A Logic of Abstract Argumentation

Guido Boella[1], Joris Hulstijn[2], and Leendert van der Torre[3]

[1] Universitá di Torino
[2] Vrije Universiteit, Amsterdam
[3] CWI Amsterdam and Delft University of Technology

Abstract. In this paper we introduce a logic of abstract argumentation capturing Dung's theory of abstract argumentation, based on connectives for attack and defend. We extend it to a modal logic of abstract argumentation to generalize Dung's theory and define variants of it. Moreover, we use the logic to relate Dung's theory of abstract argumentation to more traditional conditional and comparative formalisms, and we illustrate how to reason about arguments in meta-argumentation.

1 Introduction

Dung's theory of abstract argumentation [7] is popular in agent theory. For example, Prakken and Vreeswijk note that on the one hand it unifies theories on argumentation [14], and on the other hand it unifies theories of non-monotonic reasoning [6]. However, it has also been criticized. For example, Horty observes that the pattern called reinstatement is an integrated part of Dung's theory, whereas this pattern has been criticized in non-monotonic reasoning [9]. In multiagent systems argumentation theory is used for dialogue, for example by Parsons *et al* [12], because there is no commonly known truth to refer to. In other words, argumentation is all there is to establish agreement. This is analogous to the situation in legal reasoning.

However, argumentation theory is hardly used in agent technology. For example, it is not used for model checking agent dialogues [17]. There are two related problems. First, various researchers have claimed that the model-theoretic approach is not suitable for argumentation, such that the model-theoretic approach to argumentation has not been developed. Secondly, the relation between argumentation theory and other formal systems has not been studied, such that that existing technologies cannot be used for argumentation. There is a tendency within argumentation theory to develop specialized procedures rather than to connect to existing technologies.

In this paper we study argumentation theory from a logical point of view, using model-theoretic semantics, and we relate it to other formal systems. From a formal point of view, Dung does not consider conditionals used in traditional argumentation and non-monotonic reasoning, such as for example $a \rightarrow b$: a is an argument for (supports) b. Instead, the central concept studied in abstract argumentation is a binary *attack* relation among arguments. In this paper we represent it by '$\rhd$'. We write $a \rhd b$ for argument a attacks argument b.

Though both $a \rightarrow b$ and $a \rhd b$ are binary connectives, they have distinct logics. For example, whereas most conditional logics accept the identity rule, $a \rightarrow a$, we definitely do not have that all arguments attack itself: $a \not\rhd a$. Moreover, whereas a conditional connective '$\rightarrow$' might satisfy the transitivity or the cut rule, this does not make sense for the

S. Parsons et al. (Eds.): ArgMAS 2005, LNAI 4049, pp. 29–41, 2006.

attack connective '$\rhd$'. The latter also distinguishes the attack connective from binary comparatives like preference connectives, e.g., $p > q$ for 'p is preferred to q' [16].

Despite the popularity of Dung's framework, it seems that the logical relations among attack statements have not been studied yet. Moreover, in abstract argumentation the notion of a set of arguments defending another argument has been defined. Again the logical relations among defend statements, and their relation to attack statements, seems unexplored. However, such an analysis would be useful for several reasons. It would give insight in Dung's abstract argumentation, it would be a basis for generalizations of Dung's theory, it would enable a comparison with other formalisms, and it would support reasoning about arguments in meta-argumentation [18]. We therefore raise the following questions in this paper:

1. What is a logic for abstract argumentation?
2. What are logical properties of abstract argumentation?
3. How to use such a logic to generalize Dung?
4. How is it related to conditional and preference logics?
5. How can agents reason about arguments?

Following Besnard and Doutre [1], to study these questions we represent arguments by propositions, such that "argument a together with argument b attacks argument c" is represented by $a \wedge b \rhd c$. We introduce connective '$\oslash$' for defence, so "argument a defends argument b" is represented by $a \oslash b$. Moreover, in the relation with conditional logic, we pursue the intuition that "argument a attacks argument b" is related to "if a holds then b does not hold". We relate, e.g.,

– if a attacks b and c defends b, then c attacks a,

to the following inference:

– from $a \rightarrow_{\rhd} \neg b$ and $c \rightarrow_{\oslash} b$, derive $c \rightarrow_{\rhd} \neg a$.

At present, this relation is not only unknown, but the question could not be raised, because there was no language in which it could be expressed.

