.

However, there is, in my opinion, also an unfortunate recent development. While Dung's [1995] idea of abstract argumentation frameworks was a major breakthrough and is deservedly a key element in the formal study of argumentation-based inference, not all follow-up work is of the same generality. We have seen that several proposals for extending abstract argumentation frameworks with new elements implicitly make assumptions that are not in general satisfied. The same holds for work on the dynamics of abstract argumentation and for some work on probabilistic abstract argumentation. The resulting formalisms are thus abstract but not general in that they model special cases, such as the case in which all arguments, or all attacks, are independent of each other, or the special case in which all arguments are attackable.

It is worth noting that the word 'abstract' in Dung's [1995] notion of abstract argumentation frameworks does not qualify 'argumentation' but 'frameworks'. In Dung's terminology, it is the framework that is abstract, not the argumentation. Strictly speaking there is no such thing as abstract argumentation, just as there is no such thing as structured argumentation. All there is is argumentation, which can be studied at various levels of abstraction. And in real argumentation not arguments but things like claims, reasons and grounds are the most basic elements. There is nothing wrong in principle with abstract studies of argumentation: abstraction is an indispensable tool in any kind of research. However, one should not forget that we all study the same phenomenon, so that the various levels of abstraction should be connected. I remind the reader of my (perhaps controversial) proposal in Section 2.5.3 of a methodological guideline that every new proposal for extending abstract argumentation frameworks with new elements should in the same paper be accompanied by at least one non-trivial instantiation, in order to demonstrate the significance of the new extension. In doing so, we would respect the historic roots of the abstract study of argumentation, since in his original 1995 paper Dung respected this guideline in a way that has since never been equalled.

It is time to conclude. The formal and computational study of argumentation has established itself as a mature field of research. Argumentation is a key word or topic in all main AI conferences, papers on argumentation are published in the major AI journals, and the field has its own COMMA conference plus several workshops (CNMA, ArgMas, TAFA). Theoretically, the field is in a healthy state with much exciting research. With respect to applications this is less so, but this holds for all theoretically interesting fields of research. There is every hope to be optimistic here too, as long as a too strong focus on abstract argumentation is avoided. Unlike, for example, constraint satisfaction or model checking, argumentation is not just a technique but an important aspect of human life. There will therefore always be the need for support tools for argumentation, and our field is arguably in an excellent position to provide these tools. In any case, it provides their formal foundations.

BIBLIOGRAPHY

[Aleven and Ashley, 1991] V. Aleven and K.D. Ashley. Toward an intelligent tutoring system for teaching law students to argue with cases. In *Proceedings of the Third International Conference on Artificial Intelligence and Law*, pages 42–52, New York, 1991. ACM Press.

[Aleven, 2003] V. Aleven. Using background knowledge in case-based legal reasoning: a computational model and an intelligent learning environment. *Artificial Intelligence*, 150:183–237, 2003.

[Alexy, 1978] R. Alexy. *Theorie der juristischen Argumentation. Die Theorie des rationalen Diskurses als eine Theorie der juristischen Begründung.* Suhrkamp Verlag, Frankfurt am Main, 1978.

[Amgoud and Besnard, 2010] L. Amgoud and Ph. Besnard. A formal analysis of logic-based argumentation systems. In *Proceedings of the 4th International Conference on Scalable Uncertainty Management (SUM'10)*, number 6379 in Springer Lecture Notes in AI, pages 42–55, Berlin, 2010. Springer Verlag.

[Amgoud and Besnard, 2013] L. Amgoud and Ph. Besnard. Logical limits of abstract argumentation frameworks. *Journal of Applied Non-classical Logics*, 23:229–267, 2013.

[Amgoud and Cayrol, 1998] L. Amgoud and C. Cayrol. On the acceptability of arguments in preference-based argumentation. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 1–7, 1998.

[Amgoud and Cayrol, 2002] L. Amgoud and C. Cayrol. A model of reasoning based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34:197–215, 2002.

[Amgoud and Maudet, 2002] L. Amgoud and N. Maudet. Strategical considerations for argumentative agents (preliminary report). In *Proceedings of the Ninth International Workshop on Nonmonotonic Reasoning*, pages 399–407, Toulouse, France, 2002.

[Amgoud and Prade, 2004] L. Amgoud and H. Prade. Reaching agreement through argumentation: A possibilistic approach. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Ninth International Conference*, pages 175–182. AAAI Press, 2004.

[Amgoud and Prade, 2009] L. Amgoud and H. Prade. Using arguments for making and explaining decisions. *Artificial Intelligence*, 173:413–436, 2009.

[Amgoud and Vesic, 2010] L. Amgoud and S. Vesic. Handling inconsistency with preference-based argumentation. In *Proceedings of the 4th International Conference on Scalable Uncertainty Management (SUM'10)*, number 6379 in Springer Lecture Notes in AI, pages 56–69, Berlin, 2010. Springer Verlag.

[Amgoud *et al.*, 2000a] L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In *Proceedings of the Fourth International Conference on MultiAgent Systems*, pages 31–38, 2000.

[Amgoud *et al.*, 2000b] L. Amgoud, S. Parsons, and N. Maudet. Arguments, dialogue, and negotiation. In *Proceedings of the Fourteenth European Conference on Artificial Intelligence*, pages 338–342, 2000.

[Amgoud *et al.*, 2005] L. Amgoud, J.-F. Bonnefon, and H. Prade. An argumentation-based approach to multiple criteria decision. In *Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 05)*, number 3571 in Springer Lecture Notes in AI, pages 269–280, Berlin, 2005. Springer Verlag.

[Amgoud, 2009] L. Amgoud. Argumentation for decision making. In I. Rahwan and G.R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 301–320. Springer, Berlin, 2009.

[Angelelli, 1970] I. Angelelli. The techniques of disputation in the history of logic. *The Journal of Philosophy*, 67:800–815, 1970.

[Artikis *et al.*, 2007] A. Artikis, M.J. Sergot, and J. Pitt. An executable specification of a formal argumentation protocol. *Artificial Intelligence*, 171:776–804, 2007.

[Ashley, 1990] K.D. Ashley. *Modeling Legal Argument: Reasoning with Cases and Hypotheticals.* MIT Press, Cambridge, MA, 1990.

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

[Atkinson *et al.*, 2005a] K. Atkinson, T.J.M. Bench-Capon, and P. McBurney. Arguing about cases as practical reasoning. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law*, pages 35–44, New York, 2005. ACM Press.

[Atkinson *et al.*, 2005b] K. Atkinson, T.J.M. Bench-Capon, and P. McBurney. A dialogue game protocol for multi-agent argument over proposals for action. *Journal of Autonomous Agents and Multi-Agent Systems*, 11:153–171, 2005.

[Atkinson *et al.*, 2006] K. Atkinson, T.J.M. Bench-Capon, and P. McBurney. Computational representation of persuasive argument. *Synthese*, 152:157–206, 2006.

[Atkinson, 2005] K. Atkinson. *What Should We Do?: Computational Representation of Persuasive Argument in Practical Reasoning.* PhD Thesis, Department of Computer Science, University of Liverpool, Liverpool, UK, 2005.

[Baker and Ginsberg, 1989] A.B. Baker and M.L. Ginsberg. A theorem prover for prioritized circumscription. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 463–467, 1989.

[Baroni and Giacomin, 2007] P. Baroni and M. Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171:675–700, 2007.

[Baroni *et al.*, 2011a] P. Baroni, M. Caminada, and M. Giacomin. An introduction to argumentation semantics. *The Knowledge Engineering Review*, 26:365–410, 2011.

[Baroni *et al.*, 2011b] P. Baroni, P.E. Dunne, and M. Giacomin. On the resolution-based family of abstract argumentation semantics and its grounded instance. *Artificial Intelligence*, 175:791–813, 2011.

