Argumentation for Propositional Logic and Nonmonotonic Reasoning

Antonis Kakas
University of Cyprus, Cyprus
antonis@cs.ucy.ac.cy

Francesca Toni
Imperial College London, UK
ft@imperial.ac.uk

Paolo Mancarella
Università di Pisa, Italy
paolo@di.unipi.it

Abstract

Argumentation has played a significant role in understanding and unifying under a common framework different forms of defeasible reasoning in AI. Argumentation is also close to the original inception of logic as a framework for formalizing human argumentation and debate. In this context, the purpose of this paper is twofold: to draw a formal connection between argumentation and classical reasoning (in the form of Propositional Logic) and link this to support defeasible, Non-Monotonic Reasoning in AI. To this effect, we propose Argumentation Logic and show properties and extensions thereof.

Introduction

Over the past two decades argumentation has played a significant role in understanding and unifying under a common framework defeasible Non-Monotonic Reasoning (NMR) in AI (Lin and Shoham 1989; Dung 1995; Bondarenko et al. 1997). Moreover, a foundational role for argumentation has emerged in the context of problems requiring human-like commonsense reasoning, e.g. as found in the area of open and dynamic multi-agent systems to support context-dependent decision making, negotiation and dialogue between agents (e.g. see (Kakas and Moraitis 2003; Dung, Thang, and Toni 2008)). This foundational role of argumentation points back to the original inception of logic as a framework for formalizing human argumentation.

This paper reexamines the foundations of classical logical reasoning from an argumentation perspective, by formulating a new logic of arguments, called Argumentation Logic (AL), and showing how this relates to Propositional Logic. AL is formulated by bringing together argumentation theory from AI and the syllogistic view of logic in Natural Deduction (ND). Its definition rests on a re-interpretation of Reductio ad Absurdum (RA) through a suitable argumentation semantics. One consequence of this is that in AL the implication connective behaves like a default rule that still allows a form of contrapositive reasoning. The reasoning in AL is such that the ex-falso rule where everything can be derived from an inconsistent theory does not apply and hence an inconsistent part of a theory does not necessarily trivialize the whole reasoning with that theory.

The main motivation for studying this argumentation perspective on logical reasoning is to examine how its use to bring together classical reasoning and non-monotonic commonsense reasoning into a single unified framework. The paper presents a preliminary investigation into building such a NMR framework based on AL that integrates into the single representation framework of AL classical reasoning, as in Propositional Logic including forms of Reductio ad Absurdum, with defeasible reasoning. In particular, we study, in the context of examples, the possible use of preferences over sentences of an AL theory to capture NMR defeasible reasoning and naturally combine this with the classical reasoning of AL. Our vision is for all forms of reasoning to be captured in the argumentation reasoning of AL and its extensions with preferences.

Preliminaries on Natural Deduction

Let $\mathcal{L}$ be a Propositional Logic language and $\vdash$ denote the provability relation of Natural Deduction (ND) in Propositional Logic.[1] Throughout the paper, theories and sentences will always refer to theories and sentences wrt $\mathcal{L}$. We assume that $\bot$ stands for $\phi \wedge \neg\phi$, for any $\phi \in \mathcal{L}$.

Definition 1 *Let T be a theory and ϕ a sentence. A* direct derivation *for ϕ (from T) is a ND derivation of ϕ (from T) that does not contain any application of RA. If there is a direct derivation for ϕ (from T) we say that ϕ is* directly derived *(or* derived modulo RA*) from T, denoted $T \vdash_{MRA} \phi$.*

Example 1 *Let $T_1 = \{\alpha \rightarrow (\beta \rightarrow \gamma)\}$. The following is a direct derivation for $\alpha \wedge \beta \rightarrow \gamma$ (from T_1):*

	$\lceil \alpha \wedge \beta$	*hypothesis*
	α	$\wedge E$
	$\alpha \rightarrow (\beta \rightarrow \gamma)$	*from T*
	$\beta \rightarrow \gamma$	$\rightarrow E$
	β	$\wedge E$
	$\gamma \rfloor$	$\rightarrow E$
$\alpha \wedge \beta \rightarrow \gamma$		$\rightarrow I$

Thus, $T_1 \vdash_{MRA} \alpha \wedge \beta \rightarrow \gamma$ (and, trivially, $T_1 \vdash \alpha \wedge \beta \rightarrow \gamma$). Consider now $T_2 = \{\neg(\alpha \vee \beta)\}$. The following

	$\lceil \alpha$	*hypothesis*
	$\alpha \vee \beta$	$\vee I$
	$\neg(\alpha \vee \beta)$	*from T*
	$\bot \rfloor$	$\wedge I$
$\neg\alpha$		*RA*

[1] See the appendix for a review of the ND rules that we use.

is a ND derivation of $\neg\alpha$ that is not direct. Since there is no direct derivation of $\neg\alpha$, $T_2 \vdash \neg\alpha$ but $T_2 \nvdash_{MRA} \neg\alpha$.

Definition 2 *A theory T is* classically inconsistent *iff $T \vdash \bot$. A theory T is* directly inconsistent *iff $T \vdash_{MRA} \bot$. A theory T is* (classically or directly) consistent *iff it is not (classically or directly, respectively) inconsistent.*

Trivially, if a theory is classically consistent then it is directly consistent. However, a directly consistent theory may be classically inconsistent, as the following example shows.

Example 2 *Consider $T_1 = \{\alpha \to \bot, \neg\alpha \to \bot\}$. T_1 is classically inconsistent, since $T_1 \vdash \bot$ as follows:*

$\lceil\alpha$	*hypothesis*
$\alpha \to \bot$	*from T*
$\bot\rfloor$	$\to E$
$\neg\alpha$	*RA*
$\lceil\neg\alpha$	*hypothesis*
$\neg\alpha \to \bot$	*from T*
$\bot\rfloor$	$\to E$
α	*RA*
$\bot$	$\wedge I$

However, T_1 is directly consistent, since $T_1 \nvdash_{MRA} \bot$. Consider instead $T_2 = \{\alpha, \neg\alpha\}$. T_2 is classically and directly inconsistent, since $T_2 \vdash_{MRA} \bot$, as follows:

α	*from T*
$\neg\alpha$	*from T*
$\bot$	$\wedge I$

We will use a special kind of ND derivations, that we call *Reduction ad Absurdum Natural Deduction* (RAND). These are ND derivations with an outermost application of RA:

Definition 3 *A* RAND derivation of $\neg\phi \in \mathcal{L}$ from T *is a ND derivation of $\neg\phi$ from T of the form*

$\lceil\phi$	*hypothesis*
$\vdots$	$\vdots$
$\bot\rfloor$	$\vdots$
$\neg\phi$	*RA*

A sub-derivation (of some $\psi \in \mathcal{L}$) of a derivation d *is a* RAND sub-derivation of d *iff it is a RAND derivation.*

The ND derivation of $\neg\alpha$ given T_2 in example 1 is a RAND derivation. Also, given T_1 in example 2, the sub-derivations[2]

$\lceil\alpha$	$\lceil\neg\alpha$
$\alpha \to \bot$	$\neg\alpha \to \bot$
$\bot\rfloor$	$\bot\rfloor$
$\neg\alpha$	α

of the derivation (d) of $\bot$ are RAND sub-derivations (of d).

