On the Relationship between Carneades and Defeasible Logic

Guido Governatori
NICTA, Queensland Research Laboratory, Australia
guido.governatori@nicta.com.au

ABSTRACT

We study the formal relationships between the inferential aspects of Carneades (a general argumentation framework) and Defeasible Logic. The outcome of the investigation is that the current proof standards proposed in the Carneades framework correspond to some variants of Defeasible Logic.

1. INTRODUCTION

Argumentation emerged as one of the mainstream topics in the field of Artificial Intelligence and Law, and then it has got recognition in the broader field of Artificial Intelligence and in that of agents. In this paper we will concentrate on two argumentation frameworks, Carneades, a general framework for argumentation and Defeasible Logic, a lightweight rule based computational approach to non-monotonic reasoning with an argumentation like flavour. While these are only two among many approaches to argumentation, they have been discussed in relationship to legal reasoning and legal applications (though they are not the only ones).

The two systems exhibit some similarities (for example they both address sceptical reasoning, and both consider different inference mechanisms). Accordingly, the aim of the paper is to establish a formal relationship between Defeasible Logic and Carneades to make a precise comparison of the features presented by the approaches possible. In addition once mappings have been established it is possible to study whether theoretical results for one system transfer to the other system. Finally, we hope that the theoretical comparison of the features presented by the systems under analysis could led us to a better understanding of some phenomena of legal reasoning.

The paper is organised as follows: in Sections 2 and 3 we quickly review the two frameworks. In Section 4 we show how to reconstruct Carneades in Defeasible Logic. Specifically we focus on the current proof standards of Carneades, and we show that these correspond to some variants of Defeasible Logic. In Section 5 we point out some arguable aspects and limitations of the current version of Carneades.

2. CARNEADES

Carneades [13] is a general framework for argumentation, and it has received considerable attention since its inception: [10] reports that as of June 2010 [13] is among the ten most cited papers from articles published in the Artificial Intelligence Journal in the previous 5 years. Carneades combines many features of argumentation to capture dynamic as well as static aspects of argumentation such as switch of burden of proof, proof standards, audience and more. The Carneades models emerged from research on argumentation in legal theory and from Walton's theory of argumentation schemes.

We follow the presentation of Carneades given in [10] with a simple propositional language consisting of literals only, that is, atomic propositions and their negation. For a literal l we use $\sim l$ to indicate the complement of l. Thus if $l = p$ then $\sim l = \neg p$ and if $l = \neg p$, then $\sim l = p$.

DEFINITION 1. *A* Carneades argument *is a structure* $\langle P, E, c\rangle$, *where P and E are disjoint sets of literals, and c is a literal.*

Given an argument $\langle P, E, c\rangle$, the sets P and E are, respectively, the set of premises and the set of exceptions. If $c = p$ then the argument is *pro p*, if $c = \sim p$ the argument is *con p*. In the dynamic (full) version of Carneades arguments are evaluated in *argument evaluation structures*, where such structures consider the stage (or phase of the life-cycle of an argumentation), an audience (where the audience is a set of weights associated to the propositions and a set of given propositions, i.e., the assumptions the evaluation depends on), and a function assigning a proof standard to each proposition involved in the arguments depending on the stage of the dispute. However, as we have pointed out in the introduction the aim of this paper is just to compare the static aspects of Carneades, thus we will abstract from stages, and audience, and we will focus on the proof standards, or in other terms, the various inference mechanisms used by Carneades to determine whether arguments and the propositions associated to them are acceptable or not.

DEFINITION 2. *A* Carneades Argument Evaluation Structure *(CAES) is a structure* $\langle Arg, Ass, W, PS\rangle$ *where*

- *Arg is an acyclic*[1] *set of arguments;*
- *Ass is a consistent set of literals;*
- *W is a weight function* $W : Arg \mapsto [0,1]$, *assigning a real number in the interval* $[0,1]$ *to every argument;*
- *PS is a function mapping propositions to a proof standard.*

In the original version of Carneades [13], the weight function was replaced by a partial order on the arguments. The change, introduced in [14], was motivated by the need to introduce thresholds to model various proof standards such as 'clear and convincing evidence' and 'beyond reasonable doubt'.

[1] A set of arguments is acyclic if its dependency graph is. The dependency graph has a node for each propositional atom appearing in some argument. Furthermore, there is a link from q to p whenever p depends on q, that is, whenever there is an argument pro or con p with q or $\neg q$ in its set of premises or exceptions.

ICAIL'11 June 6–10, 2011, Pittsburgh, PA
Copyright 2011 ACM 978-1-4503-0755-0/11/06 ...$10.00.

The acceptability of a literal depends on the proof standard determining the strength of the derivation of the literal. To derive a conclusion Carneades needs to establish whether arguments are applicable or not.

DEFINITION 3. *An argument* $a = \langle P,E,c\rangle$ *is* applicable *in a CAES* $S = \langle Arg,Ass,W,PS\rangle$ *iff,* $a \in Arg$ *and*

- $p \in P$ *implies* $p \in Ass$ *or* ($\sim p \notin Ass$ *and* p *is acceptable in* S)*;*
- $p \in E$ *implies* $p \notin Ass$ *and* ($\sim p \in Ass$ *or* p *is not acceptable in* S)*.*

We have now to define when a literal is acceptable in a CAES

DEFINITION 4. *Given a CAES* $S = \langle Arg,Ass,W,PS\rangle$*, a literal* l *is* acceptable *in* S *iff the conditions associated to the proof standard for the literal* l *are satisfied.*

The proof standards currently defined in Carneades [14, 10] are:

- *scintilla of evidence*,
- *preponderance of evidence*, also called *best argument* in [13],
- *clear and convincing evidence*,
- *beyond reasonable doubt*,
- *dialectical validity*

The above proof standards are listed in order of strength; this means that the conditions to satisfy them are increasingly stringent. In addition a stronger proof standard includes the weaker ones.

DEFINITION 5 (SCINTILLA OF EVIDENCE). *The proof standard* scintilla of evidence *(se) for a literal* p *is satisfied iff there is at least one applicable argument for* p*.*

DEFINITION 6 (PREPONDERANCE OF EVIDENCE). *The proof standard* preponderance of evidence *(pe) for a literal* p *is satisfied iff* p *satisfies se and the maximum weight assigned to an applicable argument pro* p *is greater than the maximum weight assigned to an applicable argument con* p*.*

DEFINITION 7 (CLEAR AND CONVINCING EVIDENCE). *The proof standard* clear and convincing evidence *(ce) for a literal* p *is satisfied iff* p *satisfies pe and the maximum weight of applicable pro arguments exceeds some threshold* α*, and the difference between the maximum weight of the applicable pro arguments and the maximum weight of the applicable con arguments exceeds some threshold* β*.*

DEFINITION 8 (BEYOND REASONABLE DOUBT). *The proof standard* beyond reasonable doubt *(bd) for a literal* p *is satisfied iff* p *satisfies ce and the maximum weight of the applicable con arguments is less that some threshold* γ*.*

DEFINITION 9 (DIALECTICAL VALIDITY). *The proof standard* dialectical validity *(dv) for a literal* p *is satisfied iff* p *satisfies se and* $\sim p$ *does not satisfy se.*

The proof standard *dialectical validity* can also be rewritten as "there is an applicable argument pro p and there is no argument con p".

