Dialectic semantics for argumentation frameworks

H. Jakobovits and D. Vermeir
Dept. of Computer Science
Free University of Brussels, VUB
hadassa@tinf1.vub.ac.be dvermeir@vub.ac.be

Abstract

We provide a formalism for the study of dialogues, where a dialogue is a two-person game, initiated by the proponent who defends a proposed thesis. We examine several different winning criteria and several different dialogue types, where a dialogue type is determined by a set of positions, an attack relation between positions and a legal-move function. We examine two proof theories, where a proof theory is determined by a dialogue type and a winning criterion. For each of the proof theories we supply a corresponding declarative semantics.

1 Introduction

Artificial intelligence has long dealt with the challenge of modeling argumentation ([Tou84], [Fel84], [Vre97]). Abstract argumentation and formal dialectics have been developed in noteworthy works such as [Dun95], [Vre93], [KT96], [PS96], [PS97], [Ver96] and [Lou98a]. These fields are useful for the purpose of decision-making and discussion among intelligent agents, such as in [Ree97] and [PJ98]. In addition, they are important in the context of applications in AI and Law. Indeed, legal argumentation has been modeled, for example, in [McC97], [BC98], [SSK+86], and [HLL94]. Furthermore, a formalization and computational model of civil pleading such as The Pleadings Game ([Gor95]) has been implemented and tested using legal examples. Systems such as HYPO ([AR88]) and CABARET ([SR92]) have been implemented to process legal cases by integrating rules and reasoning with previous cases.

In this paper we model argumentation between two participants, i.e. *dialogue*. A dialogue is initiated by a proponent who proposes a thesis, which she then attempts to defend against any attacks which might come from the opponent. Thus, a dialogue resembles a game, in which the players successively make *moves* according to a set of *rules*, by introducing statements (arguments) or sets of statements (positions). The rules of the game can vary according to need and common consensus. As in any game, it is natural to consider those positions that can be successfully defended. Just as the rules can vary, so can the success criteria. The combination of a set of rules that govern the game, and the determination of winning criteria, constitute a *dialectic semantics* for the "theory" that underlies the player's arguments. Clearly, this notion of "dialogue game" can be regarded as a model for certain types of legal reasoning, as has been convincingly argued, for example, in [Gor95] and [PS96].

The model of argumentation which we consider is the *argumentation framework*, introduced in [Dun95], in which an argument is an abstract entity whose role is determined by its so-called *attack* relations to other arguments. Argumentation frameworks have been studied extensively, and various semantics ([Dun95], [BDKT97], [JV99b]) have been developed. Most of these semantics are formulated in a static, declarative and monological manner, despite the fact that the essence of argumentation is dialogue. In addition, these semantics specify what is admissible, without indicating how such admissible sets of arguments are to to be constructed. This has encouraged many researchers to adopt a procedural approach to argumentation. For example, [Lou98b] reports on the games underlying several programs that can argue. Also, [Lou98a] presents a game which is a model of negotiation. [Gor95] formalizes and implements civil pleading as a dialogue game which captures defeasible reasoning. Legal justification is modeled as a dialog game in [Lod98]. [HLL94] shows that there are different sets of rules that govern particular types of dialogues, and defends a procedural approach to legal reasoning.

In this paper we provide several dialectic proof theories for "winning" positions in argumentation frameworks. They are each given by a dialogue type which is governed by a set of natural rules that determine which moves players can make, and a simple winning criterion. A dialogue type is determined by a set of positions, an attack relation between

positions and a legal-move function.

The following example demonstrates some of the different possibilities which exist for determining these three parameters, and some different winning criteria:

Example 1 *Consider the following hypothetical exchange of allegations, adapted from [Gor95]. The plaintiff (***p***) and the defendant (***o***) have both loaned money to Miller for the purchase of an oil tanker, which is the collateral for both loans. Miller has defaulted on both loans, and the practical question is which of the two lenders will first be paid from the proceeds of the sale of the ship. One subsidiary issue is whether the plaintiff* ***perfected*** *his security interest in the ship or not.*

- *a* **p***: My security interest in Miller's ship was perfected. A security interest in goods may be perfected by taking possession of the collateral (UCC Article 9). I have possession of Miller's ship.*
- *b* **o***: Ships are not goods for the purposes of Article 9.*
- *c* **p***: Ships are movable, and movable things are goods according to UCC Article 9.*
- *d* **o***: According to the Ship Mortgage Act, a security interest in a ship may only be perfected by filing a financing statement.*
- *e* **p***: The Ship Mortgage Act does not apply, since the UCC is newer and therefore has precedence.*
- *f* **o***: The Ship Mortgage Act is federal law, which has precedence over state law such as UCC.*

Depending on how one defines attack between arguments, one possible argumentation framework that can be associated to this discussion is as shown in figure 1.

[Gor95]'s ship example is interesting, because it pinpoints many different issues that have to be addressed when choosing the rules of a dialogue game. For example, notice that the argument d does not attack the argument c which immediately precedes it. Indeed, the opponent gives up his line of discussion, and attempts a different one, by attacking the argument a, which the proponent mentioned earlier. In our framework, we insist that every move attacks its predecessor. This means that a player has only one chance to invalidate the adversary's move. However, in those dialogue types in which we allow multiple arguments to be introduced in one move, several attacks on the preceding move can be made at once so, in fact, this "one-shot" rule is not a restriction.

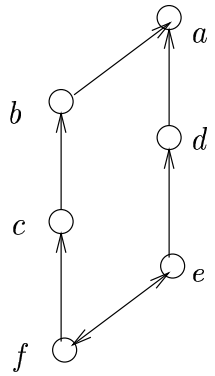

Figure 1: an argumentation framework associated to our adaptation of [Gor95]'s ship example (arrows represent "attack" relationships)

Notice also that the above discussion ends in an infinite loop, since the players have opposing views on the precedence of laws, and there is no further information to support one view over the other. Some dialogue types might allow players to repeat their moves, while some might not. Also, notice that when the argument f is introduced, it is attacked by the argument e, which was introduced earlier. Again, some dialogue types might disallow the use of previously attacked arguments. If the dialogue type does allow the use and repetition of such arguments, then dialogues can get into infinite loops. It is then the role of the winning criterion to determine the status of the position being defended. Other issues in dialogue rules are, for example, whether a player can use an argument which has already been used by the adversary, or an argument which contradicts some other argument which she has already adopted.

We will discuss further aspects of [Gor95]'s ship example in examples 7 and 9.