Argumentation Logic Frameworks

Given a propositional theory we will define a corresponding argumentation framework as follows.

Definition 4 *The* argumentation logic (AL) framework corresponding to *a theory T is the triple $\langle Args^T, Att^T, Def^T\rangle$ defined as follows:*

- *$Args^T = \{T \cup \Sigma | \Sigma \subseteq \mathcal{L}\}$ is the set of all extensions of T by sets of sentences in $\mathcal{L}$;*
- *given $a, b \in Args^T$, with $a = T \cup \Delta$, $b = T \cup \Gamma$, such that $\Delta \neq \{\}$, $(b, a) \in Att^T$ iff $a \cup b \vdash_{MRA} \bot$;*
- *given $a, d \in Args^T$, with $a = T \cup \Delta$, $(d, a) \in Def^T$ iff*
  1. *$d = T \cup \{\neg\phi\}$ ($d = T \cup \{\phi\}$) for some sentence $\phi \in \Delta$ (respectively $\neg\phi \in \Delta$), or*
  2. *$d = T \cup \{\}$ and $a \vdash_{MRA} \bot$.*

In the remainder, b attacks *a* (wrt T) *stands for $(b, a) \in Att^T$ and d* defends *or* is a defence against *a (wrt T) stands for $(d, a) \in Def^T$.*

Note also that, since T is fixed, we will often *equate arguments $T \cup \Sigma$ to sets of sentences Σ*. So, for example, we will refer to $T \cup \{\} = T$ as the *empty argument*. Similarly, we will often equate a defence to a set of sentences. In particular, when $d = T \cup D$ defends/is a defence against $a = T \cup \Delta$ we will say that *D defends/is a defence against Δ (wrt T)*.

The attack relation between arguments is defined in terms of a direct derivation of inconsistency. Note that, trivially, for $a = T \cup \Delta$, $b = T \cup \Gamma$, $(b, a) \in Att^T$ iff $T \cup \Delta \cup \Gamma \vdash_{MRA} \bot$. The following example illustrates our notion of attack:

Example 3 *Given $T_1 = \{\alpha \to (\beta \to \gamma)\}$ in example 1, $\{\alpha,\beta\}$ attacks $\{\neg\gamma\}$ (and vice-versa), $\{\alpha,\neg\gamma\}$ attacks $\{\beta\}$ (and vice-versa), $\{\alpha,\neg\alpha\}$ attacks $\{\gamma\}$ (and vice-versa) as well as any non-empty set of sentences (and vice-versa).*

Note that the attack relationship is symmetric except for the case of the empty argument. Indeed, for a, b both non-empty, it is always the case that a attacks b iff b attacks a. However, the empty argument cannot be attacked by any argument (as the attacked argument is required to be non-empty), but the empty argument can attack an argument. For example, given $T_2 = \{\alpha, \neg\alpha\}$ in example 2, $\{\}$ attacks $\{\alpha\}$ and $\{\}$ attacks $\{\beta\}$ (for any $\beta \in \mathcal{L}$), since $T \vdash_{MRA} \bot$. Finally, note that our notion of attack includes the special case of attack between a sentence and its negation, since, for any theory T, $\{\phi\}$ attacks $\{\neg\phi\}$ (and vice-versa), for any $\phi \in \mathcal{L}$.

The notion of defence is a subset of the attack relation. In the first case of the definition we defend against an argument by adopting the complement[3] of some sentence in the argument, whereas in the second case we defend against any directly inconsistent set using the empty argument. Then, in example 3, $\{\neg\alpha\}$ defends against the attack $\{\alpha, \beta\}$ and $\{\}$ defends against the (directly inconsistent) attack $\{\alpha, \neg\alpha\}$. Note that the empty argument cannot be defended against if T is directly consistent. Finally, note that when T is directly inconsistent the attack and defence relationships trivialise, in the sense that any two non-empty arguments attack each other, the empty argument attacks any argument, and any argument can be defended against by the empty argument.

Argumentation Logic

In this section we assume that T is *directly consistent*.

[2] If clear from the context, we omit to give the ND rules used.

[3] The complement of a sentence ϕ is $\neg\phi$ and the complement of a sentence $\neg\phi$ is ϕ.

As conventional in argumentation, we define a notion of acceptability of sets of arguments to determine which conclusions can be dialectically justified (or not) from a given theory. Our definition of acceptability and non-acceptability is formalised in terms of the least fix point of (monotonic) operators on the cartesian product of the set of arguments/sentences in $\mathcal{L}$, as follows:

Definition 5 *Let $\langle Args^T, Att^T, Def^T\rangle$ be the AL framework corresponding to a directly consistent theory T, and $\mathcal{R}$ the set of binary relations over $Args^T$.*

- *The* acceptability operator $\mathcal{A}_T{:}\mathcal{R}\rightarrow\mathcal{R}$ *is defined as follows: for any $acc \in \mathcal{R}$ and $a, a_0 \in Args^T$:*
  $(a, a_0) \in \mathcal{A}_T(acc)$ *iff*
  - $a \subseteq a_0$, *or*
  - *for any $b \in Args^T$ such that b attacks a wrt T,*
    - $b \not\subseteq a_0 \cup a$, *and*
    - *there exists $d \in Args^T$ that defends against b wrt T such that $(d, a_0 \cup a) \in acc$.*
- *The* non-acceptability operator $\mathcal{N}_T : \mathcal{R} \rightarrow \mathcal{R}$ *is defined as follows: for any $nacc \in \mathcal{R}$ and $a, a_0 \in Args^T$:*
  $(a, a_0) \in \mathcal{N}_T(nacc)$ *iff*
  - $a \not\subseteq a_0$, *and*
  - *there exists $b \in Args^T$ such that b attacks a wrt T and*
    - $b \subseteq a_0 \cup a$, *or*
    - *for any $d \in Args^T$ that defends against b wrt T, $(d, a_0 \cup a) \in nacc$.*

These $\mathcal{A}_T$ and $\mathcal{N}_T$ operators are monotonic wrt set inclusion and hence their repeated application starting from the empty binary relation will have a least fixed point.

