Argumentation for Propositional Logic and Nonmonotonic Reasoning

Antonis Kakas, Francesca Toni, and Paolo Mancarella

[1] University of Cyprus, Cyprus
antonis@cs.ucy.ac.cy
[2] Imperial College London, UK
ft@imperial.ac.uk
[3] 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.

1 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 [16, 10, 3]. 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 [14, 9]). 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 trivialize the whole reasoning with that theory.

272

The main motivation for studying this argumentation perspective on logical reasoning is to examine 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.

This paper is an extract of [13].

2   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.[4] 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):

$$
\begin{array}{ll}
\lceil \alpha \wedge \beta & \text{hypothesis} \\
\alpha & \wedge E \\
\alpha \rightarrow (\beta \rightarrow \gamma) & \text{from } T \\
\beta \rightarrow \gamma & \rightarrow E \\
\beta & \wedge E \\
\gamma \rfloor & \rightarrow E \\
\alpha \wedge \beta \rightarrow \gamma & \rightarrow I
\end{array}
$$

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

$$
\begin{array}{ll}
\lceil \alpha & \text{hypothesis} \\
\alpha \vee \beta & \vee I \\
\neg(\alpha \vee \beta) & \text{from } T \\
\bot \rfloor & \wedge I \\
\neg\alpha & \text{RA}
\end{array}
$$

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 $T_2 \not\vdash_{MRA} \neg\alpha$.

---

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

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:

$$
\begin{array}{ll}
\lceil \alpha & \text{hypothesis} \\
\alpha \to \bot & \text{from } T \\
\bot \rfloor & \to E \\
\neg\alpha & \text{RA} \\
\quad \lceil \neg\alpha & \text{hypothesis} \\
\quad \neg\alpha \to \bot & \text{from } T \\
\quad \bot \rfloor & \to E \\
\alpha & \text{RA} \\
\bot & \wedge I
\end{array}
$$

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:

$$
\begin{array}{ll}
\alpha & \text{from } T \\
\neg\alpha & \text{from } T \\
\bot & \wedge I
\end{array}
$$

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*

$$
\begin{array}{ll}
\lceil \phi & \text{hypothesis} \\
\vdots & \vdots \\
\bot \rfloor & \vdots \\
\neg\phi & RA
\end{array}
$$

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[5]

$$
\begin{array}{ll}
\lceil \alpha & \qquad\qquad \lceil \neg\alpha \\
\alpha \to \bot & \qquad\qquad \neg\alpha \to \bot \\
\bot \rfloor & \qquad\qquad \bot \rfloor \\
\neg\alpha & \qquad\qquad \alpha
\end{array}
$$

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

---

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

3   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 $d = T \cup \Gamma$ defends/is a defence against $a = T \cup \Delta$ we will say that Γ *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 $T_1 = \{\alpha \rightarrow (\beta \rightarrow \gamma)\}$ 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[6] 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, $\{\neg\alpha\}$

---

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

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.

4 Argumentation Logic

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

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 $\mathcal{A}_T: \mathcal{R} \to \mathcal{R}$ *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,*
    - $b \nsubseteq a_0 \cup a$*, 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 $\mathcal{N}_T : \mathcal{R} \to \mathcal{R}$ *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*
  - $a \nsubseteq a_0$*, 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 $a, a_0 \in Args^T$:

276

- $ACC^T(a, a_0)$, iff
  - $a \subseteq a_0$, or
  - for all $b \in Args^T$ such that b attacks a:
    - $b \nsubseteq a_0 \cup a$, 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
  - $a \nsubseteq a_0$ 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 choose to give the definition of non-acceptability explicitly to aid readibility. We will use these two versions of non-acceptability interchangeably.

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.

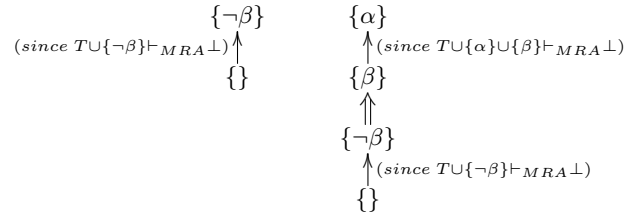The following examples illustrate non-acceptability.

$$
\begin{array}{cc}
\{\neg\beta\} & \{\alpha\} \\
{\scriptstyle(since\ T\cup\{\neg\beta\}\vdash_{MRA}\perp)}\uparrow & \uparrow{\scriptstyle(since\ T\cup\{\alpha\}\cup\{\beta\}\vdash_{MRA}\perp)} \\
\{\} & \{\beta\} \\
& \Uparrow \\
& \{\neg\beta\} \\
& \uparrow{\scriptstyle(since\ T\cup\{\neg\beta\}\vdash_{MRA}\perp)} \\
& \{\}
\end{array}
$$

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

Example 4. Let $T = \{\alpha \wedge \beta \to \perp, \neg\beta \to \perp\}$. 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)[7], since $\{\neg\beta\} \nsubseteq \{\}$, $b = \{\}$ attacks $\{\neg\beta\}$

---

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

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 $\{\} \subseteq \{\alpha, \neg\beta\}$, it follows that $NACC^T(\{\neg\beta\}, \{\alpha\})$ holds as required.