As a sample in the large range of possibilities which the above example demonstrates, we define two main dialogue types. In the first one, we adopt rules that forbid the use of "self-defeating" arguments, and the use of arguments which have already been invalidated by the adversary. As a result of these restrictions, repetitive use of the same argument is prohibited. In this dialogue type we allow the players to introduce only one argument at a time. Later we generalize this dialogue type to one in which the players can introduce sets of arguments at each move. The attack relation between positions (where a position is a consistent set of arguments) derives trivially from the attack relation between single arguments. The winning criterion in this dialogue game is quite demanding, in that it requires that the proponent win any dialogue in a finite number of moves. In the second main dialogue type which we present, we adopt a more sophisticated attack relation, in which only uncountered attacks are valid. The winning criterion which we adopt for this dialogue game is more credulous than in the first game, in that it accepts positions even if they can only be defended with

dialogues that have infinite loops.

We show that the dialogue types which we introduce can correspond to credulous or sceptical semantics, depending on which winning criterion is used. Legal reasoning is classically associated to sceptical semantics, although there are more and more authors (see, for example, [Lod98]) that emphasize the importance of procedure in legal practice. The basic idea of the procedural approach to argumentation is that a position is acceptable if it can be argued in a reasonable way. Thus, arguments in a framework do not hold because of their absolute truth, but because they can be defended in a convincing manner. In this approach, a change in the procedure can change the resulting semantics, so various credulous semantics can emerge.

Thus, because the rules in our dialogue types are natural and they reflect discussion realistically, the conclusions reached by engaging in our dialogue games are justified. However, there is no guarantee that these conclusions correspond to the set of conclusions which would be reached objectively by a third-party observer who has a global view of the "argumentation framework". Surprisingly, in each case the resulting dialectic semantics can be characterized independently of the game, in a declarative manner. These static semantics are new, but they refer to methods used in common approaches. The fact that, in each case, the proof theory corresponds to a reasonable declarative semantics, adds to its credibility. As expressed in [Fel84], concerning intuitionistic logic, "...if a new development of intuitionistic logic, based on the concepts of dialogues and strategies, shall be given, then one expects an equivalence theorem to be established which states that provability by strategies coincides with provability by one of the known calculi for intuitionistic logic." Similarly, in this paper, we reconcile the procedural approach to argumentation and the objective model-theoretical approach.

One of the added contributions of the semantics presented here is that they solve a problem posed in [Dun95] and [BDKT97], namely, how to satisfactorily deal with arguments that, directly or indirectly, contradict themselves. Indeed, as [Dun95] and [BDKT97] point out, their theories lack facilities to satisfactorily deal with such arguments. Such a contradictory argument, sometimes called a "self-defeating" argument (see [Pol94] and [PS96]), is portrayed in figure 2. [Dun95] and [BDKT97] express the desire to extend their theories in order to deal with this special case. According to [BDKT97], [Dun95], and [Pol94] [1], the only acceptable model of the argumentation framework of figure 2 is the empty set. However, both [Dun95] and [BDKT97] argue that the desired result is that a be rejected, and that b be accepted. Indeed, the argument a is irrational, and should therefore be eliminated. Moreover, a self-defeating argument such as a should be given no power to invalidate other arguments, so b is acceptable. This is the approach adopted in the declarative semantics presented in [KMD94], which attempts to eliminate self-defeating arguments and their effects on other arguments. Our game-theoretical semantics solves the problem of self-defeating arguments.

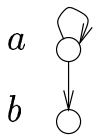

Figure 2: A self-defeating argument

[1] after suitably mapping their approaches to the present framework

The rest of the paper is organized as follows: in section 2 we define argumentation frameworks, their associated *position frameworks*, dialogues, dialogue types, and winning criteria. Section 3 considers the so-called *useful-argument* dialogue type, and its corresponding declarative semantics, which we call the *defensible semantics*. Defensibility is similar to the semantics of [KMD94] in its use of recursion, so our definition of a dialectic proof theory that corresponds to our semantics, contributes also to the understanding of the [KMD94] semantics. In section 4 we present the so-called *rational-extension* dialogue type, and its corresponding *robust semantics*, which is similar to the semantics of [JV99b], but differs in its approach to self-defeating arguments.

2 Basic definitions

We begin by recalling [Dun95]'s model of argumentation, with slightly adapted notations, as follows:

Definition 1 ([Dun95]) *An* **argumentation framework** *is a pair, $AF = (\mathcal{A}, \rightsquigarrow)$, where $\mathcal{A}$ is a set of arguments, and $\rightsquigarrow$ is a binary relation on $\mathcal{A}$, i.e. $\rightsquigarrow \subseteq \mathcal{A} \times \mathcal{A}$. An argument $a \in \mathcal{A}$ is said to* **attack** *an argument $b \in \mathcal{A}$, denoted $a \rightsquigarrow b$, iff $(a, b) \in \rightsquigarrow$. A set S of arguments attacks another set T, denoted $S \rightsquigarrow T$, if there is an argument in S which attacks an argument in T. We write $a \not\rightsquigarrow b$ if a does not attack b. Similarly, if S and T are sets of arguments, then $S \not\rightsquigarrow T$ means that no element from S attacks an element in T. The set $\{a \in \mathcal{A} \mid S \rightsquigarrow a\}$ is denoted $S^{\rightsquigarrow}$, and the set $\{a \in \mathcal{A} \mid a \rightsquigarrow S\}$ is denoted ${}^{\rightsquigarrow}S$. A set S is said to be* **consistent** *if $S \not\rightsquigarrow S$.*

An argumentation framework gives interactions between single arguments. Agents sharing a common framework can adopt *positions* within the framework, which are consistent sets of arguments which they accept. The various positions attack each other according to an attack relation which is based on the attack relation $\rightsquigarrow$ of the original framework AF. This determines a *position framework*, which gives interactions between positions, and which is the framework within which a *dialogue* can take place.