[Baroni *et al.*, 2014] P. Baroni, G. Boella, F. Cerutti, M. Giacomin, L. van der Torre, and S. Villata. On the input/output behavior of argumentation frameworks. *Artificial Intelligence*, 217:144–197, 2014.

[Baroni *et al.*, 2015] P. Baroni, M. Giacomin, and B. Lao. Dealing with generic contrariness in structured argumentation. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, pages 2727–2733, 2015.

[Baumann and Brewka, 2010] R. Baumann and G. Brewka. Expanding argumentation frameworks: Enforcing and monotonicity results. In P. Baroni, F. Cerutti, M. Giacomin, and G.R. Simari, editors, *Computational Models of Argument. Proceedings of COMMA 2010*, pages 75–86. IOS Press, Amsterdam etc, 2010.

[Baumann, 2012a] R. Baumann. Normal and strong expansion equivalence for argumentation frameworks. *Artificial Intelligence*, 193:18–44, 2012.

[Baumann, 2012b] R. Baumann. What does it take to enforce an argument? minimal change in abstract argumentation. In *Proceedings of the 20th European Conference on Artificial Intelligence*, pages 127–132, 2012.

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

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

[Bench-Capon, 1998] T.J.M. Bench-Capon. Specification and implementation of Toulmin dialogue game. In *Legal Knowledge-Based Systems. JURIX: The Eleventh Conference*, pages 5–19, Nijmegen, 1998. Gerard Noodt Instituut.

[Bench-Capon, 2002] T.J.M. Bench-Capon. The missing link revisited: the role of teleology in representing legal argument. *Artificial Intelligence and Law*, 10:79–94, 2002.

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

[Bench-Capon, 2017] T.J.M. Bench-Capon. HYPOs legacy: introduction to the virtual special issue. *Artificial Intelligence and Law*, 25:205–250, 2017.

[Benferhat *et al.*, 1993] S. Benferhat, D. Dubois, and H. Prade. Argumentative inference in uncertain and inconsistent knowledge bases. In *Proceedings of the 9th International Conference on Uncertainty in Artificial Intelligence*, pages 411–419, 1993.

[Berman and Hafner, 1993] D.H. Berman and C.D. Hafner. Representing teleological structure in case-based legal reasoning: the missing link. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*, pages 50–59, New York, 1993. ACM Press.

[Besnard and Doutre, 2004] Ph. Besnard and S. Doutre. Checking the acceptability of a set of arguments. In *Proceedings of the Tenth International Workshop on Nonmonotonic Reasoning*, pages 59–64, Chistler, Canada, 2004.

[Besnard and Hunter, 2001] Ph. Besnard and A. Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128:203–235, 2001.

[Besnard and Hunter, 2008] Ph. Besnard and A. Hunter. *Elements of Argumentation.* MIT Press, Cambridge, MA, 2008.

[Besnard and Hunter, 2014] Ph. Besnard and H. Hunter. Constructing argument graphs with deductive arguments: a tutorial. *Argument and Computation*, 5:5–30, 2014.

[Bex *et al.*, 2013a] F.J. Bex, J. Lawrence, M. Snaith, and C. Reed. Implementing the argument web. *Communications of the ACM*, 56:66–73, 2013.

[Bex *et al.*, 2013b] F.J. Bex, S.J. Modgil, H. Prakken, and C.A. Reed. On logical specifications of the argument interchange format. *Journal of Logic and Computation*, 23:951–989, 2013.

[Bisquert *et al.*, 2013] P. Bisquert, C. Cayrol, F. Dupin de Saint-Cyr, and M-C Lagasquie-Schiex. Goal-driven changes in argumentation: a theoretical framework and a tool. In *Proceedings of the 25th International Conference on Tools with Artificial Intelligence (ICTAI 2013)*, pages 610–617, 2013.

[Black and Atkinson, 2011] E. Black and K.D. Atkinson. Agreeing what to do. In P. McBurney, I. Rahwan, and S. Parsons, editors, *Argumentation in Multi-Agent Systems, 7th International Workshop, ArgMAS 2010, Toronto, Canada, May 2010. Revised Selected and Invited Papers*, number 6614 in Springer Lecture Notes in AI, pages 1–88. Springer Verlag, Berlin, 2011.

[Black and Hunter, 2007] E. Black and A. Hunter. A generative inquiry dialogue system. In *Proceedings of the Sixth International Conference on Autonomous Agents and Multiagent Systems*, pages 1010–1017, 2007.

[Black and Hunter, 2009] E. Black and A. Hunter. An inquiry dialogue system. *Journal of Autonomous Agents and Multi-Agent Systems*, 19:173–209, 2009.

[Black and Hunter, 2012] E. Black and A. Hunter. A relevance-theoretic framework for constructing and deconstructing enthymemes. *Journal of Logic and Computation*, 22:55–78, 2012.

[Bobrow, 1980] D.G. Bobrow, editor. *Artificial Intelligence*, volume 13. 1980. Special issue on Non-Monotonic Logic.

[Bochman, 2003] A. Bochman. Collective argumentation and disjunctive logic programming. *Journal of Logic and Computation*, 13:405–428, 2003.

[Bodenstaff *et al.*, 2006] L. Bodenstaff, H. Prakken, and G. Vreeswijk. On formalising dialogue systems for argumentation in the event calculus. In *Proceedings of the Eleventh International Workshop on Nonmonotonic Reasoning*, pages 374–382, Windermere, UK, 2006.

[Boella *et al.*, 2010a] G. Boella, D.M. Gabbay, L. van der Torre, and S. Villata. Support in abstract argumentation. In P. Baroni, F. Cerutti, M. Giacomin, and G.R. Simari, editors, *Computational Models of Argument. Proceedings of COMMA 2010*, pages 111–122. IOS Press, Amsterdam etc, 2010.

[Boella *et al.*, 2010b] G. Boella, S. Kaci, and L. van der Torre. Dynamics in argumentation with single extensions: attack refinement and the grounded extension (extended version). In P. McBurney, I. Rahwan, S. Parsons, and N. Maudet, editors, *Argumentation in Multi-Agent Systems, 6th International Workshop, ArgMAS 2009, Budapest, Hungary, May 12, 2009. Revised Selected and Invited Papers*, number 6057 in Springer Lecture Notes in AI, pages 150–159. Springer Verlag, Berlin, 2010.

[Bondarenko *et al.*, 1993] A. Bondarenko, R.A. Kowalski, and F. Toni. An assumption-based framework for non-monotonic reasoning. In *Proceedings of the second International Workshop on Logic Programming and Nonmonotonic Logic*, pages 171–189, Lisbon (Portugal), 1993.

[Bondarenko *et al.*, 1997] A. Bondarenko, P.M. Dung, R.A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997.

[Brewka and Woltran, 2010] G. Brewka and S. Woltran. Abstract dialectical frameworks. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference*, pages 102–111. AAAI Press, 2010.

[Brewka, 1989] G. Brewka. Preferred subtheories: An extended logical framework for default reasoning. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 1043–1048, 1989.

[Brewka, 2001] G. Brewka. Dynamic argument systems: a formal model of argumentation processes based on situation calculus. *Journal of Logic and Computation*, 11:257–282, 2001.

[Caminada and Amgoud, 2005] M. Caminada and L. Amgoud. An axiomatic account of formal argumentation. In *Proceedings of the 20th National Conference on Artificial Intelligence*, Pittsburgh, PA, 2005.

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

[Caminada *et al.*, 2012] M. Caminada, W.A. Carnielli, and P.E. Dunne. Semi-stable semantics. *Journal of Logic and Computation*, 22:1207–1254, 2012.

[Caminada *et al.*, 2015] M. Caminada, S. Sa, J. Alcantara, and W. Dvorak. On the difference between assumption-based argumentation and abstract argumentation. *IFCoLog Journal of Logic and its Applications*, 2:15–34, 2015.