The proof standards *scintilla of evidence*, *best argument* and *dialectical validity* were defined in [13] where it was claimed that *scintilla of evidence* is weaker than *best argument* and *best argument* is weaker than *dialectical validity*.

3. DEFEASIBLE LOGIC

Defeasible Logic [31] is a simple, efficient and flexible rule based approach to sceptical non-monotonic reasoning. In this paper we take as starting point the formalisation of the logic given in [3]. The logic has been generalised to provide a framework that covers several variants [5, 2] to capture the intuition behind different facets on non-monotonic reasoning. Over the years the logic has been thoroughly investigated [3, 4, 8, 28], relationships with other formalisms established [6, 4], and it can be characterised by argumentation semantics [17, 18]. Furthermore, Defeasible Logic has been proposed as a tractable computational approach for applications in the legal domain: modelling of contracts [16], representation of normative positions with time [22], modelling of norm dynamics [21, 20], and regulatory compliance [23, 34]. In addition several efficient implementations have been developed [30, 1, 7, 26].

The aim of this section is to give an introduction to the Defeasible Logic variants needed for the reconstruction of Carneades. Defeasible Logic has three kinds of rules, strict rules (for the monotonic part), defeasible rules and defeaters (for the non-monotonic part); for the sake of simplicity we restrict the presentation to defeasible rules only (this is not a real limitation, since it is possible to remove strict rules and defeaters from the non-monotonic part without affecting the set of non-monotonic consequences that can be derived [3]).

A *defeasible theory* is a structure $D = (F,R,>)$ where F is a set of facts, represented by literals, R is a set of rules, and $>$, the superiority relation, is a binary relation establishing the relative strength of rules. Thus given two rules, let us say r_1 and r_2, $r_1 > r_2$ means that r_1 is stronger than r_2, thus r_1 can override the conclusion of r_2. A rule r is an expression $A \Rightarrow c$, where A, the antecedent or body of the rule is a (possible empty) set of literals, and r, the conclusion or head of the rule, is a literal. For a literal p, the set of rules whose head is p is denoted by $R[p]$. Given a rule r we will use $A(r)$ to identify its body and $C(r)$ for the conclusion.

One of the aims of the paper is to reconstruct the proof standards currently proposed by Carneades, and eventually to point out shortcomings and propose enhancements to them. Essentially the proof standards assess the degree of uncertainty or ambiguity of a conclusion. Thus before continuing the technical presentation of Defeasible Logic we take a brief intermezzo to discuss the notion of ambiguous literal in non-monotonic reasoning.

The treatment of ambiguity is one of the aspects of non-monotonic reasoning where intuitions can clash [35, 25]. Intuitively, a literal is *ambiguous* if there is a (monotonic) chain of reasoning pro p, and another con p (or alternatively pro $\sim p$), and the superiority relation does not resolve this conflict. The first distinction is between sceptical and credulous reasoning. Sceptical reasoning refrains from concluding p and $\neg p$, while credulous reasoning explores two different alternative and incompatible sets of conclusions, one with p and the other with $\neg p$. In Defeasible Logic this distinction is captured by δ and ∂ for sceptical reasoning and σ for credulous reasoning. In sceptical reasoning we have a further distinction: ambiguities are local or ambiguities are global. In the first case we speak of ambiguity blocking, in the second of ambiguity propagation. In an ambiguity propagation setting, since we were not able to solve the conflict we want to propagate the uncertainty to conclusions depending on the ambiguous literals. In an ambiguity blocking setting, given the sceptical nature of the reasoning, the two conclusions are considered both as not provable, and we ignore the reasons why they were when we use them as premises of further arguments.

EXAMPLE 1. Consider a theory with the following rules:

$$
\begin{array}{ll}
r_1 : \Rightarrow a & \\
r_2 : \Rightarrow \neg a & r_3 : \neg a \Rightarrow b \\
 & r_4 : \Rightarrow \neg b
\end{array}
$$

Clearly a is ambiguous. There is a chain of reasoning leading to it, and a chain of reasoning leading to its complement, i.e., $\neg a$. What about b and $\neg b$. Are they ambiguous? Definitely there is a chain of reasoning leading to $\neg b$, but for b we have two options. Using ambiguity propagation, we can argue that there is a chain of reasoning for it, since there is a rule for it, and for every premise of the rule, there is a chain of reasoning for the premise. On the contrary, using ambiguity blocking, we argue that there is no chain of reasoning for b, since we are not able to solve the conflict involving some of its premises.

A *conclusion* of D is a tagged literal and can have one of the following forms[2]:

1. $+\partial p$: p is defeasibly provable in D using the ambiguity blocking variant;
2. $-\partial p$: p is defeasibly rejected in D using the ambiguity blocking variant;
3. $+\delta p$: p is defeasibly provable in D using the ambiguity propagation variant;
4. $-\delta p$: p is defeasibly rejected in D using the ambiguity propagation variant;
5. $+\sigma p$: p is supported in D, i.e., there is a chain of reasoning leading to p;
6. $-\sigma p$: p is not supported in D.

The proof tags determine the strength of a derivation. The proof tags $+\delta$, $-\delta$, $+\partial$ and $-\partial$ are for sceptical conclusions, and $+\sigma$ and $-\sigma$ capture credulous conclusions.

A proof (or derivation) P is a finite sequence $(P(1),\ldots,P(n))$ of tagged literals, satisfying the proof conditions (corresponding to inference conditions) presented in the rest of this section. The proof conditions, given a derivation $P(1),\ldots,P(n)$, describe the conditions under which we can extend the derivation to derive the tagged literal s $P(n+1)$. We use the notation $P(1..n)$ to denote the sequence of length n of a derivation P.

We are now ready to give the proof conditions for the proof tags listed above. We start with that for ambiguity blocking, i.e., ∂.

$+\partial$: If $P(n+1) = +\partial q$ then either
(1) $q \in F$ or
(2)(2.1) $\exists r \in R[q] \forall a \in A(r):\ +\partial a \in P(1..n)$ and
(2.2) $\sim q \notin F$ and
(2.3) $\forall s \in R[\sim q]$ either
(2.3.1) $\exists a \in A(s):\ -\partial a \in P(1..n)$ or
(2.3.2) $\exists t \in R[q]$ such that
$\forall a \in A(t):\ +\partial a \in P(1..n)$ and $t > s$.