Finally, reasoning about arguments in meta-argumentation is illustrated by the following dialogue:

A: I think arguments a and b defend argument c.
B: But argument d attacks argument c!
A: No problem, since argument a attacks argument d.

This dialogue illustrates how our logic contributes also to traditional argumentation theory.

The layout of this paper follows the research questions. In Section 2 we introduce a logical framework to reason about abstract argumentation. In Section 3 we consider logical properties among attack and defend statements, and in Section 4 we introduce a modal generalization of the logic to define variants and generalizations of Dung's theory. In Section 5 we consider the relation between the logic and more traditional formalisms, and in Section 6 we consider reasoning about arguments.

2 Semantics

We start with Dung's theory of argumentation. It is nowadays called a theory of abstract argumentation, because it ignores the internal structure of arguments. Here we use the presentation of Besnard and Doutre [1], who in contrast to Dung also define sets of arguments attacking other sets of arguments. Moreover, they write "argument system" where Dung writes "argument framework".

Definition 1 (Argument System). *An argument system is a pair $\langle A, R\rangle$, where A is a set (of arguments), and R is a binary relation over A which represents a notion of* attack *between arguments ($R \subseteq A \times A$). Given two arguments a and b, $(a, b) \in R$ means that a attacks b or that a is an attacker of b. A set of arguments S attacks an argument a if a is attacked by an argument of S. A set of arguments S attacks a set of arguments S' if there is an argument $a \in S$ which attacks an argument $b \in S'$.*

Dung assumes an argument system $\langle A, R\rangle$ to be given. Moreover, he gives several semantics which produce none, one or several sets of acceptable arguments called extensions. Most of these semantics depend on an additional notion of what is nowadays called defence. Instead of "S defends a", Dung says "a is acceptable with respect to S." We also define a set of arguments defending another set of arguments.

Definition 2 (Argument Semantics). *Let $\langle A, R\rangle$ be an argument system.*

- *$S \subseteq A$ is conflict free iff there are no a and b in S such that a attacks b.*
- *A conflict free set $S \subseteq A$ is a stable extension iff for each argument which is not in S, there exists an argument in S that attacks it.*
- *An argument $a \in A$ is defended by a set $S \subseteq A$ (or S defends a) iff for any argument $b \in A$, if b attacks a, then S attacks b.*
- *A conflict free set $S \subseteq A$ is admissible iff each argument in S is defended by S.*
- *A preferred extension is an admissible subset of A, which is maximal w.r.t. set inclusion.*
- *An admissible $S \subseteq A$ is a complete extension iff each argument which is defended by S is in S.*
- *The least (with respect to set inclusion) complete extension is the grounded extension.*

We say that $S \subseteq A$ defends $S' \subseteq A$ iff S defends each $a \in S'$.

The basic idea of the logic of abstract argumentation is that there is no longer a fixed argument system, in the following sense. A model of the logic represents an argument system, and such a model satisfies formulas representing that arguments attack or defend each other, or whether sets of arguments are extensions. Now, a formula is a theorem if it holds in all models, i.e., when it is true for every argument system. Theorems thus quantify over argument systems.

There are many ways to design a logic of abstract argumentation. In this section we stay close to Dung's argument system, and we generalize it in Section 4. We first assume a fixed signature or alphabet, which consists of the set of arguments A. L_0 is the set of conjunctions of atoms, representing sets of arguments, and L is the language

that contains the notions of Dung's theory of argumentation. L_1 is the fragment of L that contains only the attack and defend connectives. Note that modalities in L cannot be nested.

Definition 3 (LAA language). *Given a set of arguments $A = \{a_1, \ldots, a_n\}$, we define the set L_0 of argument sets and the set L of LAA formulas as follows.*

L_0**:** $a_i \mid p \wedge q$ $\quad (p, q \in L_0)$
L**:** $(p \rhd q) \mid (p \oslash q) \mid F(p) \mid S(p) \mid A(p) \mid P(p) \mid C(p) \mid G(p) \mid \neg\phi \mid (\phi \wedge \psi)$
$(p, q \in L_0; \phi, \psi \in L)$

We write L_1 for the fragment of L that does not contain a monadic modal operator. Moreover, disjunction $\vee$, material implication $\supset$ and equivalence $\leftrightarrow$ are defined as usual. We abbreviate formulas using the the following order on logical connectives: $\neg \mid \vee, \wedge \mid \rhd, \oslash \mid \supset, \leftrightarrow$. For example, $\neg p \rhd q \wedge r$ is short for $(\neg p \rhd (q \wedge r))$.