Definition 2 *A* ***position framework*** *corresponding with an argumentation framework $(\mathcal{A}, \rightsquigarrow)$ is an argumentation framework $(\mathcal{P}, \overset{*}{\rightsquigarrow})$ where $\mathcal{P}$ consists of consistent subsets of $\mathcal{A}$.*

*Elements of $\mathcal{P}$ are called **positions**. A **player** is a member of $\{\mathbf{p}, \mathbf{o}\}$ where $\mathbf{p}$ stands for **proponent** while $\mathbf{o}$ stands for **opponent**. For p a player, the **adversary** of p, denoted $\overline{p}$, is defined by $\overline{\mathbf{p}} = \mathbf{o}$ and $\overline{\mathbf{o}} = \mathbf{p}$. A **move in** $\mathcal{P}$ is a pair $[p, X]$ where p is a player and $X \in \mathcal{P}$. For a move $m = [p, X]$, we use $\mathbf{pl}(m)$ to denote p and $\mathbf{pos}(m)$ to denote X.*

In addition to the possible positions and the attack relation between positions, dialogue types are characterized by certain rules that govern the discussion. For example, there might be an agreement that each participant has to be consistent with herself. This consistency requirement, or any other restriction, can be included in the dialogue rules by means of the so-called *legal-move function* in the following definition:

Definition 3 *A **dialogue type** is a tuple $(\mathcal{P}, \overset{*}{\rightsquigarrow}, \phi)$ where $(\mathcal{P}, \overset{*}{\rightsquigarrow})$ is a position framework and $\phi : \mathcal{P}^* \to 2^{\mathcal{P}}$ is a "legal-move" function [2]. A **dialogue** d in $(\mathcal{P}, \overset{*}{\rightsquigarrow}, \phi)$ is any countable sequence[3] $d_0 d_1 ...$ of moves in $\mathcal{P}$ that satisfies [4], for any $i \in \mathbb{N}$,*

1. $\mathbf{pl}(d_{i+1}) = \overline{\mathbf{pl}(d_i)}$, *i.e. the players take turns,*
2. $\mathbf{pos}(d_{i+1}) \in \phi(\mathbf{pos}(d_0) \dots \mathbf{pos}(d_i))$, *i.e. the next move is legal,*
3. $\mathbf{pos}(d_{i+1}) \overset{*}{\rightsquigarrow} \mathbf{pos}(d_i)$, *and it attacks the adversary's last move, and*
4. $\mathbf{pl}(d_0) = \mathbf{p}$, *i.e. the proponent makes the first move.*

*We say that d **is about the position** $\mathbf{pos}(d_0)$.*

Example 2 *[Pra96] considers a game in which the positions are single arguments, and the attack relation (called "defeat" in [Pra96]) between positions is simply the attack relation between arguments in the original framework $AF = (\mathcal{A}, \rightsquigarrow)$. He restricts legal moves by insisting that the proponent cannot repeat herself, and that the proponent must strictly defeat the opponent's argument. In our framework, this game is represented as the dialogue type $(\mathcal{A}', \rightsquigarrow, \phi)$, where $\mathcal{A}' = \{\{a\} \mid a \in \mathcal{A}\}$ and $\phi : \mathcal{A}'^* \to 2^{\mathcal{A}'}$, with $\forall Y_0 \dots Y_i \in \mathcal{A}'^*$,*

$$\phi(Y_0 \dots Y_i) = \begin{cases} \mathcal{A}' & \text{if } i \text{ is even} \\ \mathcal{A}' \backslash (\bigcup_{j=0}^{(i-1)/2} Y_{2j} \cup Y_i^{\rightsquigarrow}) & \text{if } i \text{ is odd.} \end{cases}$$

Indeed, when i is even, it is the opponent's turn. In this case, since there are no restrictions on the opponent, $\phi(Y_0 \dots Y_i) = \mathcal{A}'$. When i is odd, it is the proponent's turn. Since the proponent cannot repeat herself, $\phi(Y_0 \dots Y_i) \cap \bigcup_{j=0}^{(i-1)/2} Y_{2j} = \emptyset$. Since the proponent must strictly *defeat the opponent's previous argument, she cannot use an argument which is attacked by the opponent's previous argument, so $\phi(Y_0 \dots Y_i) \cap Y_i^{\rightsquigarrow} = \emptyset$.*

[2] For a set X we use X^* to denote the set of finite sequences of elements from X.
[3] Note that a dialogue may be infinite.
[4] For any $i \in \mathbb{N}$ we use d_i to denote the $(i+1)$'th element of d.

We provide the following three different winning criteria. A dialogue about a position X can be *won*, there can be a *winning strategy* for X, and/or X can be a so-called *winning* position. Each of these criteria is slightly different and, together with a dialogue type, yields a particular proof theory.

Definition 4 *Let $(\mathcal{P}, \overset{*}{\rightsquigarrow}, \phi)$ be a dialogue type. A dialogue d is **won** by $\mathbf{p}$ if d is finite and ends with a move $[\mathbf{p}, X]$ such that $\overset{*}{\rightsquigarrow} X \cap \phi(d) = \emptyset$, i.e the dialogue cannot be continued. A player p that does not win d is said to have **lost** d.*

Thus, a player wins a game if her adversary has nothing to say. Dialogues which have been won are, according to this definition, finite.

Example 3 *The following is a simplification of a recent case in Flanders. The plaintiff ($\mathbf{p}$) bought a house from a previous owner, and hired a builder ($\mathbf{o}$) to renovate it. During the renovation, the builder knocked down a ceiling and discovered a valuable collection of authentic paintings hidden above the ceiling. The following arguments summarize the discussion:*

- *a* $\mathbf{p}$*: The paintings were found in my property and therefore belong to me. They should be returned to me immediately.*
- *b* $\mathbf{o}$*: It is only because I knocked down the ceiling, and discovered the paintings, that you know of their existence. I found the paintings due to my own work. Since I am the one who found them, I can keep them.*
- *c* $\mathbf{p}$*: When you discovered the paintings you were acting as my employee and manipulating my property. The fruit of your work therefore belongs to me.*
- *d* $\mathbf{o}$*: The paintings are not your property. The previous owner sold you only the house. Its movable contents remain her property.*

At this point in the discussion, it seems obvious that the builder has contradicted herself and therefore lost the argument. Indeed, her claim to ownership contradicts her claim that the previous owner of the house is the rightful owner. This discussion can be modeled with the help of the argumentation framework $AF = (\mathcal{A}, \rightsquigarrow)$, shown in figure 3. Consider the dialogue type $(\mathcal{P}, \rightsquigarrow, \phi)$, where $\mathcal{P}$ is the set of consistent subsets of $\mathcal{A}$, and ϕ is such that a move $[p, X]$ is legal if X is not attacked nor attacks anything that the player p has uttered, i.e. players must be self-consistent. The above discussion corresponds to the dialogue $[\mathbf{p}, \{a\}][\mathbf{o}, \{b\}][\mathbf{p}, \{c\}]$. The proponent wins this game since, the only positions which the opponent can adopt in order to attack the last move, contain $\{d\}$ and therefore are inconsistent with $\{b\}$.