$-\partial$: If $P(n+1) = -\partial q$ then
(1) $q \notin F$ and
(2) (2.1) $\forall r \in R[q]\ \exists a \in A(r):\ -\partial a \in P(1..n)$ or
(2.2) $\sim q \in F$ or
(2.3) $\exists s \in R[\sim q]$ such that
(2.3.1) $\forall a \in A(s):\ +\partial a \in P(1..n)$ and
(2.3.2) $\forall t \in R[q]$ either
$\exists a \in A(t):\ -\partial a \in P(1..n)$ or $t \not> s$.

The main idea of the conditions for a defeasible proof ($+\partial$) is that there is an applicable rule, i.e., a rule whose all antecedents are already defeasibly provable and for every rule for the opposite conclusion either the rule is discarded, i.e., one of the antecedents is not defeasibly provable, or the rule is defeated by a stronger applicable rule for the conclusion we want to prove. The conditions for $-\partial$ show that any systematic attempt to defeasibly prove that the conclusion fails. Notice that the above conditions characterise the notion of sceptical conclusion using ambiguity blocking [5, 2, 8].

[2]In the rest of the paper we will introduce some variants of the proof tags to capture Carneades proof standards.

EXAMPLE 1 (CONTINUED). For a and $\neg a$ clause (2.3) of $-\partial$ applies: the antecedents of r_1 and r_2 are trivially satisfy (2.3.1), similarly for clause (2.3.2) given that the superiority relation is empty. Accordingly, we have $-\partial a$ and $-\partial\neg a$. This makes clause (2.1) satisfied for b and rule r_3, thus we have $-\partial b$, and simultaneously clause (2.3.1) of $+\partial$ is satisfied for $\neg b$, thus, given that clause (2.1) holds vacuously we obtain $+\partial b$.

We pass now to the definition of the proof conditions for ambiguity blocking, i.e., δ.

$+\delta$: If $P(n+1) = +\delta q$ then
(1) $q \in F$ or
(2) (2.1) $\sim q \notin F$ and
(2.2) $\exists r \in R[q] \forall a \in A(r):\ +\delta a \in P(1..n)$ and
(2.3) $\forall s \in R[\sim q]$ either
(2.3.1) $\exists a \in A(s):\ -\sigma a \in P(1..n)$ or
(2.3.2) $\exists t \in R[q]$ such that
$\forall a \in A(t):\ +\delta a \in P(1..n)$ and $t > s$.

$-\delta$: If $P(n+1) = -\delta q$ then
(1) $q \notin F$ and
(2) (2.1) $\sim q \in F$ or
(2.2) $\forall r \in R[q]$ either $\exists a \in A(r):\ -\delta a \in P(1..n)$ or
(2.3) $\exists s \in R[\sim q]$ such that
(2.3.1) $\forall a \in A(s):\ +\sigma a \in P(1..n)$ and
(2.3.2) $\forall t \in R[q]$ either
$\exists a \in A(t):\ -\delta a \in P(1..n)$ or $t \not> s$.

The proof tags $+\delta$ and $-\delta$ capture defeasible provability using ambiguity propagation [5, 2, 8]. Their explanation is similar to that of $+\partial$ and $-\partial$. The major difference is that to prove p this time we make easier to attack it (clause 2.3). Instead of asking that the argument attacking it are justified arguments, i.e., rules whose premises are provable, we just ask for defensible arguments (i.e., credulous arguments), that is rules whose premises are just supported (i.e., there is a valid chain of reasoning leading to it). The definition of support, proof tags $+\sigma$ and $-\sigma$ is as follows:

$+\sigma$: If $P(n+1) = +\sigma q$ then either
(1) $q \in F$; or
(2) $\exists r \in R[q]$ such that
(2.1) $\forall a \in A(r), +\sigma a \in P(1..n)$, and
(2.2) $\forall s \in R[\sim q]$ either
$\exists a \in A(s), -\delta a \in P(1..n)$, or $(s \not> r)$.

$-\sigma$: If $P(n+1) = -\sigma q$ then
(1) $q \notin F$, and
(2) $\forall r \in R[q]$ either
(2.1) $\exists a \in A(r), -\sigma a \in P(1..n)$; or
(2.2) $\exists s \in R[\sim q]$ such that
$\forall a \in A(s), +\delta a \in P(1..n)$, and $s > r$.

The proof conditions above are essentially forward chaining of rules to propagate the 'support' for arguments (rules). However, we cannot propagate the support from the premises to the conclusion in case we have a rule for the complement where all the premises of the rule are provable, unless this rule is not weaker than the rule for the conclusion we want to support.

EXAMPLE 1 (CONTINUED). Similarly to the previous case, and essentially for the same reasons, we have $-\delta a$ and $-\delta\neg a$. However, in this case, since we do not have preferences over r_1 and r_2 we have $+\sigma a$ and $+\sigma\neg a$. This implies that clause (2.3.1) of $-\delta$ holds; consequently $-\delta\neg b$.

It is possible to give a weaker notion of support [19], that is, whether there is a simple forward chaining of rules leading to the conclusion

$+\sigma^-$: If $P(n+1) = +\sigma^- q$ then either

(1) $q \in F$ or
(2) $\exists r \in R[q] \forall a \in A(r) + \sigma^- a \in P(1..n)$
$-\sigma^-$: If $P(n+1) = -\sigma^- q$ then
(1) $q \notin F$ and
(2) $\forall r \in R[q] \exists a \in A(r)$ such that $-\sigma^- a \in P(1..n)$

The above proof conditions, except for σ^-, allow us to form 'teams' of rules for a conclusion to defeat other rules.

EXAMPLE 2. Consider a theory where we have the following rules

$$r_1 : \Rightarrow a \qquad r_2 : \Rightarrow a$$
$$r_3 : \Rightarrow \neg a \qquad r_4 : \Rightarrow \neg a$$

and the superiority relation is $r_1 > r_3$ and $r_2 > r_4$. All rules are applicable, so we can argue pro a using r_1, then we have to consider all possible attacks to it. r_3 is defeated by r_1 itself, and r_4 is defeated by r_2.

In case this feature, called *team defeat*, is not desired, Defeasible Logic provides variants of the proof conditions given so far to reject it. The proof conditions for the variants without team defeat can be obtained from the corresponding proof conditions given above with the following changes [8]:

- For $+\partial$ and $+\delta$, clause (2.3.2) is replaced by $r > s$; we use $+\partial^*$ and $+\delta^*$, for the proof tags thus obtained.
- For $-\partial$ and $-\delta$, clause (2.3.2) is replaced by $r \not> s$; we use $-\partial^*$ and $-\delta^*$, for the proof tags thus obtained.
- For $+\sigma$ and $-\sigma$ the occurrences of $+\delta$ and $-\delta$ must be replaced by $+\delta^*$ and $-\delta^*$, for the proof tags thus obtained we use $+\sigma^*$ and $-\sigma^*$.