A semantic structure just consists of the binary attack relation R.

Definition 4 (LAA semantics). *Let A be set of arguments, let p and q be elements of L_0 and let ϕ and ψ be elements of L, and let R be a binary relation over A. We have:*

$R \models p \rhd q$ *iff in argument system $\langle A, R\rangle$, the set of arguments in p attack the set of arguments in q.*
$R \models p \oslash q$ *iff in argument system $\langle A, R\rangle$, the set of arguments in p defend the set of arguments in q.*
$R \models F(p)$ *iff the set of arguments in p is conflict free in argument system $\langle A, R\rangle$.*
$R \models S(p)$ *iff the set of arguments in p is a stable extension in argument system $\langle A, R\rangle$.*
$R \models A(p)$ *iff the set of arguments in p is admissable in argument system $\langle A, R\rangle$.*
$R \models P(p)$ *iff the set of arguments in p is a preferred extension in arg. system $\langle A, R\rangle$.*
$R \models C(p)$ *iff the set of arguments in p is a complete extension in arg. system $\langle A, R\rangle$.*
$R \models G(p)$ *iff the set of arguments in p is a grounded extension in arg. system $\langle A, R\rangle$.*
$R \models \neg\phi$ *iff not $R \models \phi$.*
$R \models \phi \wedge \psi$ *iff $R \models \phi$ and $R \models \psi$.*

Moreover, logical notions are defined as usual, in particular:

- $R \models \{\phi_1, \ldots, \phi_n\}$ *iff* $R \models \phi_i$ *for* $1 \leq i \leq n$,
- $\models \phi$ *iff for all R, we have $R \models \phi$,*
- $S \models \phi$ *iff for all R such that $R \models S$, we have $R \models \phi$.*

In this paper we are in particular interested in logic L_1 that only contains the attack and defend connectives, which constitute the basis of Dung's theory. We believe that to understand Dung's theory, one has first to better understand these two binary connectives.

Example 1. If a attacks b and c defends b, then c attacks a,

- $\models (a \rhd b) \wedge (c \oslash b) \supset (c \rhd a)$.

3 Logical Properties

The logical relations among attack formulas are characterized by the left (LD) and right distribution (RD) properties. They follow from the definition of attack among sets of arguments in terms of attacks among individual arguments. To understand this characterization we consider two logical consequences. First, logical consequences of the distribution properties (read from right to left) are left (LS) and right strengthening (RS). Right strengthening indicates that the attack connective does not behave like a conditional connective, but it behaves in this respect like a comparative connective (see Section 5 for details). Secondly, the more remarkable logical consequences of the distribution properties (read from left to right) is that if two arguments together attack another argument, then one of these arguments individually attacks the other argument (LT and RT). These splitting properties indicate room for generalizing Dung's theory (see Section 4).

LD	$\models (a \wedge b \rhd c) \leftrightarrow (a \rhd c) \vee (b \rhd c)$	LA	$\models (a \oslash c) \vee (b \oslash c) \supset (a \wedge b \oslash c)$
RD	$\models (a \rhd b \wedge c) \leftrightarrow (a \rhd b) \vee (a \rhd c)$	RD	$\models (a \oslash b \wedge c) \leftrightarrow (a \oslash b) \wedge (a \oslash c)$
LS	$\models (a \rhd c) \supset (a \wedge b \rhd c)$	LS	$\models (a \oslash c) \supset (a \wedge b \oslash c)$
RS	$\models (a \rhd b) \supset (a \rhd b \wedge c)$	RW	$\models (a \oslash b \wedge c) \supset (a \oslash b)$
LT	$\models (a \wedge b \rhd c) \supset (a \rhd c) \vee (b \rhd c)$	RC	$\models (a \oslash b) \wedge (a \oslash c) \supset (a \oslash b \wedge c)$
RT	$\models (a \rhd b \wedge c) \supset (a \rhd b) \vee (a \rhd c)$		

The logical relations among defend relations are characterized by left additivity (LA) and right distribution (RD) properties. These properties follow from the definition of defend among sets of arguments in terms of attacks among individual arguments. The first logical consequences of these two properties (read from left to right) are left strengthening (LS) and right weakening (RW). Right weakening indicates that the defend connective behaves like a conditional connective (see Section 5 for details). Secondly, we have the conjunction property RC (read from right to left).