Definition 6 *ACC^T and $NACC^T$ denote the least fixed points of $\mathcal{A}_T$ and $\mathcal{N}_T$ respectively. We say that a* is acceptable wrt a_0 in T *iff $ACC^T(a, a_0)$, and a* is not acceptable wrt a_0 in T *iff $NACC^T(a, a_0)$.*

Thus, given the AL framework $\langle Args^T, Att^T, Def^T\rangle$ (for T directly consistent) and $a, a_0 \in Args^T$:

- $ACC^T(a, a_0)$, iff
  - $a \subseteq a_0$, or
  - for all $b \in Args^T$ such that b attacks a:
    - $b \not\subseteq a_0 \cup a$, and
    - there exists $d \in Args^T$ such that d defends against b and $ACC^T(d, a_0 \cup a)$;
- $NACC^T(a, a_0)$, iff
  - $a \not\subseteq a_0$ and
  - there exists $b \in Args^T$ such that b attacks a and
    - $b \subseteq a_0 \cup a$, or
    - for all $d \in Args^T$ such that d defends against b it holds that $NACC^T(d, a_0 \cup a)$.

We will often equate the (non-)acceptability of an argument $T \cup \Delta$ wrt an argument $T \cup \Delta_0$ to the (non-)acceptability of the set of sentences Δ wrt the set of sentences Δ_0.

Note that non-acceptability, $NACC^T(a, a_0)$, is the same as the classical negation of $ACC^T(a, a_0)$, i.e. $NACC^T(a, a_0) = \neg ACC^T(a, a_0)$. We will use these two versions of non-acceptability interchangeably.

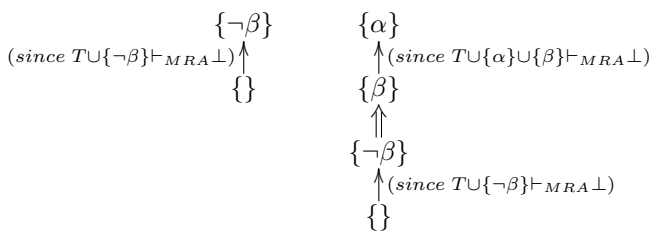

Figure 1: Illustration of $NACC^T(\{\neg\beta\}, \{\})$ (left) and $NACC^T(\{\alpha\}, \{\})$ (right), for example 4.

Note that the empty argument is always acceptable, wrt any other argument. Note also that the "canonical" attack of a sentence on its complement (i.e. of $T \cup \{\phi\}$ on $T \cup \{\neg\phi\}$ and vice-versa) does not affect the acceptability relation as it can always be defended against by this complement itself.

The following examples illustrate non-acceptability.

Example 4 *Let $T = \{\alpha \wedge \beta \rightarrow \bot, \neg\beta \rightarrow \bot\}$. T is classically and directly consistent, $T \cup \{\neg\beta\}$ is classically and directly inconsistent, and $T \cup \{\alpha\}$ is classically inconsistent but directly consistent. It is easy to see that $NACC^T(\{\neg\beta\}, \{\})$ holds, as illustrated in figure 1 (left)*[4]*, since $\{\neg\beta\} \not\subseteq \{\}$, $b = \{\}$ attacks $\{\neg\beta\}$ and $\{\} \subseteq \{\neg\beta\}$. Also, $NACC^T(\{\alpha\}, \{\})$ holds, as illustrated in figure 1 (right). Indeed:*

- *since $\{\alpha\} \not\subseteq \{\}$, $b = \{\beta\}$ attacks $\{\alpha\}$ and $\{\neg\beta\}$ is the only defence against b, to prove that $NACC^T(\{\alpha\}, \{\})$ it suffices to prove that $NACC^T(\{\neg\beta\}, \{\alpha\})$;*
- *since $\{\neg\beta\} \not\subseteq \{\alpha\}$, $b = \{\}$ attacks $\{\neg\beta\}$ and $\{\} \subseteq \{\alpha, \neg\beta\}$, $NACC^T(\{\neg\beta\}, \{\alpha\})$ holds as required.*

Note that if an argument a is attacked by the empty argument, then it is acceptable wrt any a_0 iff $a \subseteq a_0$, since there is no defence against the empty argument. This observation is used in the following example.

Example 5 *Given $T = T_1 = \{\alpha \rightarrow \bot, \neg\alpha \rightarrow \bot\}$ in example 2, $NACC^T(\{\alpha\}, \{\})$ and $NACC^T(\{\neg\alpha\}, \{\})$ both hold. Indeed, for $NACC^T(\{\alpha\}, \{\})$, $\{\alpha\}$ is attacked by $\{\}$.*

The following example shows a case of non-acceptability making use of a non-empty attack for the base case.

Example 6 *Let $T = \{\alpha \wedge \neg\beta \rightarrow \bot, \beta \wedge \gamma \rightarrow \bot, \alpha \wedge \beta \wedge \neg\gamma \rightarrow \bot\}$. T is classically (and directly) consistent, and $T \cup \{\alpha\}$ is classically inconsistent but directly consistent. $NACC^T(\{\alpha\}, \{\})$ holds, as illustrated in figure 2. Indeed:*

- *since $\{\alpha\} \not\subseteq \{\}$, $b = \{\neg\beta\}$ attacks $\{\alpha\}$ and $\{\beta\}$ is the only defence against b, to prove that $NACC^T(\{\alpha\}, \{\})$ it suffices to prove that $NACC^T(\{\beta\}, \{\alpha\})$;*
- *since $\{\beta\} \not\subseteq \{\alpha\}$, $b' = \{\gamma\}$ attacks $\{\beta\}$ and $\{\neg\gamma\}$ is the only defence against b', to prove that $NACC^T(\{\beta\}, \{\alpha\})$ it suffices to prove that $NACC^T(\{\neg\gamma\}, \{\alpha, \beta\})$;*

[4] Here and throughout the paper we adopt the following graphical convention: $\uparrow$ denotes an attack and $\Uparrow$ denotes a defence.