[Caminada, 2006] M. Caminada. On the issue of reinstatement in argumentation. In M. Fischer, W. van der Hoek, B. Konev, and A. Lisitsa, editors, *Logics in Artificial Intelligence. Proceedings of JELIA 2006*, number 4160 in Springer Lecture Notes in AI, pages 111–123, Berlin, 2006. Springer Verlag.

[Carlson, 1983] L. Carlson. *Dialogue Games: an Approach to Discourse Analysis.* Reidel Publishing Company, Dordrecht, 1983.

[Cartwright and Atkinson, 2009] D. Cartwright and K. Atkinson. Using computational argumentation to support e-participation. *IEEE Intelligent Systems*, 24:42–52, 2009. Special Issue on Transforming E-government and E-participation through IT.

[Cayrol and Lagasquie-Schiex, 2005a] C. Cayrol and M.-C. Lagasquie-Schiex. Graduality in argumentation. *Journal of Artificial Intelligence Research*, 23:245–297, 2005.

[Cayrol and Lagasquie-Schiex, 2005b] C. Cayrol and M.-C. Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation. In L. Godo, editor, *Proceedings of the 8nd European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 05)*, number 3571 in Springer Lecture Notes in AI, pages 378–389, Berlin, 2005. Springer Verlag.

[Cayrol and Lagasquie-Schiex, 2009] C. Cayrol and M.-C. Lagasquie-Schiex. Bipolar abstract argumentation systems. In I. Rahwan and G.R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 65–84. Springer, Berlin, 2009.

[Cayrol and Lagasquie-Schiex, 2013] C. Cayrol and M.-C. Lagasquie-Schiex. Bipolarity in argumentation graphs: Towards a better understanding. *International Journal of Approximate Reasoning*, 54:876–899, 2013.

[Cayrol *et al.*, 2003] C. Cayrol, S. Doutre, and J. Mengin. On decision problems related to the preferred semantics for argumentation frameworks. *Journal of Logic and Computation*, 13, 2003. 377-403.

[Cayrol *et al.*, 2010] C. Cayrol, F Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Change in abstract argumentation frameworks: adding an argument. *Journal of Artificial Intelligence Research*, 38:49–84, 2010.

[Cayrol, 1995] C. Cayrol. On the relation between argumentation and non-monotonic coherence-based entailment. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 1443–1448, 1995.

[Chesñevar *et al.*, 2004] C.I. Chesñevar, G. Simari, T. Alsinet, and L. Godo. A logic programming framework for possibilistic argumentation with vague knowledge. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, pages 76–84, 2004.

[Chesñevar *et al.*, 2006] C.I. Chesñevar, J. McGinnis, S. Modgil, I. Rahwan, C. Reed, G. Simari, M. South, G. Vreeswijk, and S. Willmott. Towards an argument interchange format. *The Knowledge Engineering Review*, 21:293–316, 2006.

[Chisholm, 1957] R. Chisholm. *Perceiving: A Philosophical Study.* Cornell University Press, Ithaca, 1957.

[Chisholm, 1974] Roderick Chisholm. Practical reason and the logic of requirement. In Stephan Körner, editor, *Practical Reason*, pages 2–13. Blackwell Publishing Company, 1974.

[Conklin *et al.*, 2001] J. Conklin, A. Selvin, S. Buckingham Shum, and M. Sierhuis. Facilitating hypertext for collective sensemaking: 15 years on from gIBIS. In *Proceedings of the The Twelfth ACM Conference on Hypertext and Hypermedia (Hypertext 2001)*, New York, 2001. ACM Press. Also available as Technical Report KMI-TR-112, Knowledge Media Institute, The Open University, UK.

[Coste-Marquis *et al.*, 2006] S. Coste-Marquis, C. Devred, and P. Marquis. Constrained argumentation frameworks. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Tenth International Conference*, pages 112–122. AAAI Press, 2006.

[Coste-Marquis *et al.*, 2007] S. Coste-Marquis, C. Devred, S. Konieczny, and P. Marquis. On the merging of dung's argumentation systems. *Artificial Intelligence*, 171:730–753, 2007.

[Coste-Marquis *et al.*, 2014] S. Coste-Marquis, S. Konieczny, J.-G. Mailly, and P. Marquis. On the revision of argumentation systems: minimal change of arguments statuses. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference*, pages 72–81. AAAI Press, 2014.

[Dimopoulos and Kakas, 1995] Y. Dimopoulos and A.C. Kakas. Logic programming without negation as failure. In *Proceedings of the Fifth International Logic Programming Symposium*, pages 369–384, Cambridge, MA, 1995. MIT Press.

[Doyle, 1979] J. Doyle. Truth maintenance systems. *Artificial Intelligence*, 12:231–272, 1979.

[Dung and Thang, 2010] P.M. Dung and P.M. Thang. Towards (probabilistic) argumentation for jury-based dispute resolution. In P. Baroni, F. Cerutti, M. Giacomin, and G.R. Simari, editors, *Computational Models of Argument. Proceedings of COMMA 2010*, pages 171–182. IOS Press, Amsterdam etc, 2010.

[Dung and Thang, 2014] P.M. Dung and P.M. Thang. Closure and consistency in logic-associated argumentation. *Journal of Artificial Intelligence Research*, 49:79–109, 2014.

[Dung *et al.*, 2007] P.M. Dung, P. Mancarella, and F. Toni. Computing ideal sceptical argumentation. *Artificial Intelligence*, 171:642–674, 2007.

[Dung, 1992] P.M. Dung. Logic programming as dialog game. Technical Report, Division of Computer Science, Asian Institute of Technology, Bangkok, 1992.

[Dung, 1993] P.M. Dung. An argumentation semantics for logic programming with explicit negation. In *Proceedings of the Tenth Logic Programming Conference*, pages 616–630, Cambridge, MA, 1993. MIT Press.

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

[Dung, 2014] P.M. Dung. An axiomatic analysis of structured argumentation for prioritized default reasoning. In *Proceedings of the 21st European Conference on Artificial Intelligence*, pages 267–272, 2014.

[Dung, 2016] P.M. Dung. An axiomatic analysis of structured argumentation with priorities. *Artificial Intelligence*, 231:107–150, 2016.

[Dunne and Bench-Capon, 2002] P.E. Dunne and T.J. Bench-Capon. Coherence in finite argumentation systems. *Artificial Intelligence*, 141:187–203, 2002.

[Dunne and Bench-Capon, 2003] P.E. Dunne and T.J. Bench-Capon. Two party immediate response disputes: Properties and efficiency. *Artificial Intelligence*, 149:221–250, 2003.

[Dunne *et al.*, 2011] P.E. Dunne, A. Hunter, P. McBurney, S. Parsons, and M. Wooldridge. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artificial Intelligence*, 175:457–486, 2011.

[Dunne, 2003] P.E. Dunne. Prevarication in dispute protocols. In *Proceedings of the Ninth International Conference on Artificial Intelligence and Law*, pages 12–21, New York, 2003. ACM Press.

[Dunne, 2006] P.E. Dunne. Suspicion of hidden agenda in persuasive argument. In P.E. Dunne and T.B.C. Bench-Capon, editors, *Computational Models of Argument. Proceedings of COMMA 2006*, pages 329–340. IOS Press, Amsterdam etc, 2006.

[Dunne, 2007] P.E. Dunne. Computational properties of argument systems satisfying graph-theoretical constraints. *Artificial Intelligence*, 171:701–729, 2007.

[Dvorak and Woltran, 2011] W. Dvorak and S. Woltran. On the intertranslatability of argumentation semantics. *Journal of Artificial Intelligence Research*, 41:445–475, 2011.