The relation among attack and defence connectives is as follows. If a set of arguments is finite, we can simply define the defend connective in terms of attack connective.

- $(a \oslash b) \leftrightarrow \bigwedge_{c \in A}((c \rhd b) \supset (a \rhd c))$

An instance of this relation, which characterizes the infinite case, is the following property already observed in Example 1. It says that the only possible defence is a direct counterattack, and thus rules out other defence tactics. This may seem counterintuitive at first sight, but it makes Dung's system effective.

- $(a \oslash b) \wedge (c \rhd b) \supset (a \rhd c)$

Though we are primarily interested in the logic L_1, the following example illustrates that the logic can be used to express well known relations among extensions.

Example 2. A stable extension is also a preferred extension, and a preferred extension is also a complete extension.

- $\models S(p) \supset P(p)$
- $\models P(p) \supset C(p)$

Two important properties are the expressive power of the language, and compactness of the logic.

Proposition 1 (Expressive power). *The logical language is expressive enough to distinguish two distinct argumentation theories based on the same set of arguments.*

Proof. If two argumentation systems are distinct, then there are two arguments a and b such that $R_1(a, b)$ holds in one argument system $\langle A, R_1\rangle$, but $R_2(a, b)$ does not hold in the other $\langle A, R_2\rangle$ – or vice versa. Then we have $R_1 \models a \rhd b$, but not $R_2 \models a \rhd b$ – or vice versa.

Proposition 2 (Compactness). *The logic is not compact, when the set of arguments A is infinite.*

Proof. Follows directly from universal quantification in the definition of the semantics. For example, assume that A is infinite. We can derive that argument a defends argument b when there is an infinite set of formulas for each argument $c \in A$ that either a attacks c or c does not attack b. However, we cannot derive that a defends b from a finite set of formulas.

A non-monotonic extension can be defined based on distinguished models and subset minimal attack relations. Sometimes such distinguished models are called preferred models and non-monotonic entailment is called preferential entailment.

Definition 5. *A model R is a distinguished model of a set of sentences S iff*

1. *$R \models S$, and*
2. *there is no $R' \subset R$ such that $R' \models S$.*

Nonmonotonic entailment is defined as usual:

- *$T \mid\!\sim \phi$ iff for all distinguished models R of T we have $R \models \phi$.*

The typical use of our logic is when an argument system is specified by a set of attack statements; we call such a set an argument specification.

Proposition 3. *An argument specification is a set of attack formulas $AS = \{p_1 \rhd q_1, \ldots, p_n \rhd q_n\}$. The distinguished model of an argument specification AS is unique.*

There are some limitations to the logic proposed here. First, the semantics leave little room for generalizations of Dung's theory. Secondly, we cannot express the characterizations in propositional logic provided by Besnard and Doutre. Thirdly, we cannot express that stable, preferred and complete semantics admit multiple extensions whereas the grounded semantics ascribes a single extension to a given argument system. In the following section we therefore discuss an extension of LAA in modal logic.

4 Modal Logic of Abstract Argumentation

To define variants and generalizations of Dung's theory, we now generalize LAA in a modal logic setting. We restrict ourselves to finite sets of arguments. Since sentences of

the logic are finite, we cannot represent and reason about infinite extensions. The logic therefore seems most suitable for finite argument systems.

Our generalization is based on an attack relation between sets of arguments. Such sets of arguments are called positions and represented in the semantics of the logic by worlds in a possible worlds model. The attack relation is thus a binary relation between worlds, that is, a standard accessibility relation of possible worlds semantics.

Our motivation is that Dung's assumption that the attack relation exists between individuals arguments instead of sets of arguments is quite strong, and that it is not warranted in cases where the cumulative weight of arguments is decisive [15, 2]. For example, in some legal cases circumstantial evidence may be used in a cumulative way. Each piece of evidence individually would not be enough to connect a suspect to the crime scene, but many pieces of evidence taken together would be enough to conclude that the suspect was present at the crime scene. So only a set of arguments taken together would attack a position in this case.