Note that if an argument a is attacked by the empty argument, then it is acceptable wrt 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 $T = T_1 = \{\alpha \to \bot, \neg\alpha \to \bot\}$ 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 $T = \{\alpha \wedge \neg\beta \to \bot, \beta \wedge \gamma \to \bot, \alpha \wedge \beta \wedge \neg\gamma \to \bot\}$. 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:
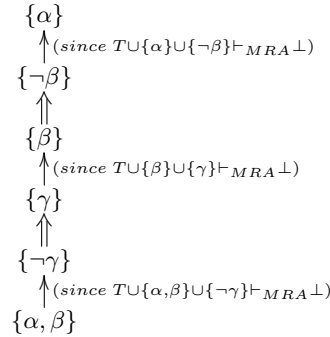
$$\{\alpha\}$$
$$\uparrow (\text{since } T \cup \{\alpha\} \cup \{\neg\beta\} \vdash_{MRA} \bot)$$
$$\{\neg\beta\}$$
$$\Uparrow$$
$$\{\beta\}$$
$$\uparrow (\text{since } T \cup \{\beta\} \cup \{\gamma\} \vdash_{MRA} \bot)$$
$$\{\gamma\}$$
$$\Uparrow$$
$$\{\neg\gamma\}$$
$$\uparrow (\text{since } T \cup \{\alpha, \beta\} \cup \{\neg\gamma\} \vdash_{MRA} \bot)$$
$$\{\alpha, \beta\}$$

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

- 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 $NACC^T(\{\beta\}, \{\alpha\})$ it suffices to prove $NACC^T(\{\neg\gamma\}, \{\alpha, \beta\})$;
- since $\{\neg\gamma\} \not\subseteq \{\alpha, \beta\}$, $b'' = \{\alpha, \beta\}$ attacks $\{\neg\gamma\}$ and $b'' \subseteq \{\alpha, \beta, \neg\gamma\}$, $NACC^T(\{\neg\gamma\}, \{\alpha, \beta\})$ and so $NACC^T(\{\beta\}, \{\alpha\})$ and $NACC^T(\{\alpha\}, \{\})$ hold.

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

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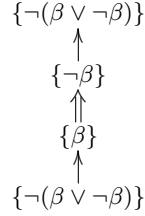
$$\{\neg(\beta \lor \neg\beta)\}$$
$$\uparrow$$
$$\{\neg\beta\}$$
$$\Uparrow$$
$$\{\beta\}$$
$$\uparrow$$
$$\{\neg(\beta \lor \neg\beta)\}$$

Fig. 3. Illustration of $NACC^T(\{\neg(\beta \lor \neg\beta)\}, \{\})$ for example 7.

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.

5 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 $T \models_{AL} \phi$.*

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.[8]

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.*[9]
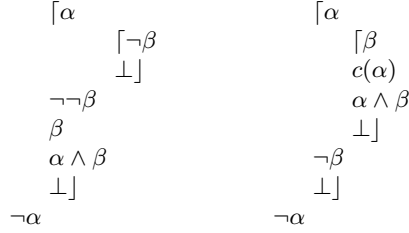

$$
\begin{array}{ll}
\lceil \alpha & \lceil \alpha \\
\quad \lceil \neg\beta & \quad \lceil \beta \\
\quad \perp \rfloor & \quad c(\alpha) \\
\neg\neg\beta & \quad \alpha \wedge \beta \\
\beta & \quad \perp \rfloor \\
\alpha \wedge \beta & \neg\beta \\
\perp \rfloor & \perp \rfloor \\
\neg\alpha & \neg\alpha
\end{array}
$$

Fig. 4. Two RAND derivations of $\neg\alpha$ in example 4: d_1 (left) and d_2 (right).

For example, the RAND derivation corresponding to the proof of $NACC^T(\{\alpha\}, \{\})$ in figure 1 is d_1 in figure 4.[10] Here, the inner RAND derivation in d_1 corresponds to the non-acceptability of the defence $\{\neg\beta\}$ against the attack $\{\beta\}$ against $\{\alpha\}$. Derivation 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, d_2 does not identify a useful attack, that cannot be defenced against, for proving non-acceptability).

6   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 $ACC^T(\{\phi\}, \{\})$ 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

---

[8] The other half of this result shows how (under some conditions) a RAND derivation of $\neg\phi$ implies $NACC^T(\{\phi\}, \{\})$, proven in [12].

[9] The proof of this theorem is included in [13].