Suppose now that the proponent had anticipated the opponent's "finders-keepers" claim b, and countered it before it was uttered. In this case the proponent would state the arguments a and c right away in the first move. Once the proponent has stated that the builder is her employee, it is useless for the builder to propose the finders-keepers argument

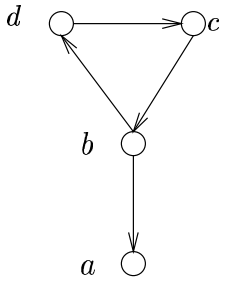

Figure 3:

b. The builder would therefore counter immediately with the argument d. The resulting dialogue is $[\mathbf{p}, \{a, c\}][\mathbf{o}, \{d\}]$. *This dialogue is lost by the proponent, since the only way that she can attack* $\{d\}$ *is to include* $\{b\}$ *in her next position, which is inconsistent with* $\{a, c\}$. *This example shows that it is sometimes in the interest of a player not to "play all of her cards at once" and, instead of revealing all of her possibilities in the beginning, saving some of her arguments for later use. Indeed, if she uses the argument c in the first move, as in the second dialogue, then it is countered and invalidated. If, on the contrary, she saves it for the third move, as in the first dialogue, it serves her to invalidate her adversary's attack and cannot be invalidated.*

The second winning criterion is as follows:

Definition 5 *Let* $D = (\mathcal{P}, \overset{*}{\rightsquigarrow}, \phi)$ *be a dialogue type. A* ***winning strategy*** S *of* D *for a position* $X \in \mathcal{P}$ *is a prefix-closed*[5] *finite set of finite dialogues about* X *in* D *that satisfies*[6]

1. $\forall d \in S, \exists d' > d$ *such that* d' *is won by* $\mathbf{p}$.
2. $\forall d[\mathbf{p}, X] \in S, \forall Y \in {}^{\overset{*}{\rightsquigarrow}}X \cap \phi(d[\mathbf{p}, X])$, $\exists d' > d[\mathbf{p}, X][\mathbf{o}, Y]$ *such that* $d' \in S$.

Essentially, a winning strategy provides a guarantee for the proponent, to be able to counterattack all attacks by the opponent, at every stage of the game. Dialogues about positions for which there are winning strategies, are finite if the proponent makes the wise choices.

The third winning criterion is specified[7] as follows:

Specification 1 *Let* $d = d_1 \ldots d_n$ *be a dialogue in a dialogue type* $D = (\mathcal{P}, \overset{*}{\rightsquigarrow}, \phi)$. d_n *is* **winning** *in* d *if* $\forall d_{n+1}$ *such that* dd_{n+1} *is a dialogue in* D,

- d_{n+1} *is not winning in* dd_{n+1}, *and*
- $\exists d_{n+2}$ *such that* d_{n+2} *is winning in* $dd_{n+1}d_{n+2}$.

A position $X \in \mathcal{P}$ *is* **winning in** D *iff the move* $[\mathbf{p}, X]$ *is winning in the dialogue* $[\mathbf{p}, X]$.

[5] A set X of sequences is prefix-closed if, for any $x \in X$, for any prefix y of x (i.e. for any y such that $yz = x$ for some z), we have $y \in X$.

[6] We write $d_1 > d_2$ to indicate that d_2 is a prefix of d_1. Concatenation of (sequences of) moves is indicated by juxtaposition, as in $d[\mathbf{p}, X][\mathbf{o}, Y]$ which consists of the sequence d followed by the moves $[\mathbf{p}, X]$ and $[\mathbf{o}, Y]$, in that order.

[7] Because of its recursive nature, this is a specification, rather than a definition, as are specifications 2 and 3. Corresponding fixpoint definitions are supplied in [JV99a].

This winning criterion allows positions to be winning, even if all dialogues about them are infinite. This is to allow for credulous reasoning. The following example presents a dialogue type which, when combined with the winning-strategy criterion of definition 5, produces a sceptical semantics and which, when combined with the winning-position criterion of specification 1, produces a credulous semantics. This shows that the same dialogue type, used with different winning criteria, determines different semantics.

Example 4 *A Belgian newspaper publishes a photograph of Gerard Depardieu and Sharon Stone, sharing a private dinner at the famous Brussels restaurant "Comme chez soi". The celebrity couple sues the newspaper for violation of privacy, while the newspaper defends itself on the basis of freedom of public information. Their disagreement is represented in the position framework of figure 4, where a is the*

Figure 4:

position "right to privacy", and b is the position "freedom of public information". Suppose that the proponent, the couple, begins the dialogue by asserting the argument a. Suppose that the rules of their dialogue allow the opponent, the newspaper, to counter this argument by asserting the argument b, and that the rules then allow the proponent to reiterate the argument a. Since the dialogue is infinite, there is no winning strategy for the assertion a, according to definition 5. Similarly, there is no winning strategy for the assertion b. Thus, the winning-strategy criterion produces a sceptical semantics. Furthermore, the proponent does not win the dialogue, according to definition 4. However, the position a is winning, according to specification 1, as is the position b, so the winning-position criterion produces a credulous semantics.

This example demonstrates that a semantics is determined by the combination of a dialogue type and a winning criterion. We have provided three winning criteria. In the next two sections we provide a sample of two specific dialogue types.

3 Useful-argument dialogues

We now formally define the first of the two dialectical proof theories presented in this paper. The proponent begins the game by proposing a position which is a single argument. Each player must then successively introduce a single argument which attacks the previously proposed argument. The rules of the game state that no player can introduce a *self-defeating* argument, i.e. an argument which attacks itself, nor can a player introduce a *useless* argument, i.e. an argument which is attacked by a previously mentioned argument. This motivates the following legal-move function:

Definition 6 *Let $(\mathcal{P}, \overset{*}{\rightsquigarrow})$ be a position framework. The legal-move function $\psi_{(\mathcal{P},\overset{*}{\rightsquigarrow})} : \mathcal{P}^* \rightarrow 2^{\mathcal{P}}$ which allows moves in $(\mathcal{P}, \overset{*}{\rightsquigarrow})$ that are not self-defeating nor useless, is defined as follows: $\forall Y_0 \ldots Y_i \in \mathcal{P}^*$,*

$$\psi_{(\mathcal{P},\overset{*}{\rightsquigarrow})}(Y_0 \ldots Y_i) = \mathcal{P} \backslash (\{X \mid X \overset{*}{\rightsquigarrow} X\} \cup \bigcup_{j=0}^{i} Y_j^{\overset{*}{\rightsquigarrow}}).$$