Formally, we define a normal bimodal semantics in which modal operator $\Box_1$ represents the attack relation, and $\Box_2$ is a universal modality used for technical reasons. Since we have right strengthening for attack connectives where normal modal operators have right weakening, we use a negation in the definition of the attack connective. Propositional formulas represent positions, i.e., sets of arguments. The logic also has negations and disjunctions in the left and right hand side of our connectives, but we do not use this in this paper. We adapt the definition of defend in terms of attack to deal with our generalized setting (s represents a set of atoms as well as a conjunction of atoms).

Definition 6 (MLAA language). *Given a set of arguments $A = \{a_1, \ldots, a_n\}$, we define the set ML of MLAA formulas as follows.*

ML**:** $a_i \mid \Box_1(\phi) \mid \Box_2(\phi) \mid F(\phi) \mid S(\phi) \mid A(\phi) \mid P(\phi) \mid C(\phi) \mid G(\phi) \mid \neg\phi \mid (\phi \wedge \psi)$ *$(\phi, \psi \in ML)$.*

We write ML_1 for the fragment of ML that contains only monadic modal operators $\Box_1$ and $\Box_2$. Moreover, disjunction $\vee$, material implication $\supset$ and equivalence $\leftrightarrow$ are defined as usual. We extend the modal logic with the definition:

- $p \rhd q = \Box_2(p \supset \Box_1 \neg q)$
- $p \oslash q = \bigwedge_{s \subseteq A}(s \rhd q \supset p \rhd s)$

We abbreviate formulas using the the following order on logical connectives: $\neg \mid \vee, \wedge \mid \rhd, \oslash \mid \supset, \leftrightarrow$.

For space reasons we only introduce a semantics for ML_1. The other modalities can be described by a non-normal modal semantics only, as they do not satisfy weakening nor strengthening. From a logical point of view, MLAA is a standard normal modal logic with universal relation. Complexity and axiomatization of this logic are well known, see for example [8].

Definition 7 (MLAA semantics). *Let A be a set of arguments. A possible worlds model M is a structure $\langle W, R, V\rangle$ where W is a set (of worlds), R is a binary (attack) relation on W, and V is a valuation function which assigns a subset of A to each element of W.*

- $M, w \models a$ *iff* $a \in V(w)$ *for all arguments* $a \in A$
- $M, w \models \neg\phi$ *iff not* $M, w \models \phi$
- $M, w \models \phi \wedge \psi$ *iff* $M \models \phi$ *and* $M \models \psi$
- $M, w \models \Box_1\phi$ *iff for all* w' *such that* $R(w, w')$ *we have* $M, w' \models \phi$.
- $M, w \models \Box_2\phi$ *if for all* $w \in W$, *we have* $M, w \models \phi$.

We assume that W *contains exactly one world for each subset of* A.

Clearly, the language L is a fragment of ML. Moreover, Dung's theory can be characterized by the properties we already discussed:

- $\models (a \wedge b \rhd c) \leftrightarrow (a \rhd c) \vee (b \rhd c)$
- $\models (a \rhd b \wedge c) \leftrightarrow (a \rhd b) \vee (a \rhd c)$

We also consider some instances of Dung's theory. As far as we know, there is no systematic study of the possible instances of Dung's theory. We consider additional axioms we can impose on the logic MLAA. The first property we consider is irreflexivity of R, which corresponds to the property that no argument can attack itself:

IR $\neg(a \rhd a)$

The second property we consider is symmetry of the attack relation, which corresponds to the property that if argument a attacks argument b, then argument b attacks argument a.

S $(a \rhd b) \leftrightarrow (b \rhd a)$

Symmetry is not accepted often, because a counterexample attacks a general rule, but a general rule does not necessarily attack a counterexample. E.g., if Swans are white (a), but in Australia they found black swans (b) then we have $b \rhd a$ without $a \rhd b$. If the attack relation is symmetric, then the defend relation becomes reflexive, that is, each argument defends itself: $a \oslash a$.

Note that when we take traditional properties of conditional logic, we do not seem to get something useful. In particular, reflexivity (R) does not hold. Transitivity (T) means that if argument a attacks argument b, and argument b attacks argument c, then argument a should attack argument c. This does not hold either. Take $a = c$ for example, then we get $a \rhd a$, which conflicts with IR.