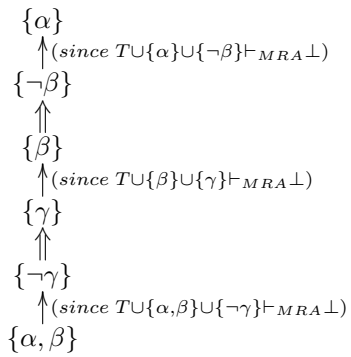

Figure 2: Illustration of $NACC^T(\{\alpha\},\{\})$ for example 6.

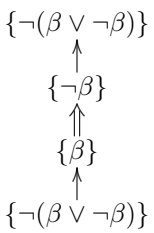

Figure 3: Illustration of $NACC^T(\{\neg(\beta\vee\neg\beta)\},\{\})$ for example 7.

- *since* $\{\neg\gamma\}\not\subseteq\{\alpha,\beta\}$, $b''=\{\alpha,\beta\}$ *attacks* $\{\neg\gamma\}$ *and* $b''\subseteq\{\alpha,\beta,\neg\gamma\}$, $NACC^T(\{\neg\gamma\},\{\alpha,\beta\})$ *holds and so* $NACC^T(\{\beta\},\{\alpha\})$ *and* $NACC^T(\{\alpha\},\{\})$ *both hold.*

The following example illustrates non-acceptability in the case of an empty (and thus classically consistent) theory.

Example 7 *For* $T=\{\}$, $NACC^T(\{\neg(\beta\vee\neg\beta)\},\{\})$ *holds, as illustrated in figure 3. Also, trivially,* $NACC^T(\{\beta\wedge\neg\beta\},\{\})$ *holds, since it is attacked by the empty argument.*

A novel, alternative notion of *entailment* can be defined for theories that are directly consistent in terms of the (non-) acceptability semantics for AL frameworks, as follows:

Definition 7 *Let* T *be a directly consistent theory and* $\phi\in\mathcal{L}$. *Then* ϕ *is* AL-entailed by T *(denoted* $T\models_{AL}\phi$*) iff* $ACC^T(\{\phi\},\{\})$ *and* $NACC^T(\{\neg\phi\},\{\})$.

This is motivated by the argumentation perspective, where an argument is held if it can be successfully defended and it cannot be successfully objected against.

In the remainder of the paper we will study properties of $\models_{AL}$ and discuss extensions thereof to support NMR.

Basic Properties

The following result gives a core property of the notion of AL-entailment wrt the notion of direct derivation in Propositional Logic, for directly consistent theories.

Proposition 1 *Let* T *be a* directly consistent *theory and* $\phi\in\mathcal{L}$ *such that* $T\vdash_{MRA}\phi$. *Then* $T\models_{AL}\phi$.

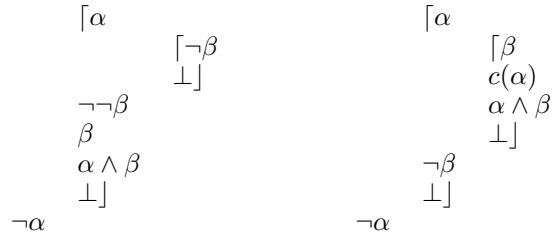

Figure 4: Two RAND derivations of $\neg\alpha$ in example 4: $\mathtt{d}_1$ (left) and $\mathtt{d}_2$ (right).

Proof: Let $a=T\cup\Delta$ be any attack against $\{\phi\}$, i.e. $T\cup\{\phi\}\cup\Delta\vdash_{MRA}\bot$. Since $T\vdash_{MRA}\phi$ then $T\cup\Delta\vdash_{MRA}\bot$. Since T is directly consistent, $\Delta\neq\{\}$. Hence any such a can be defended against by the empty argument. Since $ACC^T(\{\},\Sigma)$, for any $\Sigma\subseteq\mathcal{L}$, then $ACC^T(\{\phi\},\{\})$ holds. Moreover, since $T\vdash_{MRA}\phi$, necessarily $T\cup\{\neg\phi\}\vdash_{MRA}\bot$. Hence the empty argument attacks $\{\neg\phi\}$ and thus $NACC^T(\{\neg\phi\},\{\})$ holds. QED

The following theorem shows (one half of) the link of AL with Propositional Logic by showing how the RA rule, deleted from the ND proof system within $\vdash_{MRA}$, can be recovered back through the notion of non-acceptability.[5]

Theorem 1 *Let* T *be a* directly consistent *theory and* $\phi\in\mathcal{L}$. *If* $NACC^T(\{\phi\},\{\})$ *holds then there exists a RAND derivation of* $\neg\phi$ *from* T.[6]

For example, the RAND derivation corresponding to the proof of $NACC^T(\{\alpha\},\{\})$ in figure 1 is $\mathtt{d}_1$ in figure 4.[7] Here, the inner RAND derivation in $\mathtt{d}_1$ corresponds to the non-acceptability of the defence $\{\neg\beta\}$ against the attack $\{\beta\}$ against $\{\alpha\}$. Derivation $\mathtt{d}_2$ in figure 1 is an alternative RAND of $\neg\alpha$, but this cannot be obtained from any proof of $NACC^T(\{\alpha\},\{\})$, because there is a defence against the attack $\{\beta\}$ given by the empty set (in other words, $\mathtt{d}_2$ does not identify a useful attack, that cannot be defenced against, for proving non-acceptability).

AL for Propositional Logic

The following result gives a core property of the notion of non-acceptability for *classically consistent* theories.

Proposition 2 *Let* T *be* classically consistent *and* $\phi\in\mathcal{L}$. *If* $NACC^T(\{\neg\phi\},\{\})$ *holds then* $ACC^T(\{\phi\},\{\})$ *holds.*

Proof: By theorem 1, since $NACC^T(\{\neg\phi\},\{\})$, then $T\vdash\phi$. Suppose, by contradiction, that $ACC^T(\{\phi\},\{\})$

[5]The other half of this result shows how (under some conditions) a RAND derivation of $\neg\phi$ implies $NACC^T(\{\phi\},\{\})$, proven in (Kakas, Toni, and Mancarella 2012).

[6]The proof of this theorem is included for completeness of presentation and for inspection by the reviewers in the appendix.