The formal definition of the proof theory is as follows:

Definition 7 *Let $AF = (\mathcal{A}, \rightsquigarrow)$ be an argumentation framework. A* **useful-single-argument dialogue** *in AF is a dialogue in the dialogue type $(\mathcal{A}', \rightsquigarrow, \psi_{(\mathcal{A}',\rightsquigarrow)})$, where $\mathcal{A}' = \{\{a\} \mid a \in \mathcal{A}\}$, and $\psi_{(\mathcal{A}',\rightsquigarrow)}$ designates moves that are not* self-defeating *nor* useless*.*

Notice that, while a player cannot use useless or self-defeating arguments, she is allowed to contradict something which she said previously. This is due to the fact that the players don't have to care about forward attacks, only backward attacks. When an argument is introduced, they don't look at the things that it attacks. They don't have to "think ahead". In this respect the players are brave, or rather care-free, in that they can propose arguments without thinking of the consequences.

If $\mathcal{A}$ is finite, then any useful-argument dialogue ends after a finite number of moves. As we shall show in theorem 1, the proof theory of definition 7 corresponds to the defensible semantics, defined in specification 2, which is based on the following idea: in a discussion every argument is introduced in order to counter a previous argument. Thus, each argument is associated to the history that led to its introduction. An argument is said to be defensible with respect to its history, if none of its attackers are defensible with respect to their history. This is expressed in specification 2 as a recursive statement. In addition, the restriction on legal moves in a useful-argument dialogue, results in the fact that, no player can reuse an argument which has already been used. (This can be compared to [Pra96]'s game in example 2, in which the proponent cannot reuse her own arguments.) Indeed, if a player could reuse an argument, this would mean that the previous argument was useless. This observation provides the stopping condition in specification 2.

Specification 2 *Let $AF = (\mathcal{A}, \rightsquigarrow)$ be an argumentation framework. $\forall T \subseteq \mathcal{A}$, $\forall a \in \mathcal{A}$,*

- *$def_{AF}(a, T)$ if $a \in T$*
- *$def_{AF}(a, T)$ if $\forall b \rightsquigarrow a$, $\neg def_{AF}(b, \{a\} \cup T)$.*

The argument a is said to be **defensible** *in AF iff $def(a, \emptyset)$*[8].

The variable T plays the role of an accumulator, that holds the arguments that have been considered, i.e. the history. A similar use of an accumulator appears in the [KMD94]-acceptability relation on theories in non-monotonic reasoning frameworks.

[8] When the subscript AF is clear, it is omitted.

The following theorem shows the equivalence of the defensible semantics of specification 2, and the game-theoretical semantics of definition 7.

Theorem 1 *Let $AF = (\mathcal{A}, \rightsquigarrow)$ be an argumentation framework. $\forall a \in \mathcal{A}$, a is a defensible argument in AF iff there is a winning strategy for $\{a\}$ in the useful-single-argument dialogue type $(\mathcal{A}', \rightsquigarrow, \psi_{(\mathcal{A}',\rightsquigarrow)})$*[9].

The following example illustrates theorem 1, and demonstrates that the game-semantics of definition 7 is credulous.

Example 5 *Consider again the discussion of example 4, about the precedence of two principles in law. The proponent asserts the argument a. It is then useless for the opponent to suggest the argument b, since this is contradicted by what the proponent has already said. So, the opponent has no move and the proponent wins.*

Observe that, in keeping with theorem 1, the argument a is defensible, since $def(a, \emptyset)$ if $\neg def(b, \{a\})$ if $def(a, \{a, b\})$, which holds trivially.

The following example illustrates that the game semantics presented here solves the problem of self-defeating arguments, posed in [BDKT97] and [Dun95].

Example 6 *Consider the argumentation framework of figure 3. The proponent states the argument a. The opponent must then state b, which is the only argument that attacks a. The proponent then introduces c. The opponent cannot introduce anything, since the only argument which attacks c is the useless argument d. Thus, the proponent wins. Therefore, there is a winning strategy for $\{a\}$. This is reasonable, since the three arguments b, c and d all indirectly contradict themselves, so they are self-defeating, and therefore should not invalidate the argument a. Note that, in keeping with theorem 1, the argument a is defensible, since $def(a, \emptyset)$ if $\neg def(b, \{a\})$ if $def(c, \{a, b\})$ if $\neg def(d, \{a, b, c\})$ if $def(b, \{a, b, c, d\})$, which holds trivially.*

The following example shows that the useful-single-argument dialogue type can be generalized to a useful-multiple-argument dialogue type, simply by applying it to a position framework, rather than to the argumentation framework itself.

Example 7 *Consider the argumentation framework $AF = (\mathcal{A}, \rightsquigarrow)$ of example 1. The argument a is not defensible in AF since, in order to have $def(a, \emptyset)$ we must have $\neg def(b, \{a\})$, which requires $def(c, \{a, b\})$, which is only true if $\neg def(f, \{c, a, b\})$, which in turn requires $def(e, \{f, c, a, b\})$, which is not the case since $def(f, \{e, f, c, a, b\})$.*

In accordance with theorem 1, then, there is no winning strategy for $\{a\}$ in the dialogue type $D = (\mathcal{A}', \rightsquigarrow, \psi_{(\mathcal{A}',\rightsquigarrow)})$. Indeed, the proponent has no choice but to lose the dialogue $[\mathbf{p}, \{a\}][\mathbf{o}, \{b\}][\mathbf{p}, \{c\}][\mathbf{o}, \{f\}]$, since the only remaining move $[\mathbf{p}, \{e\}]$ is useless.

Notice, however, that if the players agree on a dialogue type which allows multiple arguments to be introduced at

[9] where $\mathcal{A}'$ is as defined in definition 7

each move, then the proponent has a way out. Consider the dialogue type $(\mathcal{P}, \rightsquigarrow, \psi_{(\mathcal{P},\rightsquigarrow)})$, where $\mathcal{P}$ is the set of consistent subsets of $\mathcal{A}$. In this dialogue type, there is no winning strategy for $\{a\}$, since the proponent loses the dialogue $[\mathbf{p}, \{a\}][\mathbf{o}, \{d, f, b\}]$. However, the proponent can immediately propose, as a first move, $[\mathbf{p}, \{a, c, e\}]$. The opponent has no further move, so there is a winning strategy for the position $\{a, c, e\}$ in the dialogue type $(\mathcal{P}, \rightsquigarrow, \psi_{(\mathcal{P},\rightsquigarrow)})$. Theorem 1 can be applied to the argumentation framework $(\mathcal{P}, \rightsquigarrow)$, so the position $\{a, c, e\}$ is defensible in the argumentation framework $(\mathcal{P}, \rightsquigarrow)$. In opposition with example 3, here it is in the interest of the proponent to "play all of her cards at once", in order to state the argument c before the opponent gets a chance to invalidate it with f.