Over the years properties of Defeasible Logic have been studied. Here we report some of the most important results for our purposes:

PROPOSITION 1. *[3, 4] For any theory D, proof tag # and literal l, it is not the case that $D \vdash +\#l$ and $D \vdash -\#l$.*

This means that Defeasible Logic is coherent, i.e., if you prove that a literal l is provable (with a particular strength) then you cannot prove that the same literal is rejected (with the same strength) by the theory (and the other way around).

PROPOSITION 2. *[3, 4] For any theory D where the transitive closure of the superiority relation is acyclic, literal l, and $\# \in \{\delta, \partial\}$, $D \vdash +\#l$ and $D \vdash +\#{\sim}l$ iff $l, {\sim}l \in F$.*

This means that the sceptical inference mechanism of Defeasible Logic produces an inconsistency only when the monotonic part of the theory is inconsistent.

Notice, however, that the above property does not hold for σ. σ corresponds to the existence of a credulous argument, thus it is possible to have credulous arguments for both p and $\neg p$.

PROPOSITION 3. *[8] For any theory D and literal l, the following implications hold:*

1. *$D \vdash +\delta^* l$ implies $D \vdash +\delta l$;*
2. *$D \vdash +\delta l$ implies $D \vdash +\partial l$;*
3. *$D \vdash +\partial l$ implies $D \vdash +\sigma l$;*
4. *$D \vdash +\sigma l$ implies $D \vdash +\sigma^* l$;*
5. *$D \vdash +\delta^* l$ implies $D \vdash +\partial^* l$;*
6. *$D \vdash +\partial^* l$ implies $D \vdash +\sigma^{*l}$;*
7. *$D \vdash +\sigma^* l$ implies $D \vdash +\sigma^- l$.*[3]

[3]The last three implications were not discussed in [8] but can be easily obtained with the same techniques. Given that the proof conditions for the negative proof tags are the strong negation of the positive ones [5, 2], we derive immediately the results for them as follows: if $D \vdash +\#_1 l$ implies $D \vdash +\#_2 l$, then $D \vdash -\#_2 l$ implies $D \vdash -\#_1 l$.

The above proposition tells us the relative strength of the various proof conditions. Graphically, the relationships among the various proof conditions are depicted in Figure 1.

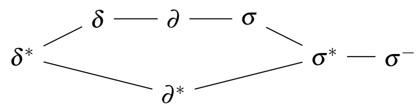

Figure 1: Relationships among the proof conditions.

Observe that the conditions for team defeat are more general than the corresponding conditions where this feature does not hold. However, when we consider the ambiguity blocking variant (i.e., ∂), the set of conclusions we can derive are in general different. This is not the case for ambiguity propagation (i.e., δ), where one set of conclusions is included in the other. From the picture in Figure 1 we can see that we have two chains of proof conditions, and that the set of conclusions we can derive from one proof tags in one chain are different from the set of conclusions we can derive from a proof tag in the other chain.

EXAMPLE 3. To see the difference among the proof tags we extend the theory in Example 2 with the following rules:

$$r_5 : a \Rightarrow b \qquad r_6 : b \Rightarrow c \qquad r_7 : \Rightarrow \neg c$$

where $r_6 > r_7$. In this theory we can prove $+\partial a$, $+\delta a$ but $-\partial^* a$ and $-\delta^* a$. Then, accordingly we have $+\partial b$ and $+\delta b$ but $-\partial^* b$ and $+\partial c$, $+\delta c$, $+\sigma c$ and $-\sigma \neg c$ but $+\partial^* \neg c$.

REMARK 1. Despite the above result, given a set of applicable rules (that is a set of rules whose premises are all applicable) the proof conditions for team defeat include those for no team defeat. Thus if we have an applicable pro rule that is stronger than any applicable con rule, then the condition for no team defeat is satisfied, and we can use that rule to instantiate the variables r, clause (2.1), and t, clause (2.3.2), in the proof conditions for the team defeat case.

The next result is about the complexity of Defeasible Logic. The original result given by Maher [28] for the ambiguity blocking (∂) variant can be easily extended to the other variants presented so far.

PROPOSITION 4. *The set of all the conclusions (extension) of a defeasible theory can be computed in time linear to the size of the theory, where the size of a theory is given by the number of symbols occurring in the theory.*

Defeasible Logic does not impose any syntactic restriction to prevent cycles in the dependency graph. Thus in theories like the theory where the only rule is $a \Rightarrow a$ we cannot obtain any conclusion about a. However, the following proposition tells us cases when this does not happen.

PROPOSITION 5. *[4] Let D be a theory where the dependency graph is acyclic, then for every literal l and proof tag #, either $D \vdash +\#l$ or $D \vdash -\#l$.*

Notice that, in general, loops in the dependency graph do not always result in not being able to derive conclusions. For example, given the theory $a \Rightarrow b$, $b \Rightarrow a$, and $\Rightarrow \neg a$, while there is a loop in the dependency graph, the loop does not prevent proving conclusions. Indeed in this theory we can prove $+\partial \neg a$, $-\partial a$ and $-\partial b$. In case one wants to ensure that a theory is always decisive (no lack of conclusions), then one could adopt the well-founded variant of the variants proposed in this paper [29, 27]. The price to pay for this is that the complexity for deriving the extension is now quadratic instead of linear.

The last result we want to present here is about the expressive power of theories with or without the superiority relation (or theories where the superiority relation is empty).

PROPOSITION 6. *[3] Let $D = (F, R, >)$ be a defeasible theory, and let H_D be the Herbrand universe of D. Then a theory $D' = (F, R', \emptyset)$ such that $D \vdash \pm\partial l$ iff $D' \vdash \pm\partial l$ exists (for $l \in H_D$).*

The meaning of this proposition is that, for ambiguity blocking, we can remove the superiority relation from a theory without changing its expressive power. As a consequence we can rewrite the proof conditions as follows:

$+\partial$: If $P(n+1) = +\partial q$ then either
 (1) $q \in F$ or
 (2)(2.1) $\exists r \in R[q] \forall a \in A(r)\colon +\partial a \in P(1..n)$ and
  (2.2) $\sim q \notin F$ and
  (2.3) $\forall s \in R[\sim q]\ \exists a \in A(s)\colon -\partial a \in P(1..n)$.