[7]Here and elsewhere in the paper, $c(\phi)$, for any $\phi\in\mathcal{L}$, indicates that ϕ is the hypothesis of an ancestor sub-derivation copied within the current sub-derivation.

does not hold. Then $NACC^T(\{\phi\},\{\})$ holds (since $NACC^T(\{\phi\},\{\})$=$\neg ACC^T(\{\phi\},\{\})$) and by theorem 1 there is a RAND derivation of $\neg\phi$ from T and thus $T \vdash \neg\phi$. This implies that T is classically inconsistent: contradiction. Hence $ACC^T(\{\phi\},\{\})$ holds. QED

Thus, in Propositional Logic, trivially AL-entailment reduces to the notion on non-acceptability:

Corollary 1 *Let T be a* classically consistent *theory and $\phi \in \mathcal{L}$. Then $T \models_{AL} \phi$ iff $NACC^T(\{\neg\phi\},\{\})$.*

The following property sanctions that AL-entailment implies classical derivability:

Corollary 2 *Let T be a* classically consistent *theory and $\phi \in \mathcal{L}$. If $T \models_{AL} \phi$ then $T \vdash \phi$.*

Proof: If $NACC^T(\{\neg\phi\},\{\})$, then, by theorem 1, there is a RAND derivation of $\neg\neg\phi$ from T and thus $T \vdash \phi$. QED

This corollary gives that consequences of a classically consistent theory under $\models_{AL}$ are classical consequences too. Proposition 1 sanctions that direct consequences are not lost by $\models_{AL}$. However, in general not all classical consequences are retrieved by $\models_{AL}$, namely the converse of corollary 2 does not hold, as the following example shows.

Example 8 *Let $T = \{\neg\alpha\}$. We show that $T \not\models_{AL} \alpha \rightarrow \beta$ by showing that $NACC^T(\{\neg(\alpha \rightarrow \beta)\},\{\})$ does not hold. A standard ND derivation of $\alpha \rightarrow \beta$ from T is:*

$$
\begin{array}{lll}
\lceil\alpha & & \\
 & \lceil\neg\beta & \\
 & c(\alpha) & \\
 & \neg\alpha & \textit{from } T \\
 & \bot\rfloor & \\
\neg\neg\beta & & \textit{RA} \\
\beta\rfloor & & \neg E \\
\alpha \rightarrow \beta & & \rightarrow I
\end{array}
$$

This does not help with determining $NACC^T(\{\neg(\alpha \rightarrow \beta)\},\{\})$. This is related to the fact that the inconsistency in the inner RAND derivation of $\neg\neg\beta$ is derived without the need of the hypothesis, $\neg\beta$, of this RAND derivation. In general, any RAND derivation of $\neg\neg(\alpha \rightarrow \beta))$ (and hence of $\alpha \rightarrow \beta$) from this theory, T, contains such a RAND sub-derivation relying on the inconsistency of the copy of α from a ($\rightarrow$I) sub-derivation, with $\neg\alpha$ from T. This means that $NACC^T(\{\neg(\alpha \rightarrow \beta)\},\{\})$ cannot hold, since, otherwise, by theorem 1, we would have a RAND derivation of $\neg\neg(\alpha \rightarrow \beta)$ without such a sub-derivation. This is because by construction of the corresponding RAND derivation given by theorem 1 the existence of such a RAND sub-derivation would violate the non-acceptability of some defence in the assumed non-acceptability of $\neg(\alpha \rightarrow \beta)$.

This example shows, in particular, that implication is not material implication under $\models_{AL}$.

AL for Non-Monotonic Reasoning-Discussion

Here we present a first investigation on how AL can be used as a basis for NMR unifying classical and defeasible reasoning, in the context of the well known $tweety$ example. Our examination is based on the (expected) need to extend AL with preferences and the observation that when a theory is (directly) inconsistent we have the possibility to reason with its sub-theories, considering these as arguments that support their conclusions under AL. For the illustration we use the following (abbreviations of) sentences:

$\phi_{bf} = [bird(tweety) \rightarrow fly(tweety)]$
$\phi_{p\neg f} = [penguin(tweety) \rightarrow \neg fly(tweety)]$
$\phi_{pb} = [penguin(tweety) \rightarrow bird(tweety)]$
$\phi_{\neg f} = [\neg fly(tweety)]$ $\quad$ $\phi_p = [penguin(tweety)]$
$\phi_{\neg b\neg p} = [\neg bird(tweety) \rightarrow \neg penguin(tweety)]$

Example 9 *Let $T = \{\phi_{bf},\phi_{pb},\phi_{\neg f}\}$ (T is classically consistent). $T \models_{AL} \neg bird(tweety)$ as $\{\}$ attacks $\{bird(tweety)\}$ and thus $NACC^T(\{bird(tweety)\},\{\})$. Similarly, $T \models_{AL} \neg penguin(tweety)$. We believe that, in absence of other information, these conclusions are legitimate and desirable.*

Note that AL does not distinguish default rules and facts and it supports contrapositive reasoning with the single form of implication it allows. In example 9, default logic (Reiter 1980) would derive the same conclusions only by labelling T as facts, but would not derive either conclusion if the first sentence were labelled as a default rule, as conventional.

Example 10 *Let $T = \{\phi_{bf}, \phi_{pb}, \phi_{\neg f}, \phi_{p\neg f}\}$ (T classically consistent, obtained by adding $\phi_{p\neg f}$ to T in example 9). $T \models_{AL} \neg bird(tweety)$ and $T \models_{AL} \neg penguin(tweety)$, as in example 9. This is counter-intuitive, as it disregards the newly added sentence and the alternative possibility for $\neg fly(tweety)$ it supports, namely $penguin(tweety)$.*

By comparison, default logic with the first and last sentences in T labelled as default rules (as conventional) would (sceptically) derive no conclusion as to whether $tweety$ is (or not) a bird or penguin. Arguably, this is too sceptical a behaviour.