[10] 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.

inconsistent: contradiction. Hence $ACC^T(\{\phi\}, \{\})$ holds.   QED

Thus, in Propositional Logic, trivially AL-entailment reduces to the notion of 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 $T \not\models_{AL} \alpha \to \beta$ by showing that $NACC^T(\{\neg(\alpha \to \beta)\}, \{\})$ does not hold. A standard ND derivation of $\alpha \to \beta$ from T is:

$$
\begin{array}{ll}
\lceil \alpha & \\
\quad \lceil \neg\beta & \\
\quad c(\alpha) & \\
\quad \neg\alpha & \text{from } T \\
\quad \bot \rfloor & \\
\neg\neg\beta & \text{RA} \\
\beta \rfloor & \neg E \\
\alpha \to \beta & \to I
\end{array}
$$

This does not help with determining $NACC^T(\{\neg(\alpha \to \beta)\}, \{\})$. 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 $\neg\neg(\alpha \to \beta)$ (and hence of $\alpha \to \beta$) from this theory, T, contains such a RAND sub-derivation relying on the inconsistency of the copy of α from a ($\to$I) sub-derivation, with $\neg\alpha$ from T. This means that $NACC^T(\{\neg(\alpha \to \beta)\}, \{\})$ cannot hold, since, otherwise, by theorem 1, we would have a RAND derivation of $\neg\neg(\alpha \to \beta)$ 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 $\neg(\alpha \to \beta)$.

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

7   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:

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

Example 9. Let $T = \{\phi_{bf}, \phi_{pb}, \phi_{\neg f}\}$ (T is classically consistent). It is easy to see that $T \models_{AL} \neg bird(tweety)$ as $\{\}$ attacks $\{bird(tweety)\}$ and therefore $NACC^T(\{bird(tweety)\}, \{\})$. Similarly, $T \models_{AL} \neg penguin(tweety)$. In absence of other information, we believe that these conclusions are legitimate/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 [19] 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). As in example 9, $T \models_{AL} \neg bird(tweety)$ and $T \models_{AL} \neg penguin(tweety)$. From a commonsense reasoning perspective, 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 [4, 5] and other approaches to supporting reasoning with priorities [7]. 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. [14, 18], 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 definition of (non-)acceptability.

8   Related Work

AL is based on a notion of acceptability of arguments which is in the same spirit as that in [8, 11] 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 [6, 20] 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 [17] the theory is that of classical logic but a complex prescription of model selection is imposed on top of the classical reasoning.

Recently, [2] 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 [1], studying the problem of handling inconsistency using argumentation with priorities over sentences, and [21], who have adapted the approach of [2] 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.

9   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 [12].

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. [4, 5] for Default Logic [19]. Also many frameworks of argumentation rely on some form of preference between arguments, e.g. [14, 15, 18] 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.

A   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).

References

1. Amgoud, L., Vesic, S.: Handling inconsistency with preference-based argumentation. In: Deshpande, A., Hunter, A. (eds.) SUM. Lecture Notes in Computer Science, vol. 6379, pp. 56–69. Springer (2010)
2. Besnard, P., Hunter, A.: Elements of Argumentation. MIT Press (2008)
3. Bondarenko, A., Dung, P.M., Kowalski, R.A., Toni, F.: An abstract, argumentation-theoretic approach to default reasoning. Artificial Intelligence 93(1–2), 63–101 (1997)
4. Brewka, G.: Reasoning about priorities in default logic. In: Hayes-Roth, B., Korf, R.E. (eds.) AAAI. pp. 940–945. AAAI Press / The MIT Press (1994)
5. Brewka, G., Eiter, T.: Prioritizing default logic. In: Hölldobler, S. (ed.) Intellectics and Computational Logic. Applied Logic Series, vol. 19, pp. 27–45. Kluwer (2000)
6. Caminada, M.W.A., Gabbay, D.M.: A logical account of formal argumentation. Studia Logica 93(2-3), 109–145 (2009)
7. Delgrande, J.P., Schaub, T., Tompits, H., Wang, K.: A classification and survey of preference handling approaches in nonmonotonic reasoning. Computational Intelligence 20(2), 308–334 (2004)
8. Dung, P.M., Kakas, A.C., Mancarella, P.: Negation as failure revisited. In: Technical Report, University of Pisa (1992)
9. Dung, P.M., Thang, P.M., Toni, F.: Towards argumentation-based contract negotiation. In: Besnard, P., Doutre, S., Hunter, A. (eds.) Proceedings of the Second International Conference on Computational Models of Argument (COMMA'08). Frontiers in Artificial Intelligence and Applications, vol. 172, pp. 134–146. IOS Press (2008)

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

11. Kakas, A.C., Mancarella, P., Dung, P.M.: The acceptability semantics for logic programs. In: ICLP. pp. 504–519 (1994)

12. Kakas, A., Toni, F., Mancarella, P.: Argumentation logic. Tech. rep., Department of Computer Science, University of Cyprus, Cyprus (April 2012)

13. Kakas, A., Toni, F., Mancarella, P.: Argumentation for propositional logic and nonmonotonic reasoning. In: Working notes of the 11th International Symposium on Logical Formalizations of Commonsense Reasoning (2013)

14. Kakas, A.C., Moraitis, P.: 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. pp. 883–890. ACM (2003)

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

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

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

18. Modgil, S., Prakken, H.: A general account of argumentation with preferences. Artificial Intelligence (2012), in Press

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

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

21. Zhang, X., Zhang, Z., Xu, D., Lin, Z.: Argumentation-based reasoning with inconsistent knowledge bases. In: Canadian Conference on AI. pp. 87–99 (2010)