R $a \rhd a$
T $(a \rhd b) \wedge (b \rhd c) \supset (a \rhd c)$

Finally, if we would add accessibility relations for the monadic modal operators, then we can deal with the remaining problems observed at the end of Section 3.

Example 3. Grounded extension is unique:

- $\models G(p) \wedge G(q) \supset \Box_2(p \leftrightarrow q)$

Characterization of conflict free sets based on satisfiability checking condition of [1]:

- $C(p) \wedge (p \rhd q) \supset \Diamond_2(p \wedge \neg q)$

5 Relation with Traditional Formal Systems

5.1 Preference Logic

Since the logic of the attack connectives satisfies left and right strengthening, it seems that it may be related to preference logic. In particular, "argument a attacks argument b" may be interpreted as "argument a is preferred to argument b". However, this is less helpful than it may seem at first sight, because the area of preference logic is characterized by lack of consensus. In this subsection we make some observations.

First, the most popular branch of preference logic, as initiated in the sixties by Von Wright [16], is concerned with ceteris paribus preferences. This means that $p > q$ is interpreted as a preference of p over q all else being equal, typically interpreted as 'under the same (or similar) circumstances'. When we consider a language that does not contain disjunction or negation, then such preferences are characterized by the following property of simultaneous left and right strengthening. This is strictly weaker than left and right strengthening of attack connectives considered in this paper.

- $p > q \supset p \wedge r > q \wedge r$

Secondly, in preference logic left and right strengthening are properties of so-called strong preferences, whereas so-called weak preferences do not satisfy left and right strengthening. The typical example of a strong preference $p > q$ is interpreted as "all $p \wedge \neg q$ worlds are strictly preferred to all $q \wedge \neg p$ worlds". Without a ceteris paribus proviso these strong preferences are known to be too strong to be useful in practice, because for example $p > \neg p$ together with $q > \neg q$ is inconsistent. The reason is that such strong preferences also satisfy the following property of asymmetry. However, for the attack connective we can easily have that argument a attacks argument b, while at the same time argument b attacks argument a.

- $p > q \rightarrow \neg(q > p)$.

Thirdly, the attack relation behaves like a non-strict preference interpreted on a partial pre-order and defined by "$p \geq q$ iff there is no $q \wedge \neg p$ world that is strictly preferred to a $p \wedge \neg q$ world". As far as we can see at this moment, this relation seems coincidental and does not seem to refer to any deep connection between the two logical systems.

5.2 Conditional Logic

The defend connective behaves like a standard conditional connective, with one important exception: it does not satisfy the identity rule. An argument a does not necessarily defend argument a, because when another argument b attacks argument a, there is no reason why argument a attacks argument b (unless the attack relation is symmetric, of course).

Consequently, to consider the defend connective we need an identity free logic, which are rare. Here we use input/output logic [10, 11], which has been proposed in philosophical logic for normative or deontic reasoning, and which have been used in artificial intelligence to characterize causal reasoning [3] and logic programming [4].

To emphasize the lack of identity, Makinson and van der Torre write their conditional "if input a, then output x" as (a, x).

The defend connective behaves like so-called simple-minded output, which is defined as a closure on a set of conditionals under replacements of logical equivalents, and the following three proof rules of strengthening of the input, weakening of the output, and the conjunction rule for the output. See the above mentioned papers for semantics of this proof system.

Definition 8. *Let AS be a set of defend formulas $\{p_1 \oslash q_1, \ldots, p_n \oslash q_n\}$. Simple-minded output is the closure of AS under replacement of logical equivalents, and the following three rules.*

$$\frac{a \oslash x}{a \wedge b \oslash x} SI \qquad \frac{a \oslash x \wedge y}{a \oslash x} WO \qquad \frac{a \oslash x, a \oslash y}{a \oslash x \wedge y} AND$$

We can also formalize the attack relation as an input/output logic, if we use the same encoding as we used in modal logic MLAA, that is, if we add a negation before the output. This is done by representing "argument a attacks b" by "if input a, then output $\neg b$". Weakening of the output is then transformed into strengthening of the output, and the right conjunction rule is transformed into right disjunction. However, the latter rule is not meaningful as we have not defined disjunctions for argument sets.