Note that we have the same counter-intuitive behaviour of deriving $\neg penguin(tweety)$ when the sentence $\neg fly(tweety)$ is deleted from the theory of example 10. In order to accommodate within AL the intuitive kind of reasoning pointed out for these examples, we can extend AL with *priorities* over sentences, so that, in particular, exceptions may override rules, in the spirit of prioritised default logic (Brewka 1994; Brewka and Eiter 2000) and other approaches to supporting reasoning with priorities (Delgrande et al. 2004). In our illustration, these priorities may be drawn from the partial order $\phi_{\neg f}, \phi_p, \phi_{pb}, \phi_{\neg b\neg p} > \phi_{p\neg f} > \phi_{bf}$. The challenge is to incorporate these priorities without imposing a separation amongst sentences (as done instead in prioritised and standard default logic) and without imposing a specific structure on the defeasible knowledge (the default rules) so as to achieve, e.g., the behaviour of AL in example 9. In example 10, the given priorities may be used to identify the sub-theory $\{\phi_{pb}, \phi_{\neg f}, \phi_{p\neg f}\}$ as the strongest and thus entail $penguin(tweety)$.

By introducing priorities we can also use preference-based argumentation, as in e.g. (Kakas and Moraitis 2003; Modgil and Prakken 2012), to distinguish between strengths of AL-entailment from sub-theories, and, in particular, allow for stronger sub-theories to dominate, as illustrated by the following example:

Example 11 *Let $T = \{\phi_{bf}, \phi_p, \phi_{\neg f}, \phi_{\neg b \neg p}\}$ (T is directly but not classically consistent). Then, correctly, in absence of other information, $T \not\models_{AL} bird(tweety)$ and $T \not\models_{AL} \neg bird(tweety)$. The sub-theories $T_1 = \{\phi_{bf}, \phi_{\neg f}\}$ and $T_2 = \{\phi_p, \phi_{\neg b \neg p}\}$ AL-entail $\neg bird(tweety)$ and $bird(tweety)$ respectively and hence dispute each other. If we now take into account $\phi_{\neg b \neg p} > \phi_{bf}$, then, under a preference-based argumentation approach, T_2 would dominate T_1 and thus T would correctly entail $bird(tweety)$.*

The core technical challenge of using priorities over sentences is to understand how these could influence the reasoning by contradiction afforded by RA in AL. In our illustrative setting we want the priorities (especially $\phi_{p\neg f} > \phi_{bf}$) to restrict the application of RA. There are other cases, however, where RA gives intuitive results and should not be restricted. For example, from the theory $\{bird(tweety), \phi_{pb}, \phi_{p\neg f}, \phi_{bf}\}$ with $\phi_{pb} > \phi_{p\neg f} > \phi_{bf}$ we expect that $\neg penguin(tweety)$ is entailed since $fly(tweety)$ is an intuitive default conclusion of this theory and then, by RA, $penguin(tweety)$ cannot be entailed (as otherwise through the stronger sentence of $\phi_{p\neg f}$, the sentence $\neg fly(tweety)$ would follow). Similarly, given the theory $\{fly(tweety), \phi_{pb}, \phi_{p\neg f}, \phi_{bf}\}$ with $\phi_{pb} > \phi_{p\neg f} > \phi_{bf}$, we expect that $\neg penguin(tweety)$ is entailed as $penguin(tweety)$ would give $\neg fly(tweety)$ due to the higher strength of $\phi_{p\neg f}$. To accommodate such cases it may be necessary to use the priority information more tightly within the definition of AL, i.e. within the definiton of (non-)acceptability.

Related Work

AL is based on a notion of acceptability of arguments which is in the same spirit as that in (Dung, Kakas, and Mancarella 1992; Kakas, Mancarella, and Dung 1994) for capturing the semantics of negation as failure in Logic Programming. These notions of acceptability are global in the sense that acceptable and non-acceptable arguments are all defined at the same time. This view has also recently been taken in (Caminada and Gabbay 2009; Wu and Caminada 2010) where the argumentation semantics is defined through the notion of a global labelling of arguments as IN, OUT or UNDECIDED.

The link of argumentation to NMR has been the topic of extensive study for many years. Most of these studies either separate in the language the classical reasoning from the defeasible part of the theory (e.g. in Default Logic) or restrict the classical reasoning (e.g. in LP with NAF) or indeed as in the case of circumscription (McCarthy 1980) the theory is that of classical logic but a complex prescription of model selection is imposed on top of the classical reasoning.

Recently, (Besnard and Hunter 2008) proposed an argumentation framework based upon classical logic with the aim (that we share) to use argumentation to reason with possibly inconsistent classical theories, beyond the realms of classical logic. In their approach, arguments are defined in terms of sub-theories of a given (typically inconsistent) theory and they have minimal and consistent supports (wrt the full classical consequence relation). Attacks are defined in terms of a notion of canonical undercut that relies on arguments for the negation of the support of attacked argument. Further, the evaluation of arguments is given through a related tree structure of defeated or undefeated nodes.

Other works that aim for a tighter link between classical and defeasible reasoning include the work of Amgoud and Vesic (Amgoud and Vesic 2010), studying the problem of handling inconsistency using argumentation with priorities over sentences, and (Zhang et al. 2010), who have adapted the approach of (Besnard and Hunter 2008) to Description Logic and have proposed an argumentation-based operator to repair inconsistencies. Our approach differs from these works in that it starts with providing an alternative understanding of Propositional Logic in argumentation terms on which to base any further development of reasoning with inconsistent or defeasible theories. In comparison with our approach, these other works can be seen more as a form of belief revision, based on argumentation, for classically inconsistent theories rather than a re-examination of classical logic through argumentation to provide a uniform basis for classical and defeasible reasoning.

Conclusion and Future Work

We have presented Argumentation Logic (AL) and shown how it allows us to understand classical reasoning in Propositional Logic in terms of argumentation. Its definition rests on capturing semantically the Reductio ad Absurdum rule through a suitable notion of acceptability of arguments. One property of the ensuing AL is that the interpretation of implication is different from that of material implication. Further results on the relationship between AL and Propositional Logic including how AL can completely capture the entailment of PL are given in (Kakas, Toni, and Mancarella 2012).

Given the significant role that argumentation has played in understanding under a common framework NMR in AI we have examined the problem of how we could unify classical reasoning and NMR within the framework of AL. In this context, we have considered the following questions: How could we use AL as the underlying logic to build a NMR framework? Can AL with its propositional language provide a single representation framework for classical and defeasible reasoning without any distinctions on the type of sentences allowed in a given theory? In particular, can we understand AL as a NMR framework with sentences that would behave as default rules but also as classical rules, with a form of contrapositive reasoning with these rules allowed? In this paper we have identified this problem and the challenges it poses, and studied these questions in the context of examples.