$-\partial$: If $P(n+1) = -\partial q$ then
 (1) $q \notin F$ and
 (2) (2.1) $\forall r \in R[q]\ \exists a \in A(r)\colon -\partial a \in P(1..n)$ or
  (2.2) $\sim q \in F$ or
  (2.3) $\exists s \in R[\sim q]\ \forall a \in A(s)\colon +\partial a \in P(1..n)$.

Thus, for a literal l we can conclude $+\partial l$ if there is an applicable rule for it and there are no applicable rules for its complement.

Before concluding this section we give a quick overlook at the relationships between Defeasible Logic and argumentation. In [17] we give a Dung like argumentation semantics for the original ambiguity blocking variant (∂) of Defeasible Logic. In [18] we extended the results of [17] and we proved that the ambiguity propagation of Defeasible Logic (δ) is characterised by Dung's grounded semantics.

4. MAPPING CARNEADES TO DEFEASIBLE LOGIC

For mapping Carneades to Defeasible Logic we will use ideas from Modal Defeasible Logic [19]; in particular we introduce modal operators corresponding to provability operators (for the various proof standards). Thus the key concept is that if we derive p with a particular proof condition, for example, let us say we derive $+\partial p$, then we can assert $\Box^{ps} p$ for the proof standard ps corresponding to the proof condition, $+\partial$ in the example. Accordingly, the first thing to do is to extend the language to capture this.

DEFINITION 10. *If l is a literal, then $\Box^{ps} l$ and $\neg\Box^{ps} l$ are modal literals for $ps \in \{se, pe, ce, bd, dv\}$.*

The second thing to do is to adjust the definition of rule.

DEFINITION 11. *A rule is an expression $A \Rightarrow c$, where A is a (possibly empty) set of modal literals and c is a literal.*

The final step is to revise the notion of theory to accommodate the changes above and to extend the definition of superiority relation to capture the various proof standards.

DEFINITION 12. *A theory D is a structure $(F, R, >^{PS})$ where*

- *F the set of facts is a set of literals;*
- *R is a set of rules;*
- $>^{PS} = \{>^{ps}\colon ps \in \{pe, ce, bd, dv\}\}$.

The major difference with what presented in the previous section is the superiority relation: instead of a single superiority relation we have a set of superiority relations, one for each proof standard (except for the proof standard *scintilla of evidence*).

After the language we have to adjust the proof conditions. The adjustments are as follows: the proof conditions depend on whether rules are applicable or discarded. The general format for a rule to be applicable is:

 a rule r is applicable iff $\forall a \in A(r) : +\#a \in P(1..n)$.

A rule is applicable in a derivation if all the elements of the antecedent of the rule have been proved (with a particular strength). Similarly the general format for discarding a rule is:

 a rule r is discarded iff $\exists a \in A(r)$ such that $-\#a \in P(1..n)$.

A rule is discarded if at least one of the elements of the antecedent has been rejected (with a particular strength).

Now the elements of the antecedents are modal literals and we have to specify what it means for a modal literal to be derivable or rejected.

DEFINITION 13. *Let P be a derivation. A rule r is applicable at $P(1+i)$ iff $\forall ml \in A(r)$:*

- *if $ml = \Box^{se} l$ then $+\sigma^- l \in P(1..i)$;*
- *if $ml = \neg\Box^{se} l$ then $-\sigma^- l \in P(1..i)$;*
- *if $ml = \Box^{pe} l$ then $+\partial^*_{pe} l \in P(1..i)$;*
- *if $ml = \neg\Box^{pe} l$ then $-\partial^*_{pe} l \in P(1..i)$;*
- *if $ml = \Box^{ce} l$ then $+\partial^*_{ce} l \in P(1..i)$;*
- *if $ml = \neg\Box^{ce} l$ then $-\partial^*_{ce} l \in P(1..i)$;*
- *if $ml = \Box^{bd} l$ then $+\partial^*_{bd} l \in P(1..i)$;*
- *if $ml = \neg\Box^{bd} l$ then $-\partial^*_{bd} l \in P(1..i)$;*
- *if $ml = \Box^{dv} l$ then $+\partial^*_{dv} l \in P(1..i)$;*
- *if $ml = \neg\Box^{dv} l$ then $-\partial^*_{dv} l \in P(1..i)$.*

The definition for when a rule is discarded is similar, namely:

DEFINITION 14. *Let P be a derivation. A rule r is discarded at $P(1+i)$ iff $\exists ml \in A(r)$ such that:*

- *if $ml = \Box^{se} l$ then $-\sigma^- l \in P(1..i)$;*
- *if $ml = \neg\Box^{se} l$ then $+\sigma^- l \in P(1..i)$;*
- *if $ml = \Box^{pe} l$ then $-\partial^*_{pe} l \in P(1..i)$;*
- *if $ml = \neg\Box^{pe} l$ then $+\partial^*_{pe} l \in P(1..i)$;*
- *if $ml = \Box^{ce} l$ then $-\partial^*_{ce} l \in P(1..i)$;*
- *if $ml = \neg\Box^{ce} l$ then $+\partial^*_{ce} l \in P(1..i)$;*
- *if $ml = \Box^{bd} l$ then $-\partial^*_{bd} l \in P(1..i)$;*
- *if $ml = \neg\Box^{bd} l$ then $+\partial^*_{bd} l \in P(1..i)$;*
- *if $ml = \Box^{dv} l$ then $-\partial^*_{dv} l \in P(1..i)$;*
- *if $ml = \neg\Box^{dv} l$ then $+\partial^*_{dv} l \in P(1..i)$.*

To complete the modifications required for the proof conditions we have to take care of the instances of the superiority relations. The general rule is that in a derivation for a tag $\#_{ps}$ we have to use the corresponding superiority relation, i.e., $>^{ps}$. The proof conditions for Carneades Defeasible Logic are obtained from the proof conditions for the variants of Defeasible Logic presented in Section 3 by replacing the definitions of acceptable and discarded with the notions defined in Definitions 13 and 14, and replacing the instances of the superiority relation.

To illustrate the construction we provide the full proof conditions for $+\partial^*_{pe}$.

$+\partial^*_{pe}$: If $P(n+1) = +\partial q$ then either
 (1) $q \in F$ or
 (2)(2.1) there is a rule $r \in R[q]$ applicable at $P(n+1)$ and
  (2.2) $\sim q \notin F$ and
  (2.3) for all rules $s \in R[\sim q]$ s is discarded at $P(n+1)$ or $r >^{pe} s$.

Notice that the above construction has the effect of nullifying the distinction between ambiguity blocking (∂) and ambiguity propagation (δ).