This example motivates the following dialogue type:

Definition 8 *Let $AF = (\mathcal{A}, \rightsquigarrow)$ be an argumentation framework. Let $\mathcal{P}$ be the set of consistent subsets of $\mathcal{A}$. The dialogue type $(\mathcal{P}, \rightsquigarrow, \psi_{(\mathcal{P},\rightsquigarrow)})$ is called the* **useful-multiple-argument** *dialogue type.*

4 Rational-extension dialogues

Example 8 *Consider a case in family law, in which the custody of the baby of a divorced couple must be determined. Imagine the following discussion between* $\mathbf{p}$*, the lawyer of the mother, and* $\mathbf{o}$*, the lawyer of the father:*

- $\mathbf{p}$*: My client, the mother of the child, has the emotional stability and the financial capacity to raise this child. The child is presently being breastfed and should not be separated from her mother. The mother therefore requests custody over the child.*
- $\mathbf{o}$*: Right now the child is small and cannot know what she wants. My client is sure that in a few years she will prefer living with her father.*

In this example the mother's lawyer begins by giving her opinion on the status of those arguments which she considers relevant. The father's lawyer then introduces a further element, the will of the child. At the moment when the judge must decide, the will of the child is unknown. The judge must therefore take a decision which is independent of the will of the child. She might decide to attribute custody to the mother for the first few years and then to re-assess the situation later, when the will of the child can be determined. This decision is reasonable, no matter what the will of the child turns out to be.

This example demonstrates that legal reasoning is typically an evolving process. Legal reasoners are often obliged to take decisions with incomplete information; events which affect the outcome can happen during the decision process; information which is relevant to the case can be made available progressively. It is for this reason that defeasible argumentation, which is concerned with drawing conclusions that might be overridden at a later stage when new information becomes available, is an appropriate paradigm for the analysis of legal reasoning. It is for this reason, too, that when a legal reasoner makes a decision which might be affected by further events, she must choose a position which will remain valid once the further information is available. In this section we introduce a dialogue type for which the winning positions are precisely those positions which are stable under extension of available information.

For this dialogue type, we will constrain positions to be *rational*, which is a stronger (but reasonable) requirement than mere consistency. Rational positions, such as that proposed in the statement made by the mother's lawyer above, give the opinion of the participant in the discussion, on the status of all the relevant arguments. They are defined using so-called *labelings*, as follows:

Definition 9 *Let $AF = (\mathcal{A}, \rightsquigarrow)$ be an argumentation framework. A* **labeling of** *AF is a total mapping*

$$l : A \to 2^{\mathcal{A}} \setminus \{\emptyset\}$$

that satisfies the following conditions:

1. *$\forall a \in \mathcal{A}$, $l(a) \ni +$ iff $\forall b \rightsquigarrow a$, $l(b) \ni -$.*
2. *$\forall a \in \mathcal{A}$, if $l(a) \ni -$ then $\exists b \rightsquigarrow a$ such that $l(b) \ni +$.*

A position $X \subseteq \mathcal{A}$ is **rational** *in AF iff there is a labeling l of AF such that $X \subseteq l^+$.*[10] *In this case l is said to* ***correspond*** *to X. For l_1, l_2 labelings, we say that l_1 is* ***less defined*** *than l_2, denoted $l_1 \sqsubseteq l_2$, iff $\forall a \in \mathcal{A}$, $l_1(a) \supseteq l_2(a)$.*

The intuition behind labelings is extremely simple: "$+$" stands for "support", while "$-$" stands for "doubt". Condition 1 states that an argument has support (i.e. contains a "$+$") iff all of its attackers are doubtful (i.e. contain a "$-$"). Condition 2 states that arguments are "supported by default"; i.e. to cast doubt on an argument, one needs support for one of its attackers. This means that there are no grounds to doubt the validity of an argument unless it is attacked by an argument which has some support. Notice that an argument labeled $\pm$ is both supported and doubtful; i.e. it is undecided. Also, whether a mapping $l : A \to 2^{\{+,-\}}$ is a labeling, is determined "locally": it suffices that each argument be labeled in a way that is consistent with (the label of) its neighbors. Clearly, a labeling then defines a *rational* position, which consists of all arguments that have no doubt. In addition, we observe that any position X has a unique superset, which includes the *necessary conditions* of X, and its *inevitable consequences*. Indeed, the existence of such a minimal rational set including X, called the *completion* of X, is guaranteed by the following:

Theorem 2 *Let $\mathcal{L}_X$ be the set of labelings that correspond to a rational position X in an argumentation framework AF. $(\mathcal{L}_X, \sqsubseteq)$ is a $\wedge$-lattice.*

[10] For a labeling l, $l^+ = \{a \mid l(a) = +\}$. l^- and $l^\pm$ are defined similarly.

Definition 10 *Let $\mathcal{L}_X$ be the set of labelings that correspond to a rational position X in an argumentation framework. The* **least-defined labeling** *which corresponds to X, denoted l_X, is $\wedge\mathcal{L}_X$. The* **rational completion** *of X, denoted X^c, is the set $(l_X)^+$.*

According to [Dun95], a position is admissible iff it counterattacks all attacks against itself. Thus, a real threat on a position is one which is not counterattacked. In the attack relation of our dialogue type, therefore, we consider attacks that are not counterattacked. As for the legal-move function, a player cannot adopt any argument which has already been used in a previous move, nor can she use any argument which is attacked by a previous move, since it is already invalidated and therefore useless. These considerations motivate the following dialogue type:

Definition 11 *Let $AF = (\mathcal{A}, \rightsquigarrow)$ be an argumentation framework. A* **rational-extension dialogue** *in AF is a dialogue in the dialogue type $(\mathcal{R}, \rightharpoonup, \gamma)$, where*

- *$\mathcal{R}$ designates* rational completions *in AF, i.e.*

$$\mathcal{R} \equiv \bigcup_{X \in 2^{\mathcal{A}}} \{Y^c \mid Y \text{ is a rational position in } AF\,|_X\}$$[11]