Our preliminary investigation suggests the need for an extension of AL to accommodate preferences amongst sentences. Many existing frameworks for NMR use, either explicitly or implicitly, preferences to capture defeasible reasoning, e.g. (Brewka 1994; Brewka and Eiter 2000) for Default Logic (Reiter 1980). Also many frameworks of argumentation rely on some form of preference between arguments, e.g. (Kakas and Moraitis 2003; Kowalski and Toni 1996; Modgil and Prakken 2012) to capture a notion of (relative) strength of arguments through which the attack relation between arguments can be realized. One way therefore

to study this problem of integrating classical and defeasible reasoning is to use some form of preference on the sentences of AL theories, and adapt existing approaches of reasoning with preferences to AL.

Naturally, the question arises as to where these preferences would come from. As we have suggested these could be in the form of priority orderings which need not to be total, expressing only whatever priority information is known. As such, these partial priority orderings can be incrementally learned thus making the reasoning more complete as more learning is performed[8] In particular, for commonsense reasoning knowledge natural preferences between types of information (that can then be mapped into argument preferences) have already been identified, e.g. that causal information is stronger than persistence, or that forward persistence from later information is stronger than that from earlier information, or that specific case information is stronger than general information etc. Furthermore, such preference schemas as well as more specific preferences amongst commonsense knowledge can be learned by exploiting the corpus of information over the Web using and adapting existing learning frameworks for semi-autonomous preference elicitation, e.g. (Dimopoulos, Michael, and Athienitou 2009; Michael 2011). In fact, such learning methods can be used more generally to learn over the Web not only the preferences but the whole of AL theories as theories of commonsense knowledge, in line with recent studies (Doppa et al. 2011; Michael 2010; 2011) that have suggested that web-extracted knowledge can be seen as a form of commonsense knowledge of (default) associations between concepts.

As a consequence, a natural domain of application of AL and its unified extension for defeasible reasoning is that of *textual entailment* (Dagan, Glickman, and Magnini 2006; Michael 2009) and text comprehension more generally. This is a challenge for testing and evaluating the suitability of AL and more generally the argumentation perspective for automating commonsense reasoning. We have already begun to investigate this in the particular context of Narrative Test Comprehension (Mueller 2002) where we are interested in the specific form of textual entailment of elaborative inferences from the (relevant) background commonsense knowledge, expressed as unified extended AL theories, under the narrative information given in a story text (Diakidoy et al. 2013).

Appendix: Natural Deduction

We use the following rules, for any $\phi, \psi, \chi \in \mathcal{L}$:

$$\wedge I : \frac{\phi, \psi}{\phi \wedge \psi} \quad \wedge E : \frac{\phi \wedge \psi}{\phi} \quad \wedge E : \frac{\phi \wedge \psi}{\psi} \quad \vee I : \frac{\phi}{\phi \vee \psi}$$

$$\vee I : \frac{\psi}{\phi \vee \psi} \quad \rightarrow I : \frac{\lceil \phi \dots \psi \rfloor}{\phi \rightarrow \psi} \quad \neg E : \frac{\neg\neg\phi}{\phi} \quad \neg I : \frac{\lceil \phi \dots \bot \rfloor}{\neg\phi}$$

$$\vee E : \frac{\phi \vee \psi, \lceil \phi \dots \chi \rfloor, \lceil \psi \dots \chi \rfloor}{\chi} \quad \rightarrow E : \frac{\phi, \phi \rightarrow \psi}{\psi}$$

where $\lceil \zeta, \dots \rfloor$ is a (sub-)derivation with ζ referred to as the *hypothesis*. $\neg I$ is also called Reduction ad Absurdum (RA).

Appendix: Proof of theorem 1

We will use the following lemma:

Lemma 1 *For any theory $T \subseteq \mathcal{L}$ and for any set of sentences $\Delta \subseteq \mathcal{L}$ such that $T \cup \Delta$ is directly consistent,*
if $NACC^T(\{\phi\}, \Delta)$ holds
then there exists a RAND derivation of $\neg\phi$ from $T \cup \Delta$.

Proof of lemma 1: We use induction on the number of iterations of the $\mathcal{N}_T$ operator whose least fixed point defines $NACC^T$ (see definition 5).

Base Case: $NACC^T(\{\phi\}, \Delta)$ holds at the first iteration of $\mathcal{N}_T$. Then, there exists A such that A attacks $\{\phi\}$ (namely $T \cup A \cup \{\phi\} \vdash_{MRA} \bot$) and $A \subseteq \Delta \cup \{\phi\}$. Thus, $T \cup \Delta \cup \{\phi\} \vdash_{MRA} \bot$ and, trivially, there exists a RAND derivation $\lceil \phi \dots \bot \rfloor$ (with no RAND sub-derivations) of $\neg\phi$ from $T \cup \Delta$.

Induction Hypothesis: For any $\psi \in \mathcal{L}$, for any $\mathcal{E}$ such that $T \cup \mathcal{E}$ is directly consistent, if $NACC^T(\{\psi\}, \mathcal{E})$ holds after k iterations of $\mathcal{N}_T$, then there exists a RAND derivation of $\neg\psi$ from $T \cup \mathcal{E}$.

Inductive Step: Assume $NACC^T(\{\phi\}, \Delta)$ holds after $k+1$ iterations of $\mathcal{N}_T$, for some Δ such that $T \cup \Delta$ is directly consistent. Then there exists A such that

(i) A attacks $\{\phi\}$ (namely $T \cup A \cup \{\phi\} \vdash_{MRA} \bot$), but $A \not\subseteq \Delta \cup \{\phi\}$; and

(ii) for each defence D against A, $NACC^T(D, \Delta \cup \{\phi\})$ holds after k iterations of $\mathcal{N}_T$.

Since $A \not\subseteq \Delta \cup \{\phi\}$, $A \neq \{\}$. Also, by compactness of $\vdash_{MRA}$ (holding by compactness of $\vdash$), we can assume that A is finite. Let $A = \{\psi_1, \dots, \psi_n\}$. Then, $D_i = \{\neg\psi_i\}$, for any $i = 1, \dots, n$, is a defence against A and hence satisfies property (ii) above, i.e. $NACC^T(D_i, \Delta \cup \{\phi\})$ holds after k iterations. Note that $T \cup \Delta \cup \{\phi\}$ is directly consistent, as otherwise Δ attacks $\{\phi\}$ wrt T and $NACC^T(\{\phi\}, \Delta)$ would hold at the first iteration.

