Argumentation and Propositional Logic

Antonis Kakas[1], Francesca Toni[2] and Paolo Mancarella[3]

[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 Artificial Intelligence (AI). Argumentation is also close to the original inception of logic as a framework for formalizing human debate and dialogue. The purpose of this paper is to draw a formal connection between argumentation and classical reasoning, as supported by Propositional Logic. To this effect, we propose Argumentation Logic and show properties 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 [7, 3, 1]. 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 [6, 2]). 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 (PL). 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.

2 Preliminaries on Natural Deduction

Let $\mathcal{L}$ be a PL language and $\vdash$ denote the provability relation of ND in PL.[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) without 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 $T \vdash_{MRA} \phi$.*

[1] See appendix A for a review of the ND rules we use, including $\neg I$/Reduction ad Absurdum (RA).

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/directly consistent *iff it is not classically/directly, respectively, inconsistent.*

Trivially, if a theory is classically consistent then it is directly consistent. However, a directly consistent theory may be classically inconsistent.

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. We will adopt the following notation:

Notation 1. *Given a RAND derivation $\mathtt{d}$ and a RAND (sub-)derivation $\mathtt{d}'$ of $\neg\phi$ in $\mathtt{d}$ (possibly $\mathtt{d}' = \mathtt{d}$), $\mathtt{d}'$ is denoted by $\lceil \phi : c(\phi_1), \ldots, c(\phi_k); \neg\psi_1, \ldots, \neg\psi_l : \bot \rfloor$ where $k, l \geq 0$ and*

- *ϕ is the hypothesis of $\mathtt{d}'$;*
- *$\forall i, j = 1, \ldots k$, if $i \neq j$ then $\phi_i \neq \phi_j$; ϕ_i is the hypothesis of an ancestor RAND (sub-)derivation of $\mathtt{d}'$ in $\mathtt{d}$; ϕ_i is "copied" ($c(\phi_i)$) in $\mathtt{d}'$;*
- *$\forall i, j = 1, \ldots l$, if $i \neq j$ then $\psi_i \neq \psi_j$; ψ_j is the hypothesis of a child RAND sub-derivation of $\mathtt{d}'$ in $\mathtt{d}$.*

3 Argumentation Logic Frameworks

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

Definition 3. *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 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 1. *Given $T_1 = \{\alpha \to (\beta \to \gamma)\}$, $\{\alpha, \beta\}$ 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 relation 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. 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[2] 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 1, $\{\neg\alpha\}$ 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.

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 4. *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.

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

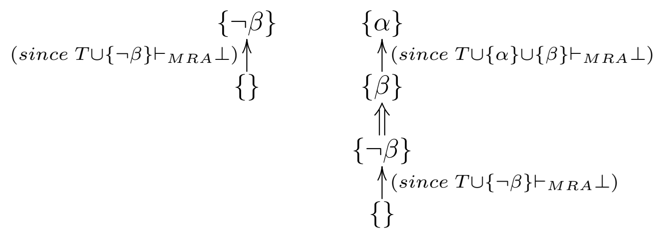

Figure 1: Illustration of $NACC^T(\{\neg\beta\}, \{\})$ (left) and $NACC^T(\{\alpha\}, \{\})$ (right), for example 2.

Definition 5. *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 $NACC^T(a, a_0)$.*

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. The following examples illustrate non-acceptability.

Example 2. *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)*[3]*, 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 3. *Given $T = \{\alpha \rightarrow \bot, \neg\alpha \rightarrow \bot\}$, $NACC^T(\{\alpha\}, \{\})$ and $NACC^T(\{\neg\alpha\}, \{\})$ both hold: $NACC^T(\{\alpha\}, \{\})$ holds as $\{\alpha\}$ is attacked by $\{\}$; $NACC^T(\{\neg\alpha\}, \{\})$ holds as $\{\neg\alpha\}$ is attacked by $\{\}$.*

The following example illustrates non-acceptability in the case of an empty theory.

Example 4. *For $T = \{\}$, $NACC^T(\{\neg(\beta \vee \neg\beta)\}, \{\})$ holds, as illustrated in figure 2. 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 6. *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.