[Egly *et al.*, 2010] U. Egly, S. Gaggl, and S. Woltran. Answer-set programming encodings for argumentation frameworks. *Argument and Computation*, 1:147–177, 2010.

[Elvang-Göransson *et al.*, 1993] M. Elvang-Göransson, J. Fox, and P. Krause. Acceptability of arguments as logical uncertainty. In *Proceedings of the 2nd European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 93)*, pages 85–90, Berlin, 1993. Springer Verlag.

[Fan and Toni, 2013] X. Fan and F. Toni. Decision making with assumption-based argumentation. In E. Black, S. Modgil, and N. Oren, editors, *Theories and Applications of Formal Argumentation - Second International Workshop, TAFA 2013*, number 8306 in Springer Lecture Notes in AI, pages 127–142, Berlin, 2013. Springer Verlag.

[Fox and Parsons, 1997] J. Fox and S. Parsons. On using arguments for reasoning about actions and values. In *Proceedings of the AAAI Spring Symposium on Qualitative Preferences in Deliberation and Practical Reasoning*, Stanford, CA, 1997.

[Fox *et al.*, 2007] J. Fox, D. Glasspool, D. Greca, S. Modgil, M. South, and V. Patkar. Argumentation-based inference and decision making – a medical perspective. *IEEE Intelligent Systems*, 22:34–41, 2007.

[Franqueira *et al.*, 2011] V.N.L. Franqueira, T.T. Tun, Y.Yu, R. Wieringa, and B. Nuseibeh. Risk and argument: a risk-based argumentation method for practical security. In *Proceedings of the 19th IEEE International Requirements Engineering Conference*, pages 239–248, Trento, Italy, 2011.

[Gabbay, 1985] D. Gabbay. Theoretical foundations for non-monotonic reasoning in expert systems. In Krzysztof R. Apt, editor, *Logics and Models of Concurrent Systems*, pages 439–457. Springer-Verlag New York, Inc., New York, NY, USA, 1985.

[Gabbay, 2011] D. Gabbay. Introducing equational semantics for argumentation networks. In *Proceedings of the 11th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2011)*, number 6717 in Springer Lecture Notes in Computer Science, pages 19–35, Berlin, 2011. Springer Verlag.

[Gabbay, 2015] D. Gabbay. Editorial comment about "on the difference between ABA and AA. *IFCoLog Journal of Logic and its Applications*, 2:1–14, 2015.

[Garcia and Simari, 2004] A.J. Garcia and G.R. Simari. Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming*, 4:95–138, 2004.

[Garcia *et al.*, 1998] A.J. Garcia, G.R. Simari, and C.I. Chesñevar. An argumentative framework for reasoning with inconsistent and incomplete information. In *Proceedings of the ECAI'98 Workshop on Practical Reasoning and Rationality*, Brighton, UK, 1998.

[Geffner and Pearl, 1992] H. Geffner and J. Pearl. Conditional entailment: bridging two approaches to default reasoning. *Artificial Intelligence*, 53:209–244, 1992.

[Geffner, 1991] H. Geffner. Beyond negation as failure. In *Principles of Knowledge Representation and Reasoning: Proceedings of the second International Conference*, pages 218–229. AAAI Press, 1991.

[Geffner, 1992] H. Geffner. *Default reasoning: causal and conditional theories.* MIT Press, Cambridge, MA, 1992.

[Gordon and Walton, 2009] T.F. Gordon and D.N. Walton. Proof burdens and standards. In I. Rahwan and G.R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 239–258. Springer, Berlin, 2009.

[Gordon *et al.*, 2007] T.F. Gordon, H. Prakken, and D.N. Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171:875–896, 2007.

[Gordon, 1994] T.F. Gordon. The Pleadings Game: an exercise in computational dialectics. *Artificial Intelligence and Law*, 2:239–292, 1994.

[Gordon, 1995] T.F. Gordon. *The Pleadings Game. An Artificial Intelligence Model of Procedural Justice*. Kluwer Academic Publishers, Dordrecht/Boston/London, 1995.

[Gordon, 2011] T.F. Gordon. The policy making tool of the IMPACT argumentation toolbox. In *Proceedings of the Jurix Workshop on Modelling Policy-Making (MPM 2011)*, pages 29–38, 2011.

[Gorogiannis and Hunter, 2011] N. Gorogiannis and A. Hunter. Instantiating abstract argumentation with classical-logic arguments: postulates and properties. *Artificial Intelligence*, 175:1479–1497, 2011.

[Governatori *et al.*, 2004] G. Governatori, M.J. Maher, G. Antoniou, and D. Billington. Argumentation semantics for defeasible logic. *Journal of Logic and Computation*, 14:675–702, 2004.

[Grasso *et al.*, 2001] F. Grasso, A. Cawsey, and R. Jones. Dialectical argumentation to solve conflicts in advice giving: a case study in the promotion of healthy nutrition. *International Journal of Human-Computer Studies*, 53:1077–1115, 2001.

[Grossi and Modgil, 2015] D. Grossi and S. Modgil. On the graded acceptability of arguments. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, pages 868–874, 2015.

[Grossi, 2010] D. Grossi. On the logic of argumentation theory. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, pages 1147–1154, 2010.

[Hadjinikolis *et al.*, 2013] C. Hadjinikolis, S. Modgil, E. Black, P. McBurney, and Y. Siantos. Opponent modelling in persuasion dialogues. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, pages 164–170, 2013.

[Haenni *et al.*, 2000] R. Haenni, J. Kohlas, and N. Lehmann. Probabilistic argumentation systems. In J. Kohlas and S. Moral, editors, *Defeasible reasoning and uncertainty management systems: algorithms*. Oxford University Press, Oxford, 2000.

[Hage *et al.*, 1993] J.C. Hage, R.E. Leenes, and A.R. Lodder. Hard cases: a procedural approach. *Artificial Intelligence and Law*, 2:113–166, 1993.

[Hage, 1993] J.C. Hage. Monological reason-based logic: a low-level integration of rule-based reasoning and case-based reasoning. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*, pages 30–39, New York, 1993. ACM Press.

[Haley *et al.*, 2008] C. Haley, R. Laney, J. Moffett, and B. Nuseibeh. Security requirements engineering: A framework for representation and analysis. *IEEE Transactions on Software Engineering*, 34(1):133–153, 2008.

[Hamblin, 1970] C.L. Hamblin. *Fallacies*. Methuen, London, 1970.

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

[Hanks and McDermott, 1986] S. Hanks and D. McDermott. Default reasoning, nonmonotonic logics and the frame problem. In *Proceedings of the 5thth National Conference on Artificial Intelligence*, pages 328–333, 1986.

[Hart, 1949] H.L.A. Hart. The ascription of responsibility and rights. In *Proceedings of the Aristotelean Society*, pages 99–117, 1949. Reprinted in *Logic and Language. First Series*, ed. A.G.N. Flew, 145–166. Oxford: Basil Blackwell.

[Horty *et al.*, 1990] J. Horty, R.H. Thomason, and D.S. Touretzky. A skeptical theory of inheritance in nonmonotonic semantic networks. *Artificial Intelligence*, 42:311–348, 1990.

[Horty, 2011] J. Horty. Rules and reasons in the theory of precedent. *Legal Theory*, 17:1–33, 2011.

[Hunter and Thimm, 2016] A. Hunter and M. Thimm. On partial information and contradictions in probabilistic abstract argumentation. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference*, pages 53–62. AAAI Press, 2016.

[Hunter and Williams, 2012] A. Hunter and M. Williams. Aggregating evidence about the positive and negative effects of treatments. *Artificial Intelligence in Medicine*, 56:173–190, 2012.

[Hunter and Woltran, 2013] A. Hunter and S. Woltran. Structural properties for deductive argument systems. In *Proceedings of the 12th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2013)*, number 7958 in Springer Lecture Notes in Computer Science, pages 278–289, Berlin, 2013. Springer Verlag.