We are now ready to give the mapping to translate a CAES into a defeasible theory. We begin with the transformation of an argument. Many notions of argument have been discussed it the argumentation literature where an argument can range from being an atomic element without any internal structure [11], to a set of proof trees [17] passing by a relationship between a set of assumptions/premises and a conclusion [37]. Carneades arguments seem to fall in this last category. But contrary to the construction in [37], the inferential mechanisms of Carneades suggest that Carneades arguments can be combined to generate proof trees. It has been advanced that a rule is just a relationship between a set of premises and a conclusion, and they can be combined to form proof trees. Accordingly, Carneades arguments are mapped to defeasible rules[4].

DEFINITION 15 (ARGUMENT MAPPING). *Given a CAES $S = \langle Arg, Ass, W, PS\rangle$, an argument mapping is a function marg that transforms an argument $a \in Arg$ into a defeasible rule such that given an argument $a = \langle P, E, c\rangle$*

$$marg(a) = \{\Box^{PS(p_i)} p_i : p_i \in P\} \cup \{\neg\Box^{PS(e_j)} e_j : e_j \in E\} \Rightarrow c$$

The second step is to take the weights associated with arguments and to generate the superiority relations associated with the proof conditions.

DEFINITION 16 (WEIGHT MAPPING). *Given a CAES $S = \langle Arg, Ass, W, PS\rangle$ and the threshold α, β and γ as in Definitions 7 and 8, a weight mapping is a function mwgt that takes as input a weight function and produces a set of superiority relations as follows:*

Let $a = \langle P_a, E_a, c_a\rangle$ and $b = \langle P_b, E_b, c_b\rangle$ be two arguments in Arg such that $c_a = \sim c_b$.

- $marg(a) >^{pe} marg(b)$ *iff* $W(a) > W(b)$;
- $marg(a) >^{ce} marg(b)$ *iff*
  1. $W(a) > \alpha$ *and*
  2. $W(a) - W(b) > \beta$;
- $marg(a) >^{bd} marg(b)$ *iff:*
  1. $W(a) > \alpha$ *and*
  2. $W(a) - W(b) > \beta$ *and*
  3. $W(b) < \gamma$;
- $>^{dv} = \emptyset$.

Finally, the full transformation of a Carneades Argument Evaluation Structure is given by the following definition.

DEFINITION 17. *Given a CAES $S = \langle Arg, Ass, W, PS\rangle$, the defeasible theory obtained from S is*

$$map(S) = (Ass, \{marg(a) : a \in Arg\}, mwgt(W))$$

We are now able to give the result about the relationship between Carneades and Defeasible Logic.

THEOREM 7. *Let S be a Carneades Argument Evaluation Structure, and $D = map(S)$, then*

1. *p is acceptable in S using proof standard* scintilla of evidence *iff $D \vdash +\sigma^- p$;*
2. *p is acceptable in S using proof standard $ps \in \{pe, ce, bd, dv\}$ iff $D \vdash +\partial^*_{ps} p$.*

PROOF. (sketch) The proof is by induction on the length of a derivation in Defeasible Logic and height of a proof tree in Carneades[5]. The inductive base is straightforward given that the base of acceptability for Carneades is whether a literal is an assumption or not, and for Defeasible Logic is being a fact or not. But facts in the defeasible theory corresponding to a CAES are the assumptions in the CAES.

For the inductive step, as usual we assume that the property holds for derivations of length up to n in one direction and for proof trees of height up to n in the other direction. The key point is the weight mapping. To illustrate the core of the proof we sketch the proof for the case of the *beyond reasonable doubt* proof standard.

Suppose that we have a Carneades argument a pro p such that the proof standard bd for p is satisfied. This means that the argument is applicable, thus by the inductive hypothesis the rule $marg(a)$ is applicable as well, satisfying clause (2.1) of $+\partial^*_{bd}$. This also means that the weight $W(a) > \alpha$, and then for any applicable argument b con p we have $W(a) - W(b) > \beta$ and $W(b) < \gamma$. Thus from *mwgt* we have that for every rule corresponding to $marg(b)$, we have $marg(a) >^{bd} marg(b)$, this implies that clause (2.3.2) of $+\partial^*_{bd}$ is satisfied for p. If there are other rules con p then these correspond to non applicable arguments in CAES and thus by inductive hypothesis those rules are discarded (satisfying clause (2.3.1) of $+\partial^*_{bd}$), and thus the two disjuncts of condition (2.3) of $+\partial^*_{bd}$ are satisifed, and thus we have $+\partial^*_{bd} p$. The argument for the proof in the other direction is similar. The proofs for the other proof standards are similar. □

A consequence of the above result, in conjunction with Proposition 4, is the answer to the question raised by Gordon and Walton [14] about the computational complexity of acceptability of a proposition given the current Carneades proof standards. Proposition 4 tells us that the complexity of computing the extension of Defeasible Logic has linear complexity. It is immediate to see that the mapping from a CAES to the corresponding defeasible theory is, in the worst case, quadratic (linear for the *dv* proof standard), given that we have to consider the relationships between weights of the arguments to derive the superiority relations.

COROLLARY 8. *Acceptability of a proposition in Carneades can be computed in polynomial time.*

The results in Theorem 7 shows that the inference mechanism of Carneades, based on the current proof standards, corresponds to a simple combination of defeasible logic theories (sharing the same rules and facts but with different superiority relations) where the conclusions for each theory are computed using the ambiguity blocking no team defeat variants of Defeasible Logic. In addition it shows that the proof standards *preponderance or evidence*, *clear and convincing evidence*, *beyond reasonable doubt* are all just instances of one and the same inference mechanism (this can also be easily seen from Carneades alone by setting the threshold α and β to 0 and γ to ∞). What about *dialectical validity*? Based on Proposition 6 it seems it corresponds to ∂, so it corresponds to the ambiguity blocking variant of defeasible logic. In addition since the superiority relation is not used for this proof standard, the distinction between team defeat and no-team defeat is irrelevant. However, we have shown in Example 3 that the two variants (∂ and ∂^*) are in general different[6]. Nevertheless, the use of the modal operators in the antecedents of rules (or alternatively the reliance of applicability of arguments) allows us apply the discussion in Remark 1 to specify that the proof conditions for ∂_{ps} are a generalisation of that of ∂^*_{ps}, and thus the proof standard *dv* is not stronger than the other proof standards. Here, we believe, the issue is on knowledge representation and not on logical mechanisms. It is on how to encode arguments. However, we refrain from investigating this issue any further in this paper.

[4] The transformation given in Definition 15 seems to suffer from a minor drawback. In Carneades it is possible to change the proof standard associated to a proposition without changing the arguments, but the transformation apparently has to change all arguments where the literal occurs. However, this is not the case if the function $PS(l)$ is instantiated during a proof in Defeasible Logic.

[5] In addition we prove that p is not acceptable in S iff $D \vdash -\#p$, where # is the tag for accepting S.

[6] The mapping given in [3] to remove the superiority relation does produce the same result for ∂^*.