[3]Here and throughout the paper, ↑ denotes an attack and ⇑ denotes a defence.

$$\begin{array}{c} \{\neg(\beta \vee \neg\beta)\} \\ \uparrow \\ \{\neg\beta\} \\ \Uparrow \\ \{\beta\} \\ \uparrow \\ \{\neg(\beta \vee \neg\beta)\} \end{array}$$

Figure 2: Illustration of $NACC^T(\{\neg(\beta \vee \neg\beta)\}, \{\})$ for example 4.

$\mathtt{d}_1$	$\mathtt{d}_2$
$\lceil \alpha$	$\lceil \alpha$
$\quad \lceil \neg\beta$	$\quad \lceil \beta$
$\quad \bot \rfloor$	$\quad c(\alpha)$
$\neg\neg\beta$	$\quad \alpha \wedge \beta$
β	$\quad \bot \rfloor$
$\alpha \wedge \beta$	$\neg\beta$
$\bot \rfloor$	$\bot \rfloor$
$\neg\alpha$	$\neg\alpha$

Figure 3: Two RAND derivations of $\neg\alpha$ in example 2: $\mathtt{d}_1$ (left) and $\mathtt{d}_2$ (right).

5 Basic Properties of AL

The following result gives a core property of the notion of AL-entailment wrt the notion of direct derivation in PL, 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 $T \models_{AL} \phi$.*[4]

The following theorem shows how the RA rule, deleted from the ND proof system within $\vdash_{MRA}$, is brought back through the notion of non-acceptability. This theorem will be used, in section 6, to prove (one half of) the link between AL and PL.

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.*

For example, the RAND derivation corresponding to the proof of $NACC^T(\{\alpha\}, \{\})$ in figure 1 is $\mathtt{d}_1$ in figure 3. 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).

6 From AL to PL and back

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

[4] The proof of this result as well as all other omitted proofs in the paper can be found in [4] and/or in [5].

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

Thus, in PL, 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 $T \vdash \phi$.*

This corollary gives that consequences of a classically consistent theory under $\models_{AL}$ are classical consequences too. Although proposition 1 sanctions that all *direct* consequences are retrieved by $\models_{AL}$, in general not all classical consequences are retrieved by $\models_{AL}$, namely the converse of corollary 2 does not hold. For example, $\{\neg\alpha\} \not\models_{AL} \alpha \to \beta$, namely, under $\models_{AL}$, implication is not material implication. However, if we restrict attention to theories *expressed using connectives $\wedge$ and $\neg$ only* (without loss of generality), then all classical consequences that can be derived by a *special kind of RAND derivations* are retrieved by $\models_{AL}$ (see corollary 3 below). These special kinds of derivations are defined as follows:

Definition 7. *Let $\mathtt{d} = \lceil \phi : c(\phi_1), \ldots, c(\phi_k); \neg\psi_1, \ldots, \neg\psi_l : \bot \rfloor$ be a RAND (sub-)derivation from T. Then* d satisfies the genuine absurdity property *(wrt T) iff*

$$T \cup \{\phi_1, \ldots, \phi_k\} \cup \{\neg\psi_1, \ldots, \neg\psi_l\} \not\vdash_{MRA} \bot.$$

d fully satisfies the genuine absurdity property *(wrt T) iff it satisfies the genuine absurdity property (wrt T) and all its sub-derivations fully satisfy the genuine absurdity property (wrt T).*

Namely, the genuine absurdity property is satisfied by a (sub-)derivation when its hypothesis ϕ is necessary for its direct derivation of $\bot$. This property is illustrated by example 2: $\mathtt{d}_1$ and $\mathtt{d}_2$ in figure 3 are both RAND derivations of $\neg\alpha$, but only $\mathtt{d}_1$ fully satisfies the genuine absurdity property (wrt T). Indeed, in $\mathtt{d}_2$, α is not necessary in the outer RAND direct derivation of $\bot$.