- *$\rightharpoonup$ designates* uncountered attacks, *i.e.* $\forall X, Y \in \mathcal{R}$,

$$Y \rightharpoonup X \text{ iff } \exists S \subseteq Y \text{ such that } S \rightsquigarrow X \wedge X \not\rightsquigarrow S.$$

- *γ designates moves that are not* repeated *nor* useless, *i.e.*

$$\gamma : \mathcal{R}^* \to 2^{\mathcal{R}}$$
$$\forall X_0 \ldots X_i \in \mathcal{R}^*, \gamma(X_0 \ldots X_i) = \mathcal{R} \setminus \bigcup_{j=0}^{i} (X_j \cup X_j^{\rightharpoonup})$$

As we shall show in theorem 3, the semantics which results from this proof theory is the *robust semantics*. The idea of the robust semantics is to determine those rational positions that respect the stability of the decided arguments. When a rational position is proposed, the so-called "undecided part" of the argumentation framework, consists of the arguments which have not yet been decided (i.e. the arguments in $l^{\pm}$, where l is the corresponding labeling), with their interactions. The information contained in this restricted argumentation framework may permit further conclusions to be added to the decided arguments (i.e. the arguments in $l^+ \cup l^-$). These additional decisions are expressed by a labeling of the undecided part of the argumentation framework. In this case, the original set of arguments may or may not be compatible with the newly added conclusions, in that their combination may or may not correspond to a labeling. We shall say that a labeling l is *robust*, if the decided arguments can remain unchanged, when some undecided arguments become decided. These notions are formalized in the following definition. The notion of an *extension* conveys the idea of extending a present knowledge state, by adding further conclusions which concern undecided arguments. The notion of *compatibility* is a criterion which determines whether the combination of two successive sets of decisions forms one coherent set of decisions.

Definition 12 *Let l be a labeling of an argumentation framework $AF = (A, \rightsquigarrow)$. A labeling l' of $AF\,|_{l^{\pm}} = (l^{\pm}, \rightsquigarrow|_{l^{\pm} \times l^{\pm}})$ is said to be an* **extension** *of l. The labeling l is said to be* **compatible** *with l' iff $l \dagger l'$ is a labeling of AF, where*

$$l \dagger l'(a) = \begin{cases} l'(a) & \text{if } a \in l^{\pm} \\ l(a) & \text{otherwise} \end{cases}$$

In order for a set of decisions to be considered "valid", it must be compatible with any "valid" extension of itself. Thus, any extension which is not compatible must be itself invalidated by a further extension which is valid. This motivates the following specification for robust labelings:

Specification 3 *Let AF be an argumentation framework.*

- *A labeling l of AF is* **robust** *iff*
  1. *it is compatible with all robust extensions of itself, and*
  2. *any incompatible extension l' of l has a robust extension which is incompatible with l'.*
- *A set $T \subseteq \mathcal{A}$ of arguments is said to be* **robust** *iff it corresponds to a robust labeling of AF.*

This specifies the robust semantics, which corresponds to the proof theory given by the rational-extension dialogue type, as follows:

Theorem 3 *Let X be a rational position in an argumentation framework. X is a winning position in the rational-extension dialogue type $(\mathcal{R}, \rightharpoonup, \gamma)$, iff l_X is robust.*

Example 9 *Consider again the argumentation framework of example 1. Recall that the discussion ends in an infinite loop, since the players have opposing views on the precedence of laws, and there is no further information to support one view over the other. Suppose that the proponent would like to defend her position $\{a, c\}$, even though there is no way to resolve the conflict over law precedence. This is equivalent to adopting the position which corresponds to the labeling l shown in figure 5, in which the arguments d, e and f are left undecided. This labeling is not robust, since it is incompatible with the labeling l' shown in the figure. This means that if it is later revealed that the Ship Mortgage Act has precedence over the UCC, the proponent's claim will be incompatible with this new information. By theorem 3, then, the move $[\mathbf{p}, \{a, c\}]$ is not winning in the rational-extension dialogue type. This is indeed the case, since the dialogue $d = d_1 d_2 = [\mathbf{p}, \{a, c\}][\mathbf{o}, \{d, f\}]$ has no further move, so it*

[11] $AF\,|_X$ is the argumentation framework $(X, \rightsquigarrow|_{X \times X})$.

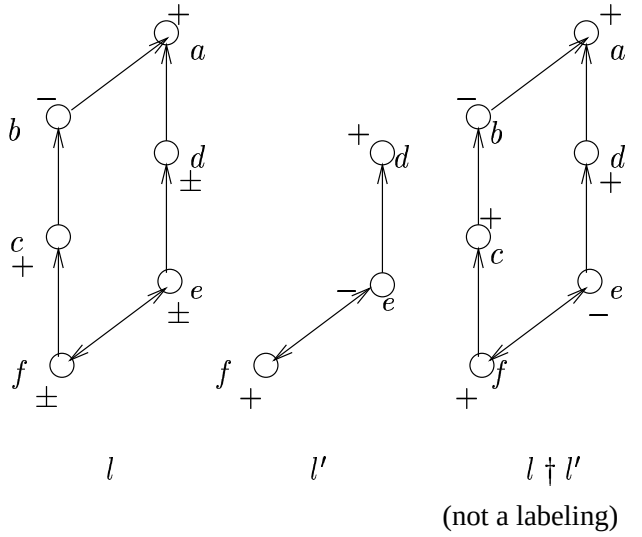

Figure 5: l is not robust

is lost by the proponent. The move d_2 is therefore winning in $d_1 d_2$, so d_1 is not winning in d_1 and therefore not winning in the rational-extension dialogue type.

Notice that, although $[\mathbf{p}, \{a, c\}]$ is not winning, this does not mean that $\mathbf{p}$ loses every dialogue about $\{a, c\}$. Indeed, the dialogue $[\mathbf{p}, \{a, c\}][\mathbf{o}, \{d\}][\mathbf{p}, \{e\}]$ is won by the proponent. This shows that in order to act wisely, the opponent has to state f right away, before it becomes invalidated by the argument e in the proponent's next move. In opposition to example 3 and as was the case in example 7, in order to win the proponent has to "play her cards quickly". Similarly, if the proponent starts with the move $[\mathbf{p}, \{a, c, e\}]$, the opponent has no move, so the position $\{a, c, e\}$ is winning and its corresponding labeling is robust.

5 Directions for further research

In this work several sample dialogue types have been presented. Other dialogue types, defined using similar considerations, can be formulated, with the objective of understanding how the parameters of the dialogue type affect the set of conclusions that can be proven.