Hence, by the induction hypothesis, there exists a RAND derivation of $\neg\neg\psi_i$, for any $i = 1, \dots, n$. from $T \cup \Delta \cup \{\phi\}$. We can construct a RAND derivation, $\mathtt{d}$, of $\neg\phi$ from $T \cup \Delta$, with top derivation $\mathtt{d} : \lceil \phi \dots \bot \rfloor$ using the RAND derivations of $\neg\neg\psi_i$ from $T \cup \Delta \cup \{\phi\}$ as child sub-derivations. Note that in the top derivation we can use the $\neg E$ rule to derive ψ_i from each $\neg\neg\psi_i$, and hence, by definition of the attack A, the derivation $\mathtt{d}$ indeed leads directly to inconsistency from $T \cup \Delta$.

The resulting $\mathtt{d}$ is a RAND derivation of $\neg\phi$ from $T \cup \Delta$ as any use of ϕ in the sub-derivations of $\neg\neg\psi_i$ from $T \cup \Delta \cup \{\phi\}$ can now be replicated using the copy operation of ϕ from the top derivation $\mathtt{d}$. QED

To prove the theorem, assume now that $NACC^T(\{\phi\}, \{\})$ holds. Directly from lemma 1 with $\Delta = \{\}$, if T is directly consistent then there is a RAND derivation $\mathtt{d}$ of $\neg\phi$ from T.

[8] Note that in most argumentation approaches when priority information is missing from an argumentation theory this generally gives a form of non-determinism preventing to draw sceptical conclusions: arguments and counter-arguments have the same strength and hence the argumentation reasoning cannot arbitrate between them.

References

Amgoud, L., and Vesic, S. 2010. Handling inconsistency with preference-based argumentation. In Deshpande, A., and Hunter, A., eds., *SUM*, volume 6379 of *Lecture Notes in Computer Science*, 56–69. Springer.

Besnard, P., and Hunter, A. 2008. *Elements of Argumentation*. MIT Press.

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

Brewka, G., and Eiter, T. 2000. Prioritizing default logic. In Hölldobler, S., ed., *Intellectics and Computational Logic*, volume 19 of *Applied Logic Series*, 27–45. Kluwer.

Brewka, G. 1994. Reasoning about priorities in default logic. In Hayes-Roth, B., and Korf, R. E., eds., *AAAI*, 940–945. AAAI Press / The MIT Press.

Caminada, M. W. A., and Gabbay, D. M. 2009. A logical account of formal argumentation. *Studia Logica* 93(2-3):109–145.

Dagan, I.; Glickman, O.; and Magnini, B. 2006. The pascal recognizing textual entailment challenge. In *Machine Learning Challenges, LNCS 3944*, 177–190.

Delgrande, J. P.; Schaub, T.; Tompits, H.; and Wang, K. 2004. A classification and survey of preference handling approaches in nonmonotonic reasoning. *Computational Intelligence* 20(2):308–334.

Diakidoy, I.; Kakas, A.; Michael, L.; and Miller, R. 2013. Narrative text comprehension: From psychology to ai. In *The 11th International Symposium on Logical Formalizations of Commonsense Reasoning, Proceedings*.

Dimopoulos, Y.; Michael, L.; and Athienitou, F. 2009. Ceteris paribus preference elicitation with predictive guarantees. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI'09)*, 1890–1895.

Doppa, J.; Sorower, M.; Nasresfahani, M.; Irvine, J.; Orr, W.; Dietterich, T.; Fern, X.; and Tadepalli, P. 2011. Learning rules from incomplete examples via implicit mention models. In *Proceedings of the Asian Conference on Machine Learning, JMLR: Workshop and Conference Proceedings 20*, 197–212.

Dung, P. M.; Kakas, A. C.; and Mancarella, P. 1992. Negation as failure revisited. In *Technical Report, University of Pisa*.

Dung, P. M.; Thang, P. M.; and Toni, F. 2008. Towards argumentation-based contract negotiation. In Besnard, P.; Doutre, S.; and Hunter, A., eds., *Proceedings of the Second International Conference on Computational Models of Argument (COMMA'08)*, volume 172 of *Frontiers in Artificial Intelligence and Applications*, 134–146. IOS Press.

Dung, P. 1995. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. *Artif. Intell.* 77:321–357.

Kakas, A. C., and Moraitis, P. 2003. Argumentation based decision making for autonomous agents. In *The Second International Joint Conference on Autonomous Agents & Multiagent Systems, AAMAS 2003, July 14-18, 2003, Melbourne, Victoria, Australia, Proceedings*, 883–890. ACM.

Kakas, A. C.; Mancarella, P.; and Dung, P. M. 1994. The acceptability semantics for logic programs. In *ICLP*, 504–519.

Kakas, A.; Toni, F.; and Mancarella, P. 2012. Argumentation logic. Technical report, Department of Computer Science, University of Cyprus, Cyprus.

Kowalski, R. A., and Toni, F. 1996. Abstract argumentation. *Artificial Intelligence and Law* 4(3–4):275–296.

Lin, F., and Shoham, Y. 1989. Argument systems: A uniform basis for nonmonotonic reasoning. In Brachman, R. J.; Levesque, H. J.; and Reiter, R., eds., *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning (KR'89). Toronto, Canada, May 15-18 1989*, 245–255. Morgan Kaufmann.

McCarthy, J. 1980. Circumscription - a form of non-monotonic reasoning. *Artificial Intelligence* 13(1-2):27–39.

Michael, L. 2009. Reading between the lines. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI'09)*, 1525–1530.

Michael, L. 2010. Partial observability and learnability. *Artificial Intelligence* 174(11):639–669.

Michael, L. 2011. Causal learnability. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI'11)*, 1014–1020.

Modgil, S., and Prakken, H. 2012. A general account of argumentation with preferences. *Artificial Intelligence*. In Press.

Mueller, E. 2002. Story understanding. In *Encyclopedia of Cognitive Science*. Macmillan Reference.

Reiter, R. 1980. A logic for default reasoning. *Artificial Intelligence* 13(1-2):81–132.

Wu, Y., and Caminada, M. 2010. A labelling-based justification status of arguments. *Studies in Logic* 3(4):12–29.

Zhang, X.; Zhang, Z.; Xu, D.; and Lin, Z. 2010. Argumentation-based reasoning with inconsistent knowledge bases. In *Canadian Conference on AI*, 87–99.