[Hunter, 2004] A. Hunter. Making arguments more believable. In *Proceedings of the 19th National Conference on Artificial Intelligence*, pages 6269–274, 2004.

[Hunter, 2013] A. Hunter. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning*, 54:47–81, 2013.

[Hunter, 2014] A. Hunter. Probabilistic qualification of attack in abstract argumentation. *International Journal of Approximate Reasoning*, 55:607–638, 2014.

[Jakobovits and Vermeir, 1999a] H. Jakobovits and D. Vermeir. Dialectic semantics for argumentation frameworks. In *Proceedings of the Seventh International Conference on Artificial Intelligence and Law*, pages 53–62, New York, 1999. ACM Press.

[Jakobovits and Vermeir, 1999b] H. Jakobovits and D. Vermeir. Robust semantics for argumentation frameworks. *Journal of Logic and Computation*, 9:215–261, 1999.

[Jakobovits, 2000] H. Jakobovits. *On the Theory of Argumentation Frameworks.* Doctoral dissertation Free University Brussels, 2000.

[Kaci *et al.*, 2007] S. Kaci, L. van der Torre, and E. Weydert. On the acceptability of incompatible arguments. In *Proceedings of the 9th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 07)*, number 4724 in Springer Lecture Notes in AI, page 247258, Berlin, 2007. Springer Verlag.

[Kaci, 2010] S. Kaci. Refined preference-based argumentation frameworks. In P. Baroni, F. Cerutti, M. Giacomin, and G.R. Simari, editors, *Computational Models of Argument. Proceedings of COMMA 2010.* IOS Press, Amsterdam etc, 2010.

[Kakas and Moraitis, 2003] A.C. Kakas and P. Moraitis. Argumentation based decision making for autonomous agents. In *Proceedings of the Second International Conference on Autonomous Agents and Multiagent Systems*, pages 883–890, 2003.

[Kakas *et al.*, 1992] A.C. Kakas, R.A. Kowalski, and F. Toni. Abductive logic programming. *Journal of Logic and Computation*, 2:719–770, 1992.

[Kirschner *et al.*, 2003] P.A. Kirschner, S.J. Buckingham Schum, and C.S. Carr. *Visualizing Argumentation. Software Tools for Collaborative and Educational Sense-Making.* Springer, London etc, 2003.

[Kok *et al.*, 2011] E. Kok, J.-J.Ch. Meyer, H. Prakken, and G. Vreeswijk. A formal argumentation framework for deliberation dialogues. In P. McBurney, I. Rahwan, and S. Parsons, editors, *Argumentation in Multi-Agent Systems, 7th International Workshop, ArgMAS 2010, Toronto, Canada, May 2010. Revised Selected and Invited Papers*, number 6614 in Springer Lecture Notes in AI, pages 31–48. Springer Verlag, Berlin, 2011.

[Kok, 2013] E.M. Kok. *Exploring the Practical Benefits of Argumentation in Multi-Agent Deliberation.* Doctoral dissertation Department of Information and Computing Sciences, Utrecht University, 2013.

[Konolige, 1988] K. Konolige. Defeasible argumentation in reasoning about events. In Z.W. Ras and L. Saitta, editors, *Methodologies for Intelligent Systems*, page 380390. Elsevier, Amsterdam, 1988.

[Kowalski and Toni, 1996] R.A. Kowalski and F. Toni. Abstract argumentation. *Artificial Intelligence and Law*, 4:275–296, 1996.

[Kraus *et al.*, 1990] S. Kraus, D.J. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[Kraus *et al.*, 1998] S. Kraus, K. Sycara, and A. Evenchik. Reaching agreements through argumentation: a logical model and implementation. *Artificial Intelligence*, 104:1–69, 1998.

[Krause *et al.*, 1995] P. Krause, S. Ambler, M. Elvang-Gøransson, and J. Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11(1):113–131, 1995.

[Li *et al.*, 2012] H. Li, N. Oren, and T. Norman. Probabilistic argumentation frameworks. In S. Modgil, N. Oren, and F. Toni, editors, *Theorie and Applications of Formal Argumentation. First International Workshop, TAFA 2011. Barcelona, Spain, July 16-17, 2011, Revised Selected Papers*, number 7132 in Springer Lecture Notes in AI, pages 1–16, Berlin, 2012. Springer Verlag.

[Lin and Shoham, 1989] F. Lin and Y. Shoham. Argument systems. A uniform basis for nonmonotonic reasoning. In *Principles of Knowledge Representation and Reasoning: Proceedings of the First International Conference*, pages 245–255, San Mateo, CA, 1989. Morgan Kaufmann Publishers.

[Lippi and Torroni, 2016] M. Lippi and P. Torroni. Argument mining: state of the art and emerging trends. *ACM Transactions on Internet Technology*, 10(2), 2016.

[Lodder, 1999] A.R. Lodder. *DiaLaw. On Legal Justification and Dialogical Models of Argumentation*. Law and Philosophy Library. Kluwer Academic Publishers, Dordrecht/-Boston/London, 1999.

[Lorenzen and Lorenz, 1978] P. Lorenzen and K. Lorenz. *Dialogische Logik*. Wissenschaftliche Buchgesellschaft, Darmstadt, 1978.

[Loui *et al.*, 1993] R.P. Loui, J. Norman, J. Olson, and A. Merrill. A design for reasoning with policies, precedents, and rationales. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*, pages 202–211, New York, 1993. ACM Press.

[Loui, 1987] R.P. Loui. Defeat among arguments: a system of defeasible inference. *Computational Intelligence*, 2:100–106, 1987.

[Loui, 1995] R.P. Loui. Hart's critics on defeasible concepts and ascriptivism. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*, pages 21–30, New York, 1995. ACM Press.

[Loui, 1998] R.P. Loui. Process and policy: resource-bounded non-demonstrative reasoning. *Computational Intelligence*, 14:1–38, 1998.

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

[Mackenzie, 1990] J.D. Mackenzie. Four dialogue systems. *Studia Logica*, 51:567–583, 1990.

[Makinson, 1994] D. Makinson. General patterns in nonmonotonic reasoning. In D. Gabbay, C.J. Hogger, and J.A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, pages 35–110. Clarendon Press, Oxford, 1994.

[Martinez *et al.*, 2006] D.C. Martinez, A.J. Garcia, and G.R. Simari. On acceptability in abstract argumentation frameworks with an extended defeat relation. In P.E. Dunne and T.B.C. Bench-Capon, editors, *Computational Models of Argument. Proceedings of COMMA 2006*, pages 273–278. IOS Press, Amsterdam etc, 2006.

[Matt and Toni, 2008] P.-A. Matt and F. Toni. A game-theoretic measure of argument strength for abstract argumentation. In S. Hoelldobler, C. Lutz, and H. Wansing, editors, *Proceedings of the 11th European Conference on Logics in Artificial Intelligence (JELIA 2008)*, number 5293 in Springer Lecture Notes in AI, pages 285–297, Berlin, 2008. Springer Verlag.

[Maudet and Moore, 1999] N. Maudet and D. Moore. Dialogue games for computer-supported collaborative argumentation. In *Proceedings of the Workshop on Computer-Supported Collaborative Argumentation for Learning Communities*, Stanford, 1999.

[McBurney and Parsons, 2001] P. McBurney and S. Parsons. Representing epistemic uncertainty by means of dialectical argumentation. *Annals of Mathematics and Artificial Intelligence*, 32:125–169, 2001.

[McBurney and Parsons, 2002] P. McBurney and S. Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language and Information*, 13:315–343, 2002.

[McBurney *et al.*, 2002] P. McBurney, S. Parsons, and M. Wooldridge. Desiderata for agent argumentation protocols. In *Proceedings of the First International Conference on Autonomous Agents and Multiagent Systems*, pages 402–409, 2002.