Definition 9. *Let AS be an argument specification. Simple-minded output is the closure of AS under replacement of logical equivalents, and the following three rules.*

$$\frac{a \rhd x}{a \wedge b \rhd x} SI \qquad \frac{a \rhd x}{a \rhd x \wedge y} SO \qquad \frac{a \rhd x, a \rhd y}{a \rhd x \vee y} OR$$

At this point, it is very tempting to define both attack and defend in a single conditional logic to study their interaction. In other words, it is tempting to consider an input/output logic in which a conditional (p, q) is read as 'p defends q', and $(p, \neg q)$ is read as 'p attacks q', for p and q conjunctions of atomic propositions. This is formalized in the following definition of the input/output logic of abstract argumentation.

Definition 10. *Let IOLAA be simple minded output, together with the following two definitions for p and q conjunctions of atomic propositions.*

- $p \rhd q = (p, \neg q)$
- $p \oslash q = (p, q)$

Let us now consider the relation between attack and defend in IOLAA. The characteristic axiom that a defends b implies that if c attacks b, then a also attacks c, is given by the following unusual rule: $\frac{(a,b),(c,\neg b)}{(a,\neg c)}$. However, clearly we do not want to derive that a defends b implies that if c attacks b, then c also attacks a, that is: $\frac{(a,b),(c,\neg b)}{(c,\neg a)}$. Note that the distinction between these two inference rules is whether the formulas start with a negation symbol. Consequently we cannot accept one without the other, unless we add additional syntactic constraints. This illustrates that the formalization of argumentation theory in input/output logic needs further investigation.

6 Use in Argumentation

The typical approach in argumentation is that each participant makes an argument, and then argumentation theory is used to determine grounded, stable, or preferred sets of arguments. Reasoning of the agents occurs only at the level of constructing arguments, and the role of logic has been restricted to the internal structure of arguments.

A: I think p.
B: I think not p, because q.
A: But not q, because of r.

In the analysis of argumentation, reasoning about arguments has been restricted to meta-rules such as, for example, the order of arguments, or the choice of words. However, Dung [7] has shown that reasoning about arguments can also be based on concepts such as attack and defence. A typical example may be:

A: I think p and q defend r.
B: But s attacks r.
C: No problem, since p attacks s.

Note that the agents do not enumerate the complete argument system, that is, they do not list the complete attack relation R of the argument system $\langle A, R\rangle$. To formalize this example we therefor cannot assume a fixed argument system $\langle A, R\rangle$, as Dung does. We need the logical language to quantify over argument systems.

In this example, the agents make arguments like "p and q defend r" which themselves refer to arguments p, q and r. The former may therefore be called meta-arguments. The logic that formalizes or characterizes the reasoning of agents about arguments, when they construct meta-arguments, is therefore at first sight quite different from the logic typically used in argumentation. We therefore believe that the confinement of logic to the internal structure of arguments is too limited; there is also a role of logic in the formalization of reasoning about arguments.

We agree with Wooldridge *et al* that meta-argumentation is particularly useful for agent theory [18]. This meta-level could be used, potentially, to speed up argumentations by means of a kind of "caching" function. Just like in chess (Polish opening), you can use patterns of arguments, give them a name, and know that such a pattern attacks or defends another pattern. If you respect your opponent, there is no need to "play out" the whole argument.

Wooldridge *et al* [18] propose a hierarchical first-order meta-logic, which enables them to distinguish object level statements, arguments made about these object level statements, and statements about arguments. Such a distinction is commonly accepted in dialogue systems [13]. However, as a consequence their formal system appears to be more complex, and it is less clear how to relate their formal system to other formal approaches in the way we have related our system to reasoning about preferences or conditionals. A comparison between the two approaches is left for further research.

7 Concluding Remarks

In this paper we introduce a logic of abstract argumentation called LAA, with two properties usually not considered in formal theories of argumentation: it formalizes logical relations among attack and defend formulas, and it formalizes reasoning which does not assume a fixed argument system, but which quantifies over argument systems. We first define a logical system that is very close to Dung's theory. We show some properties of this logical system, focussing on the logical relations among attack and defend formulas. We also relate the logic and its properties to more traditional preference and conditional logics. To generalize Dung's setting, we turn the logic LAA into a normal bimodal logic MLAA. We define positions as sets of arguments, and define the attack relation as a binary relation between positions. We suggest that in reasoning about arguments, Dung's assumption that an argument system is given is too strong.