5. PROOF STANDARDS: CARNEADES AND DEFEASIBLE LOGIC

In the previous section we have seen how to capture in Defeasible Logic the proof standards currently provided by Carneades> They correspond to a single inference mechanism in Defeasible Logic. In Section 3 we defined a variety of proof conditions. Thus in this section we discuss some issues with the proof standards that could be perceived as drawbacks on the current proof standards, and we advance how Defeasible Logic and its proof conditions could address them.

In Section 3 we gave a few versions of the notion of support, namely $\sigma^-, \sigma, \sigma^*$. and we have seen that σ^- corresponds to the Carneades proof standard *scintilla of evidence*. While the definition of this standard seems to conform to its use in some Common Law jurisdictions[7], Defeasible Logic allows us to capture different nuances of the notion of support. So what about the proof standard corresponding to σ? We believe that σ could be used to model *substantial evidence*, i.e., 'relevant evidence as a reasonable mind might accept as adequate to support a conclusion' [9]. Consider a conclusion proved with proof tag $+\sigma$: this means that there is a chain of reasoning leading to the conclusion, and for every argument con, if the argument con is stronger than the argument pro, then we have to show that the premises of the argument con do not hold (are not acceptable). We believe that it would be unreasonable to support a conclusion defeated by an argument deemed valid. For example, consider the following scenario: a drunk person claims that he had a glimpse of the accused (and the accused was not know to him before) in a place different from the crime scene, but footage from several high definition security cameras located at the crime scene clearly shows pictures of the accused at the crime scene at the time of the crime. In addition it has been assessed that the footage from the cameras has not been tampered with. In this case there is a *scintilla of evidence* that the accused was not at the crime scene, but it would be unreasonable to claim that there is *substantial evidence* about it, while, we think it reasonable to say that there is *substantial evidence* that the accused was at the crime scene at the time of the crime. The scenario can be modelled by the following rules:

$$r_1 : drunk \Rightarrow \neg crimeScene \qquad r_2 : camera \Rightarrow crimeScene$$

where $r_1 < r_2$ and the antecedents of the rules are given as facts. Then we derive $+\sigma^- \neg crimeScene$ but $-\sigma crimeScene$ and $+\sigma crimeScene$.

As we have alluded to in Section 2 in the version of Carneades proposed in [13] the comparison of the strength of the arguments depended on a preference relation (a partial order on the set of arguments) while in successive versions [14, 10] it was replaced by a weight function. One of the reasons for this move was the need to capture additional proof standards, in particular *clear and convincing evidence* and *beyond reasonable doubt*. However, in the previous section we have seen that this is not the case, and that it is possible to capture these two proof standards (as defined in Carneades) just using different superiority relations defining the relative strength of arguments. In general, we believe, a qualitative approach to preferences is more intuitive than a quantitative one. In our view it is hard to explain what is the difference between two arguments with, let us say, weight of 0.65 the first and 0.57 the second, and why one should accept arguments whose weight is above, let us say, 0.58 instead of 0.56.

The second reason, claimed by the authors, is to aggregate multiple arguments. The authors rightly argue that considering the sum of arguments pro versus the sum of arguments con often leads to counter-intuitive results, in particular when the arguments in one set might not be all independent from each other. To obviate this problem Carneades adopts the strategy of considering the maximum of the arguments pro and the maximum of arguments con. However, the use of weights cannot take into account fully independent arguments.

EXAMPLE 4. Consider a theory with the following rules

$$r_1 : a_1 \Rightarrow b \qquad r_2 : a_2 \Rightarrow b$$
$$r_3 : a_3 \Rightarrow \neg b \qquad r_4 : a_4 \Rightarrow \neg b$$

where $r_1 > r_3$ and $r_2 > r_4$.[8]

The idea here is that we have two independent chains of reasoning. On one side we have r_1 and r_3 where r_1 overrides r_3; on the other side we have r_2 and r_4 where r_2 overrides r_4. Besides that r_1 and r_4 are independent, and so are r_2 and r_3, meaning that in absence of any other information we cannot take a decision about the conclusion. Thus if a_1 and a_3 are given then we conclude b, and so if a_2 and a_4 are given. Also this should be the case when all antecedents are given. However, if either a_1 and a_4, or a_2 and a_3 are given we cannot conclude b; similarly when we have three of the antecedents but not both a_1 and a_2.

Notice that this example essentially has the same structure of Example 2. Formally, given a_1, a_2, a_3, a_4 we derive $+\partial b$ and $+\delta b$ but $-\partial b$.

We would like to point out that the above example can be modelled in the original version of Carneades [13], but it cannot be represented in version of Carneades using the weight functions.

Consider a Carneades Argument Evaluation Structure with the following arguments

- a_1 pro b
- a_2 pro b
- a_3 con b
- a_4 con b

where $W(a_1) > W(a_3)$ and $W(a_2) > W(a_4)$. Furthermore, we assume that $W(a_1) - W(a_3) > \gamma$, $W(a_2) - W(a_4) > \gamma$ but $W(a_1) - W(a_4) \leq \gamma$ and $W(a_2) - W(a_3) \leq \gamma$, and that $W(a_1), W(a_2) > \delta$, where $\gamma, \delta > 0$; δ and γ are the threshold for accepting arguments pro and for the minimum difference in weight between the max for arguments pro and max for arguments con.

From the description, it is immediate to see that when a_1 and a_2 are both applicable (but the other two arguments are not) we accept b; similarly when a_2 and a_4 are the applicable arguments, but we cannot accept it when the applicable arguments are either a_1 and a_4, or a_2 and a_3.

Carneades, however, is not able to accept b when all four arguments are applicable. Suppose it does. Thus $\max\{W(a_1), W(a_2)\} > \delta$ and $\max\{W(a_1), W(a_2)\} - \max\{W(a_3), W(a_4)\} > \gamma$.

Suppose that $W(a_1) = \max\{W(a_1), W(a_2)\}$; this means that $W(a_4) = \max\{W(a_3), W(a_4)\}$ (since $W(a_1) - W(a_2) > \gamma$ and $W(a_1) - W(a_4) \leq \gamma$). Thus we have $W(a_1) \geq W(a_2)$, $W(a_2) - W(a_4) > \gamma$, therefore $W(a_1) - W(a_4) > \gamma$ and thus we get a contradiction.

Suppose that $W(a_2) = \max\{W(a_1), W(a_2)\}$; this means that $W(a_3) = \max\{W(a_3), W(a_4)\}$ (since $W(a_2) - W(a_4) > \gamma$ and $W(a_2) - W(a_3) \leq \gamma$). Thus we have $W(a_2) \geq W(a_1)$, $W(a_1) -$

[7][9] defines *scintilla of evidence* as 'the slightest bit of evidence tending to support a material issue in a lawsuit'.