[McBurney *et al.*, 2007] P. McBurney, D. Hitchcock, and S. Parsons. The eightfold way of deliberation dialogue. *International Journal of Intelligent Systems*, 22:95–132, 2007.

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

[McCarty, 1995] L.T. McCarty. An implementation of Eisner v. Macomber. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*, pages 276–286, New York, 1995. ACM Press.

[Modgil and Caminada, 2009] S. Modgil and M. Caminada. Proof theories and algorithms for abstract argumentation frameworks. In I. Rahwan and G.R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 105–129. Springer, Berlin, 2009.

[Modgil and Prakken, 2012] S. Modgil and H. Prakken. Resolutions in structured argumentation. In B. Verheij, S. Woltran, and S. Szeider, editors, *Computational Models of Argument. Proceedings of COMMA 2012*, pages 310–321. IOS Press, Amsterdam etc, 2012.

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

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

[Modgil *et al.*, 2013] S. Modgil, F. Toni, F. Bex, I. Bratko, C. Chesñevar, W. Dvorak, M. Falappa, X. Fan, S. Gaggl, A. Garcia, M. Gonzalez, T. Gordon, J. Leite, M. Mozina, C. Reed, G. Simari, S. Szeider, P. Torroni, and S. Woltran. The added value of argumentation. In S. Ossowski, editor, *Agreement Technologies*, pages 357–403. Springer, 2013.

[Modgil, 2006] S. Modgil. Hierarchical argumentation. In M. Fischer, W. van der Hoek, B. Konev, and A. Lisitsa, editors, *Logics in Artificial Intelligence. Proceedings of JELIA 2006*, number 4160 in Springer Lecture Notes in AI, pages 319–332, Berlin, 2006. Springer Verlag.

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

[Moore, 1993] D. Moore. *Dialogue game theory for intelligent tutoring systems.* PhD Thesis, Leeds Metropolitan University, 1993.

[Nielsen and Parsons, 2007a] S.H. Nielsen and S. Parsons. An application of formal argumentation: fusing Bayesian networks in multi-agent systems. *Artificial Intelligence*, 171:754–775, 2007.

[Nielsen and Parsons, 2007b] S.H. Nielsen and S. Parsons. A generalization of Dung's abstract framework for argumentation: Arguing with sets of attacking arguments. In S. Parsons, N. Maudet, and I. Rahwan, editors, *Argumentation in Multi-Agent Systems. Third International Workshop, ArgMAS 2006, Hakodate, Japan, May 8, 2006, Revised Selected and Invited Papers*, number 4766 in Springer Lecture Notes in AI, pages 54–73. Springer Verlag, Berlin, 2007.

[Nouioua and Risch, 2011] F. Nouioua and V. Risch. Argumentation frameworks with necessities. In *Proceedings of the 4th International Conference on Scalable Uncertainty Management (SUM'11)*, number 6929 in Springer Lecture Notes in AI, pages 163–176, Berlin, 2011. Springer Verlag.

[Nute, 1988] Donald Nute. Defeasible reasoning: A philosophical analysis in prolog. In J. Fetzer, editor, *Aspects of Artificial Intelligence*, pages 251–288. Kluwer Academic Publishers, 1988.

[Nute, 1994] D. Nute. Defeasible logic. In D. Gabbay, C.J. Hogger, and J.A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, pages 253–395. Clarendon Press, Oxford, 1994.

[Oren and Norman, 2008] N. Oren and T.J. Norman. Semantics for evidence-based argumentation. In Ph. Besnard, S. Doutre, and A. Hunter, editors, *Computational Models of Argument. Proceedings of COMMA 2008*, pages 276–284, Amsterdam etc, 2008. IOS Press.

[Oren and Norman, 2010] N. Oren and T. Norman. Arguing using opponent models. In P. McBurney, I. Rahwan, S. Parsons, and N. Maudet, editors, *Argumentation in Multi-Agent Systems, 6th International Workshop, ArgMAS 2009, Budapest, Hungary, May 12, 2009. Revised Selected and Invited Papers*, number 6057 in Springer Lecture Notes in AI, pages 160–174. Springer Verlag, Berlin, 2010.

[Palau and Moens, 2009] R. Mochales Palau and M.-F. Moens. Argumentation mining: the detection, classification and structure of arguments in text. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law*, pages 98–107, New York, 2009. ACM Press.

[Parsons and Jennings, 1996] S. Parsons and N. Jennings. Negotiation through argumentation - a preliminary report. In *Proceedings of the 2nd International Conference in Multiagent Systems*, pages 267–274, Kyoto, Japan, 1996.

[Parsons *et al.*, 1998] S. Parsons, C. Sierra, and N.R. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8:261–292, 1998.

[Parsons *et al.*, 2002] S. Parsons, M. Wooldridge, and L. Amgoud. An analysis of formal interagent dialogues. In *Proceedings of the First International Conference on Autonomous Agents and Multiagent Systems*, pages 394–401, 2002.

[Parsons *et al.*, 2003] S. Parsons, M. Wooldridge, and L. Amgoud. Properties and complexity of some formal inter-agent dialogues. *Journal of Logic and Computation*, 13, 2003. 347-376.

[Parsons, 1998a] S. Parsons. On precise and correct qualitative probabilistic reasoning. *International Journal of Approximate Reasoning*, 35:111–135, 1998.

[Parsons, 1998b] S. Parsons. A proof-theoretic approach to qualitative probabilistic reasoning. *International Journal of Approximate Reasoning*, 19:265–297, 1998.

[Perelman and Olbrechts-Tyteca, 1969] Ch. Perelman and L. Olbrechts-Tyteca. *The New Rhetoric. A Treatise on Argumentation.* University of Notre Dame Press, Notre Dame, Indiana, 1969.

[Pinkwart and McLaren, 2012] N. Pinkwart and B.M. McLaren, editors. *Educational Technologies for Teaching Argumentation Skills.* Bentham Science Publishers, Sharjah, UAE, 2012.

[Pollock, 1970] John Pollock. The structure of epistemic justification. In *Studies in the Theory of Knowledge*, American Philosophical Quarterly Monograph Series, number 4, pages 62–78. Basil Blackwell Publisher, Inc., 1970.

[Pollock, 1974] J.L. Pollock. *Knowledge and Justification.* Princeton University Press, Princeton, 1974.

[Pollock, 1987] J.L. Pollock. Defeasible reasoning. *Cognitive Science*, 11:481–518, 1987.

[Pollock, 1994] J.L. Pollock. Justification and defeat. *Artificial Intelligence*, 67:377–408, 1994.

[Pollock, 1995] J.L. Pollock. *Cognitive Carpentry. A Blueprint for How to Build a Person.* MIT Press, Cambridge, MA, 1995.

[Pollock, 2002] J.L. Pollock. Defeasible reasoning with variable degrees of justification. *Artificial Intelligence*, 133:233–282, 2002.

[Pollock, 2007a] J.L. Pollock. Reasoning and probability. *Law, Probability and Risk*, 6:43–58, 2007a.

[Pollock, 2007b] J.L. Pollock. Defeasible reasoning. In J. Adler and L. Rips, editors, *Reasoning: Studies of Human Inference and its Foundations*, pages 451–470. Cambridge, Cambridge University Press, 2007b.

[Pollock, 2010] J.L. Pollock. Defeasible reasoning and degrees of justification. *Argument and Computation*, 1:7–22, 2010.

[Poole, 1989] D.L. Poole. Explanation and prediction: an architecture for default and abductive reasoning. *Computational Intelligence*, 5:97–110, 1989.

[Prakken and Horty, 2012] H. Prakken and J. Horty. An appreciation of John Pollock's work on the computational study of argument. *Argument and Computation*, 3:1–19, 2012.