There are several issues for further research. For example, the modal logic represents negations and disjunctions in the left and right hand side of the attack and defend connectives. Can they be given a useful interpretation? When 'a' means that argument a has been made, then '$\neg a$' may mean that a is withdrawn in the sense that there is no longer a commitment to defend it. In such a setting, one may look into properties like:

$$\frac{a \rhd b \wedge c, a \rhd b \wedge \neg c}{a \rhd b} \qquad \frac{a \rhd b, \neg b \rhd c}{a \rhd c}$$

Moreover, the content level of argumentation contains much more than just propositional logic, including attacks and defend expressions. In a straightforward extension of our logic, reasoning about arguments such as "$a \rhd b$ attacks $c \oslash b$" can be studied using nested connectives:

$$(a \rhd b) \rhd (c \oslash b)$$

Reasoning becomes argumentation once there are two agents with opposing views. Such agents may have have beliefs and goals. Moreover, in a goal-based dialogue with sub-goals to achieve it, there may be dialogue fragments representing each of the sub-goals, probably in the same order. The use and extension of our logic for such dialogues is another topic for further research.

The modal characterization outlined in the paper raises an interesting issue, which might be worth exploring in future research. At first sight it opens the door for the application of model-checking techniques, initially used for automatically verifying a Kripke structure (describing the execution of a program) against a number of 'correctness' requirements. It is natural to ask if such techniques can be applied to argumentation. The Kripke structure to be model-checked describes an argument system (or, if you wish, a dialogue) rather than the execution of a program. And the "correctness" requirement is expressed as a formula f in MLAA rather than a temporal formula. For instance, termination seems to be an essential property of both programs and dialogues.

Finally, Bochman [5] recently introduced a logic of propositional argumentation based on the assumption-based argumentation framework of Bondarenko *et al.* [6]. A comparison is left for further research.

References

1. P. Besnard and S. Doutre. Checking the acceptability of a set of arguments. In *Procs. of NMR04*, pages 59–64, 2004.
2. A. Bochman. Collective argumentation and disjunctive logic programming. *Journal of Logic and Computation*, 13:405–428, 2003.
3. A. Bochman. A causal approach to nonmonotonic reasoning. *Artificial Intelligence*, 160(1-2):105–143, 2004.
4. A. Bochman. A causal logic of logic programming. In *Procs. of KR 2004*, pages 427–437, 2004.
5. A. Bochman. Propositional argumentation and causal reasoning. In *Procs. of IJCAI95*, pages 388–393, 2005.
6. A. Bondarenko, P. Dung, R. Kowalski, and F. Toni. An abstract, argumentation based approach to default reasoning. *Artificial Intelligence*, 93(1-2):63 – 101, 1997.
7. P. Dung. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 76:321–358, 1995.
8. E. Hemaspaandra. The price of universality. *Notre Dame Journal of Formal Logic*, 37(2):174–203, 1996.
9. J. Horty. Argument construction and reinstatement in logics for defeasible reasoning. *Artificial Intelligence and Law*, 9:1–28, 2001.
10. D. Makinson and L. van der Torre. Input-output logics. *Journal of Philosophical Logic*, 29:383–408, 2000.
11. D. Makinson and L. van der Torre. Constraints for input-output logics. *Journal of Philosophical Logic*, 30(2):155–185, 2001.
12. S. D. Parsons, C. Sierra, and N. R. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261–292, 1998.
13. H. Prakken and G. Sartor. Modelling reasoning with precedents in a formal dialogue game. *Artificial Intelligence and Law*, 6:231–287, 1998.
14. H. Prakken and G. Vreeswijk. Logics for defeasible argumentation. In D. Gabbay and F. Guenthner, editors, *Handbook of philosophical logic*, pages 218–319. Kluwer, Dordrecht, 2002.
15. B. Verheij. Accrual of arguments in defeasible argumentation. In *Proceedings of the Second Dutch/German Workshop on Nonmonotonic Reasoning*, pages 217–224, 1995.
16. Georg Henrik Von Wright. *The Logic of Preference: an Essay*. Edinburgh University Press, 1963.
17. C. Walton. Model checking agent dialogues. In *Proceedings of DALT'04*, LNCS, pages 132–147. Springer, 2005.
18. M. Wooldridge, P. McBurney, and S. Parsons. On the meta-logic of arguments. In *Proceedings of AAMAS'05*, pages 560–567, 2005.