Theorem 2. *Let T be a directly consistent theory and $\phi \in \mathcal{L}$, both expessed using only $\wedge$ and $\neg$. If there exists a RAND derivation of $\neg\phi$ from T that fully satisfies the genuine absurdity property (wrt T) then $NACC^T(\{\phi\}, \{\})$ holds.*

This theorem and corollary 1 imply that all classical consequences obtainable from RAND derivations fully satisfying the genuine absurdity property are retrieved by $\models_{AL}$:

Corollary 3. *Let T be a directly consistent theory expessed using only $\wedge$ and $\neg$, and $\phi \in \mathcal{L}$. If there exists a RAND derivation of $\neg\phi$ from T that fully satisfies the genuine absurdity property (wrt T) then $T \models_{AL} \phi$.*

7 Conclusions

We have presented Argumentation Logic (AL) and shown how it allows us to understand classical reasoning in PL in terms of argumentation. Its definition rests on capturing semantically the Reductio ad Absurdum (RA) rule through a suitable notion of acceptability of arguments. AL gives an alternative view of RA, in that it does not allow the use of RA when the inconsistency that it (directly) derives does not depend on the hypothesis posed when we apply the rule. As such the interpretation of implication in AL is different from that of material implication.

We have given a "weak" correspondence between AL and classical PL. However, in [5], we give further results on the relationship between AL and PL including how AL *completely* captures the entailment of PL (based on the guaranteed existence of a RAND derivations fully satisfying the genuine absurdity property given any RAND derivation for the same sentence).

A discussion on how to extend AL to capture non-monotonic reasoning can be found in [4]. Further work is needed to explore the extension of AL to capure, within the same unified setting, both classical and defeasible reasoning. AL incorporates its own mechanism for belief revision, in the presence of inconsistencies, when reasoning with directly consistent theories, as it isolates and thus "removes" inconsistencies. Another important direction for future work is the study of this form of belief revision in the context of AL.

Clearly, there is a link between AL and other logics such as Intuitionistic and Relevance Logics. We are currently exploring these possible links.

A Appendix: Natural Deduction

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

$$\wedge I : \frac{\phi, \psi}{\phi \wedge \psi} \quad \wedge E : \frac{\phi \wedge \psi}{\phi} \quad \wedge E : \frac{\phi \wedge \psi}{\psi} \quad \vee I : \frac{\phi}{\phi \vee \psi} \quad \vee I : \frac{\psi}{\phi \vee \psi} \quad \rightarrow I : \frac{\lceil \phi \ldots \psi \rfloor}{\phi \rightarrow \psi}$$

$$\neg E : \frac{\neg\neg\phi}{\phi} \quad \neg I/RA : \frac{\lceil \phi \ldots \bot \rfloor}{\neg\phi} \quad \vee E : \frac{\phi \vee \psi, \lceil \phi \ldots \chi \rfloor, \lceil \psi \ldots \chi \rfloor}{\chi} \quad \rightarrow E : \frac{\phi, \phi \rightarrow \psi}{\psi}$$

where $\lceil \zeta, \ldots \rfloor$ is a (sub-)derivation with ζ referred to as the *hypothesis*.

References

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

[2] Phan Minh Dung, Phan Minh Thang, and Francesca Toni. Towards argumentation-based contract negotiation. In Philippe Besnard, Sylvie Doutre, and Anthony Hunter, editors, *Proceedings of the Second International Conference on Computational Models of Argument (COMMA'08)*, volume 172 of *Frontiers in Artificial Intelligence and Applications*, pages 134–146. IOS Press, 2008.

[3] P.M. Dung. 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.

[4] A.C. Kakas, F. Toni, and P. Mancarella. Argumentation for propositional logic and nonmonotonic reasoning. In *The 11th International Symposium on Logical Formalizations of Commonsense Reasoning, Proceedings*, 2013.

[5] Antonis Kakas, Francesca Toni, and Paolo Mancarella. Argumentation logic. Technical report, Department of Computer Science, University of Cyprus, Cyprus, April 2012.

[6] Antonis C. Kakas and Pavlos Moraitis. 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*, pages 883–890. ACM, 2003.

[7] Fangzhen Lin and Yoav Shoham. Argument systems: A uniform basis for nonmonotonic reasoning. In Ronald J. Brachman, Hector J. Levesque, and Raymond Reiter, editors, *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning (KR'89). Toronto, Canada, May 15-18 1989*, pages 245–255. Morgan Kaufmann, 1989.