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

[Prakken and Sartor, 1997] H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-classical Logics*, 7:25–75, 1997.

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

[Prakken and Sartor, 2009] H. Prakken and G. Sartor. A logical analysis of burdens of proof. In H. Kaptein, H. Prakken, and B. Verheij, editors, *Legal Evidence and Proof: Statistics, Stories, Logic*, pages 223–253. Ashgate Publishing, Farnham, 2009.

[Prakken and Sartor, 2015] H. Prakken and G. Sartor. Law and logic: A review from an argumentation perspective. *Artificial Intelligence*, 227:214–225, 2015.

[Prakken and Vreeswijk, 2002] H. Prakken and G.A.W. Vreeswijk. Logics for defeasible argumentation. In D. Gabbay and F. Günthner, editors, *Handbook of Philosophical Logic*, volume 4, pages 219–318. Kluwer Academic Publishers, Dordrecht/Boston/London, second edition, 2002.

[Prakken *et al.*, 2015] H. Prakken, A.Z. Wyner, T.J.M. Bench-Capon, and K. Atkinson. A formalisation of argumentation schemes for legal case-based reasoning in ASPIC+. *Journal of Logic and Computation*, 25:1141–1166, 2015.

[Prakken, 1993] H. Prakken. An argumentation framework in default logic. *Annals of Mathematics and Artificial Intelligence*, 9:91–132, 1993.

[Prakken, 1999] H. Prakken. Dialectical proof theory for defeasible argumentation with defeasible priorities (preliminary report). In J.-J.Ch. Meyer and P.-Y. Schobbens, editors, *Formal Models of Agents*, number 1760 in Springer Lecture Notes in AI, pages 202–215, Berlin, 1999. Springer Verlag.

[Prakken, 2001a] H. Prakken. Modelling reasoning about evidence in legal procedure. In *Proceedings of the Eighth International Conference on Artificial Intelligence and Law*, pages 119–128, New York, 2001. ACM Press.

[Prakken, 2001b] H. Prakken. Relating protocols for dynamic dispute with logics for defeasible argumentation. *Synthese*, 127:187–219, 2001.

[Prakken, 2005] H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 15:1009–1040, 2005.

[Prakken, 2008] H. Prakken. A formal model of adjudication dialogues. *Artificial Intelligence and Law*, 16:305–328, 2008.

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

[Prakken, 2012] H. Prakken. Some reflections on two current trends in formal argumentation. In *Logic Programs, Norms and Action. Essays in Honour of Marek J. Sergot on the Occasion of his 60th Birthday*, pages 249–272. Springer, Berlin/Heidelberg, 2012.

[Prakken, 2014] H. Prakken. On support relations in abstract argumentation as abstractions of inferential relations. In *Proceedings of the 21st European Conference on Artificial Intelligence*, pages 735–740, 2014.

[Rahwan and Larson, 2008] I. Rahwan and K. Larson. Mechanism design for abstract argumentation. In *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems*, pages 1031–1038, 2008.

[Rahwan *et al.*, 2009] I. Rahwan, K. Larson, and F. Thomé. A characterisation of strategy-proofness for grounded argumentation semantics. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, pages 251–256, 2009.

[Rao and Georgeff, 1991] A.S. Rao and M.P. Georgeff. Modelling rational agents within a BDI-architecture. In *Principles of Knowledge Representation and Reasoning: Proceedings of the second International Conference*, pages 473–484. AAAI Press, 1991.

[Raz, 1975] J. Raz. *Practical Reason and Norms.* Princeton University Press, Princeton, 1975.

[Reed and Norman, 2003] C. Reed and T.J. Norman, editors. *Argumentation Machines. New Frontiers in Argument and Computation*, volume 9 of *Argumentation Library.* Kluwer Academic Publishers, Boston/Dordrecht/London, 2003.

[Reed *et al.*, 2007] C. Reed, D. Walton, and F. Macagno. Argument diagramming in logic, law and artificial intelligence. *The Knowledge Engineering Review*, 22:87–109, 2007.

[Reiter, 1980] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[Rescher, 1976] N. Rescher. *Plausible Reasoning.* Van Gorcum, Assen, 1976.

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

[Rienstra *et al.*, 2013] T. Rienstra, M. Thimm, and N. Oren. Opponent modelling with uncertainty for strategic argumentation. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, pages 332–338, 2013.

[Rissland and Ashley, 1987] E.L. Rissland and K.D. Ashley. A case-based system for trade secrets law. In *Proceedings of the First International Conference on Artificial Intelligence and Law*, pages 60–66, New York, 1987. ACM Press.

[Riveret *et al.*, 2007] R. Riveret, A. Rotolo, G. Sartor, H. Prakken, and B. Roth. Success chances in argument games: a probabilistic approach to legal disputes. In A.R. Lodder and L. Mommers, editors, *Legal Knowledge and Information Systems. JURIX 2007: The Twentieth Annual Conference*, pages 99–108. IOS Press, Amsterdam etc., 2007.

[Riveret *et al.*, 2008] R. Riveret, H. Prakken, A. Rotolo, and G. Sartor. Heuristics in argumentation: a game-theoretical investigation. In Ph. Besnard, S. Doutre, and A. Hunter, editors, *Computational Models of Argument. Proceedings of COMMA 2008*, pages 324–335. IOS Press, Amsterdam etc, 2008.

[Ross, 1930] W.D. Ross. *The Right and the Good.* Oxford University Press, Oxford, 1930.

[Rotstein *et al.*, 2008] N.D. Rotstein, M.O. Moguillansky, M.A. Falappa, A.J. Garcia, and G.R. Simari. Argument theory change: Revision upon warrant. In Ph. Besnard, S. Doutre, and A. Hunter, editors, *Computational Models of Argument. Proceedings of COMMA 2008*, pages 336–347, Amsterdam etc, 2008. IOS Press.

[Scheuer *et al.*, 2010] O. Scheuer, F. Loll, N. Pinkwart, and B.M. McLaren. Computer-supported argumentation: A review of the state-of-the-art. *International Journal of Computer-Supported Collaborative Learning*, 5:43–102, 2010.

[Schneider *et al.*, 2013] J. Schneider, T. Groza, and A. Passant. A review of argumentation for the social semantic web. *Semantic Web*, 4:159–218, 2013.

[Searle, 1969] J.R. Searle. *Speech Acts. An Essay in the Philosophy of Language.* Cambride University Press, Cambridge, UK, 1969.

[Simari and Loui, 1992] G.R. Simari and R.P. Loui. A mathematical treatment of defeasible argumentation and its implementation. *Artificial Intelligence*, 53:125–157, 1992.

[Snaid and Reed, 2012] M. Snaid and C. Reed. TOAST: Online ASPIC+ argumentation. In B. Verheij, S. Woltran, and S. Szeider, editors, *Computational Models of Argument. Proceedings of COMMA 2012*, pages 509–510. IOS Press, Amsterdam etc, 2012.

[Strass, 2013] H. Strass. Instantiating knowledge bases in abstract dialectical frameworks. In J. Leite, T.C. Son, P. Torroni, L. van der Torre, and S. Woltran, editors, *Proceedings of the 14th International Workshop on Computational Logic in Multi-Agent Systems (CLIMA XIV)*, number 8143 in Springer Lecture Notes in AI, pages 86–101, Berlin, 2013. Springer Verlag.

[Sycara, 1985] K.P. Sycara. Arguments of persuasion in labour mediation. In *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, pages 294–296, 1985.

[Sycara, 1990] K.P. Sycara. Persuasive argumentation in negotiation. *Theory and Decision*, 28:203–242, 1990.

[Tang and Parsons, 2005] Y. Tang and S. Parsons. Argumentation-based dialogues for deliberation. In *Proceedings of the Fourth International Conference on Autonomous Agents and Multiagent Systems (AAMAS-05)*, pages 552–559, 2005.