One particular legal-move function which we are examining is the self-consistency function. Indeed, in [JV99a] we introduce a so-called *self-consistency* dialogue type in which the set of arguments uttered by any player has to be consistent. Unlike useful-argument dialogues, in which players only consider backward attacks, self-consistency dialogues require players to consider both backward and forward attacks of the positions which they adopt. It seems that the positions that can be successfully argued, correspond to the intersection of all maximal admissible sets.

Furthermore, procedural models for argumentation, such as the one presented in this paper, are appropriate for capturing defeasibility in argumentation, and allow the addition of new information. An argumentation system, because of its nature, cannot be deterministic or complete since the acceptance of certain arguments invites the introduction of further arguments. We therefore would like to abandon the assumption that the available arguments are known before the discussion. This suggests defining the argumentation framework during the dialogue. We would like to determine the resulting semantics of an argumentation framework which is defined *by means of* a discussion.

As has been demonstrated in [PS96], legal reasoning can be modeled as a logical system for defeasible argumentation, with a logic-programming-like language. An application of the present work is to generate, e.g. with the help of the mapping from argumentation frameworks to logic programs, provided in [JV96], a dialectic proof theory for semantics of logic programs which is appropriate for legal reasoning.

In addition, a computational model of our dialogue games could be implemented and tested using examples from legal reasoning.

6 Acknowledgements

The authors are grateful to an anonymous referee for useful comments.

References

[AR88] K. D. Ashley and E. L. Rissland. A case-based approach to modelling legal expertise. *IEEE Expert*, 3(3):70–77, 1988.

[BC98] T. Bench-Capon. Specification and implementation of toulmin dialogue game. In J. C. Hage, T.J. M. Bench-Capon, A.W. Koers, C.N.J. de Vey Mestdagh, and C. A. F. M. Grutters, editors, *Proceedings of the Eleventh International Conference on Legal Knowledge-Based Systems (Jurix)*, December 1998.

[BDKT97] A. Bondarenko, P.M. Dung, R. A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93(1–2):63–101, 1997.

[Dun95] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[Fel84] W. Felscher. Dialogues as a foundation for intuitionistic logic. In D. Gabbay and F. Guenther, editors, *Handbook of Philospohical Logic III*, pages 341–372. D. Reidel, 1984.

[Gor95] Thomas F. Gordon. *The Pleadings Game: an Artificial Intelligence Model of Procedural Justice*. Kluwer, 1995.

[HLL94] Jaap C. Haag, Ronald Leenes, and Arno R. Lodder. Hard cases: a procedural approach. *Artificial Intelligence and Law*, 2:113–167, 1994.

[JV96] H. Jakobovits and D. Vermeir. Contradiction in argumentation frameworks. In *Proceedings of the IPMU conference*, pages 821–826, 1996.

[JV99a] H. Jakobovits and D. Vermeir. Dialectic vs. static semantics for argumentation frameworks. in preparation, 1999.

[JV99b] H. Jakobovits and D. Vermeir. Robust semantics for argumentation frameworks. *Journal of Logic and Computation*, 6(2):215–261, 1999.

[KMD94] A. C. Kakas, P. Mancarella, and Phan Minh Dung. The acceptability semantics for logic programs. In P. Van Hentenrijck, editor, *Proceedings of the 11th International Conference on Logic Programming*, pages 504–519. MIT Press, 1994.

[KT96] R. Kowalski and F. Toni. Abstract argumentation. *Artificial Intelligence and Law Journal, Special Issue on Logical Models of Argumentation*, 4:275–296, 1996. reprinted in H. Prakken and G. Sartor (eds.), Logical Models of Legal Argument. Dordrecht: Kluwer Academic Publishers, pp. 119-140.

[Lod98] A. Lodder. Dialaw: On legal justification and dialog games. Phd thesis, Universiteit Maastricht, Department of Metajuridica, 1998.

[Lou98a] Ronald P. Loui. A better negotiation game. *Negotiation Journal*, 1998. submitted.

[Lou98b] Ronald P. Loui. Process and policy: resource-bounded non-demonstrative argument. *Computational Intelligence*, 1:92–43, 1998.

[McC97] L.T. McCarty. Some arguments about legal arguments. In *Proceedings of the Sixth International Conference on Artificial Intelligence and Law*, July 1997.

[PJ98] S. Parsons and N. R. Jennings. Argumentation and multi-agent decision making. In *AAAI Spring Symposium on Mixed-Initiative Decision Theoretic Systems*, Stanford, 1998.

[Pol94] John Pollock. Justification and defeat. *Artificial Intelligence*, 67:377–407, 1994.

[Pra96] H. Prakken. Dialectical proof theory for defeasible argumentation with defeasible priorities. In *Proceedings of the biannual International Conference on Formal and Applied Practical Reasoning (FAPR) workshop*, 1996. available at http://nathan.gmd.de/projects/zeno/fapr/programme.html.

[PS96] H. Prakken and G. Sartor. A dialectical model of assessing conflicting arguments in legal reasoning. *Artificial Intelligence and Law*, 4:331–368, 1996.

[PS97] H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. *Applied Non-Classical Logics*, 7:25–75, 1997. special issue on 'Handling inconsistency in knowledge systems'.

[Ree97] C.A. Reed. Argumentation theory and multi-agent systems. Notes for an invited seminar presented at QMW, April 1997.

[SR92] D. B. Skalak and E. L. Rissland. Arguments and cases: an inevitable intertwining. *Artificial Intelligence and Law*, 1:3–44, 1992.

[SSK+86] M. Sergot, F. Sadri, R. Kowalski, F. Kriwaczek, P. Hammond, and H. Cory. The british nationality act as a logic program. *Communications of the Association for Computing Machinery*, 19(5):370–386, 1986.

[Tou84] S. Toulmin. *The Uses of Arguments*. Cambridge University Press, Cambridge, MA, 1984.

[Ver96] B. Verheij. Two approaches to dialectical argumentation: admissible sets and argumentation stages. In *Proceedings of the biannual International Conference on Formal and Applied Practical Reasoning (FAPR) workshop*, 1996. available at http://nathan.gmd.de/projects/zeno/fapr/programme.html.

[Vre93] G. Vreeswijk. Defeasible dialectics: A controversy-oriented approach towards defeasible argumentation. *Journal of Logic and Computation*, 3(3):317–334, June 1993.

[Vre97] Gerard A.W. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90(1-2):225–279, 1997.