[ter Berg *et al.*, 2009] T. ter Berg, T. van Gelder, F. Patterson, and S. Teppema. *Critical Thinking: Reasoning and Communicating with Rationale.* Pearson Education Benelux, Amsterdam, 2009.

[Thimm and Garcia, 2010] M. Thimm and A.J. Garcia. Classification and strategical issues of argumentation games on structured argumentation frameworks. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, pages 1247–1254, 2010.

[Thimm, 2014] M. Thimm. Strategic argumentation in multi-agent systems. *Kuenstliche Intelligenz*, 28:159–168, 2014.

[Timmer *et al.*, 2017] S. Timmer, J.-J.Ch. Meyer, H. Prakken, S. Renooij, and B. Verheij. A two-phase method for extracting explanatory arguments from Bayesian networks. *International Journal of Approximate Reasoning*, 80:475–494, 2017.

[Tolchinsky *et al.*, 2006] P. Tolchinsky, U. Cortes, S. Modgil, F. Caballero, and A. Lopez-Navidad. Increasing human-organ transplant availability: argumentation-based agent deliberation. *IEEE Intelligent Systems*, 21:30–37, 2006.

[Tolchinsky *et al.*, 2012] P. Tolchinsky, S. Modgil, K.D. Atkinson, P. McBurney, and U. Cortes. Deliberation dialogues for reasoning about safety critical actions. *Journal of Autonomous Agents and Multi-Agent Systems*, 25:209–259, 2012.

[Toni, 2014] F. Toni. A tutorial on assumption-based argumentation. *Argument and Computation*, 5:89–117, 2014.

[Toulmin, 1958] S.E. Toulmin. *The Uses of Argument.* Cambridge University Press, Cambridge, 1958.

[Touretzky, 1984] D.S. Touretzky. Implicit ordering of defaults in inheritance systems. In *Proceedings of the 4th National Conference on Artificial Intelligence*, pages 322–325, 1984.

[Touretzky, 1986] David Touretzky. *The Mathematics of Inheritance Systems.* Morgan Kaufmann, 1986.

[van der Weide *et al.*, 2011] T. van der Weide, F. Dignum, J.-J.Ch. Meyer, H. Prakken, and G. Vreeswijk. Arguing about preferences and decisions. In P. McBurney, I. Rahwan, and S. Parsons, editors, *Argumentation in Multi-Agent Systems, 7th International Workshop, ArgMAS 2010, Toronto, Canada, May 2010. Revised Selected and Invited Papers*, number 6614 in Springer Lecture Notes in AI, pages 68–85. Springer Verlag, Berlin, 2011.

[van der Weide, 2011] T.L. van der Weide. *Arguing to Motivate Decisions.* Doctoral dissertation Department of Information and Computing Sciences, Utrecht University, 2011.

[van Veenen and Prakken, 2006] J. van Veenen and H. Prakken. A protocol for arguing about rejections in negotiation. In S. Parsons, N. Maudet, P. Moraitis, and I. Rahwan, editors, *Argumentation in Multi-Agent Systems: Second International Workshop, ArgMAS 2005, Utrecht, Netherlands, July 26, 2005, Revised Selected and Invited Papers*, number 4049 in Springer Lecture Notes in AI, pages 138–153. Springer Verlag, Berlin, 2006.

[Verheij, 1996] B. Verheij. Two approaches to dialectical argumentation: admissible sets and argumentation stages. In *Proceedings of the Eighth Dutch Conference on Artificial Intelligence (NAIC-96)*, pages 357–368, Utrecht, The Netherlands, 1996.

[Verheij, 2003] B. Verheij. DefLog: on the logical interpretation of prima facie justified assumptions. *Journal of Logic and Computation*, 13:319–346, 2003.

[Verheij, 2007] B. Verheij. A labeling approach to the computation of credulous acceptance in argumentation. In *Proceedings of the 2oth International Joint Conference on Artificial Intelligence (IJCAI-07)*, pages 623–628, 2007.

[Vreeswijk and Prakken, 2000] G.A.W. Vreeswijk and H. Prakken. Credulous and sceptical argument games for preferred semantics. In *Proceedings of the 7th European Workshop on Logics in Artificial Intelligence (JELIA'2000)*, number 1919 in Springer Lecture Notes in AI, pages 239–253, Berlin, 2000. Springer Verlag.

[Vreeswijk, 1991] G.A.W. Vreeswijk. The feasibility of defeat in defeasible reasoning. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, pages 526–534, San Mateo, CA, 1991. Morgan Kaufmann Publishers.

[Vreeswijk, 1993a] G.A.W. Vreeswijk. Defeasible dialectics: a controversy-oriented approach towards defeasible argumentation. *Journal of Logic and Computation*, 3:317–334, 1993.

[Vreeswijk, 1993b] G.A.W. Vreeswijk. *Studies in Defeasible Argumentation.* Doctoral dissertation Free University Amsterdam, 1993.

[Vreeswijk, 1997] G.A.W. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90:225–279, 1997.

[Vreeswijk, 2006] G.A.W. Vreeswijk. An algorithm to compute minimally grounded and admissible defence sets in argumentation sytems. In P.E. Dunne and T.B.C. Bench-Capon, editors, *Computational Models of Argument. Proceedings of COMMA 2006*, pages 109–120. IOS Press, Amsterdam etc, 2006.

[Walton and Krabbe, 1995] D.N. Walton and E.C.W. Krabbe. *Commitment in Dialogue. Basic Concepts of Interpersonal Reasoning.* State University of New York Press, Albany, NY, 1995.

[Walton, 1984] D.N. Walton. *Logical dialogue-games and fallacies.* University Press of America, Inc., Lanham, MD, 1984.

[Walton, 1996] D.N. Walton. *Argumentation Schemes for Presumptive Reasoning.* Lawrence Erlbaum Associates, Mahwah, NJ, 1996.

[Walton, 2006a] D.N. Walton. *Fundamentals of Critical Argumentation.* Cambridge University Press, Cambridge, 2006.

[Walton, 2006b] D.N. Walton. Metadialogues for resolving burden of proof disputes. *Argumentation*, 2006. To appear.

[Williams and Williamson, 2006] M. Williams and J. Williamson. Combining argumentation and Bayesian Nets for breast cancer prognosis. *Journal of Logic, Language and Information*, 15:155–178, 2006.

[Woods and Walton, 1978] J. Woods and D.N. Walton. Arresting circles in formal dialogues. *Journal of Philosophical Logic*, 7:73–90, 1978.

[Wooldridge and Parsons, 2000] M. Wooldridge and S. Parsons. Languages for negotiation. In *Proceedings of the Fourteenth European Conference on Artificial Intelligence*, pages 393–400, 2000.

[Wooldridge *et al.*, 2006] M. Wooldridge, P. McBurney, and S. Parsons. On the meta-logic of arguments. In S. Parsons, N. Maudet, P. Moraitis, and I. Rahwan, editors, *Argumentation in Multi-Agent Systems: Second International Workshop, ArgMAS 2005, Utrecht, Netherlands, July 26, 2005, Revised Selected and Invited Papers*, number 4049 in Springer Lecture Notes in AI, pages 42–56. Springer Verlag, Berlin, 2006.

[Wu and Podlaszewski, 2015] Y. Wu and M. Podlaszewski. Implementing crash-resistence and non-interference in logic-based argumentation. *Journal of Logic and Computation*, 25:303–333, 2015.

[Yuan, 2004] T. Yuan. *Human-computer debate, a computational dialectics approach.* PhD Thesis, Leeds Metropolitan University, 2004.

Henry Prakken
Department of Information and Computing Sciences, Utrecht University, The Netherlands
Faculty of Law, University of Groningen, The Netherlands
Email: H.Prakken@uu.nl