[8]For a less abstract instance of the theory in this example, consider these rules: r_1 : *monotreme* $\Rightarrow$ *mammal*, r_2 : *hasFur* $\Rightarrow$ *mammal*, r_3 : *laysEggs* $\Rightarrow \neg$*mammal*, r_4 : *hasBill* $\Rightarrow \neg$*mammal*. An then we have that a platypus is a mammal, has fur, lays eggs, and has a bill.

$W(a_3) > \gamma$, therefore $W(a_2) - W(a_3) > \gamma$ and thus we get a contradiction.

The impossibility to represent this scenario in the case of preponderance of evidence follows from the fact that we need to have $W(a_1) = W(a_4)$ and $W(a_2) = W(a_3)$, which then gives $W(a_1) = W(a_2)$ and $W(a_2) = W(a_3)$.

The second aspect we discuss here is about what we called combination of proof standards. Carneades associates to each stage and each proposition a proof standard to determine whether the proposition is acceptable at that stage. In addition the applicability of an argument depends on whether its premises are acceptable or not (where acceptable means acceptable at that stage). Thus, according to the formal definitions, it is possible to have an argument where some of the premises have to be accepted with one proof standard, let us say with the *scintilla of evidence* proof standard, while the conclusion has to be proved without reasonable doubt.

EXAMPLE 5. Consider a CAES with the following arguments

$$a_1 : \langle \emptyset, \emptyset, a \rangle \qquad a_2 : \langle \emptyset, \emptyset, \neg a \rangle \qquad a_3 : \langle a, \emptyset, b \rangle$$

where $W(a_1) < W(a_2)$, and the $PS(a) = se$ and $PS(b) = bd$. For the moment let us assume that the thresholds are satisfied. It is easy to see that the conclusion a is acceptable since, trivially, there is an applicable argument. Then, the argument a_3 is applicable, the appropriate thresholds are reached and there are no arguments against it.[9]

The above example poses a few questions. Is it appropriate to state that the proof standard for b has been met? From one point of view, the proof standard has been attained, since all we needed was to establish that there were a glimpse of evidence for a, the premise of the argument for b. From another perspective, the perspective where an argument is not reduced to a rule as it is done in Carneades, but an argument is a chain of reasoning, or better a proof tree, the argument (chain) a, b is attacked and defeated, so while there is no doubt about the connection between a and b, the validity of the premise is seriously questioned, and there is no substantial evidence to support b. While the example is specifically about the combination of the weakest and the strongest proof standards (to accentuate the possibly counterintuitive results) the aim of the discussion is whether combining proof standards is meaningful. not.

The formal machinery of Carneades allows for the combination of proof standards. Mathematically, this is the most general option. However, in [14] Carneades' method is explicitly motivated by the view that the *scintilla of evidence* proof standard pertains to a different proof burden than the other proof standards and is therefore relevant at a different stage in a legal proceeding [33]. More precisely, [33] argues that the *scintilla of evidence* proof standard is relevant for the burden of production while the other proof standards are relevant for the burden of persuasion. In [14] the authors propose to differentiate several stages of an argumentation dialogue, and at the last stage of the argumentation phase (the stage where burden of production is decided) all propositions have associated to them the *scintilla of evidence* proof standard, while at the last stage of the closing dialogue (when the the burden of persuasion is decided) each proposition is mapped to its applicable proof standard for this type of dialogue.

The second question is again about the use of thresholds. Suppose that the weight of argument a_3 does exceed α. The proof standard *dialectical validity* is satisfied, but the standard *beyond reasonable doubt* is not. It is possible to argue that the above interpretation is not appropriate: no matter how feeble the argument for b is there are no applicable arguments against it, so how can one doubt it if no reasons for the opposite have been put forward?

[9] Technically if the set of arguments con is empty the maximum is not defined, thus we assume that in such a case the maximum is 0, and we further assume that $\alpha > \beta$.

What about the mapping to Defeasible Logic? The mapping given in Definition 17 follows strictly the approach preposed by Carneades. The CAES of Example 5 is mapped to the following theory:

$$r_1 : \Rightarrow a \qquad r_2 : \Rightarrow \neg a \qquad r_3 : \Box^{se} a \Rightarrow b$$

From r_1 we can derive $+\sigma^- a$, thus we can assert $\Box^{se} a$, which is what we need to establish that r_2 is applicable.

On the other hand, all variants of Defeasible Logic presented in Section 3 require all elements of the antecedent of a rule to be provable with at least the same strength of the strength of the conclusion we want to prove.

The final aspect we want to discuss regards the *beyond reasonable doubt* proof standard. This issue we are going to discuss is related to the distinction between ambiguity blocking and ambiguity propagation.

REMARK 2. Defeasible Logic is neutral about ambiguity blocking and ambiguity propagation. Indeed variants to capture both intuitions have been defined. [13] claims that in [18] we argue that ambiguity propagation is counterintuitive. The argument in [18] is that, similarly to what we are going to argue here, it is possible to justify both views on ambiguity, and that both ambiguity propagation and ambiguity blocking have their own sphere of applicability, and there are applications where ambiguity blocking is counterintuitive as well as applications where ambiguity propagation is counterintuitive, and finally applications where we need both. Thus the outcome of our discussion is that a (sceptical) non-monotonic formalism should be able to accommodate both. To the best of our knowledge, currently, Defeasible Logic is the only formalism able to do that.

Let us illustrate again the distinction between ambiguity blocking and ambiguity propagation with the help of the following example.

EXAMPLE 6. Let us suppose that a piece of evidence A suggests that the defendant in a legal case is not responsible while a second piece of evidence B indicates that he/she is responsible; moreover the sources are equally reliable. According to the underlying legal system a defendant is presumed innocent (i.e., not guilty) unless responsibility has been proved (without reasonable doubt).

The above scenario is encoded in the following defeasible theory:

$$r_1 : \mathit{evidenceA} \Rightarrow \neg \mathit{responsible},$$
$$r_2 : \mathit{evidenceB} \Rightarrow \mathit{responsible}, \qquad r_3 : \mathit{responsible} \Rightarrow \mathit{guilty},$$
$$r_4 : \Rightarrow \neg \mathit{guilty}.$$

Given both *evidenceA* and *evidenceB*, the literal *responsible* is ambiguous. There are two applicable rules (r_1 and r_2) with the same strength, each supporting the negation of the other. As a consequence r_3 is not applicable, and so there is no applicable rule supporting the *guilty* verdict. Thus according to ambiguity blocking we obtain $+\partial \neg \mathit{guilty}$. In contrast, in an ambiguity propagation setting we propagate the ambiguity of *responsible* thus the literals *guilty* and $\neg$*guilty* are ambiguous; hence an undisputed conclusion cannot be drawn, so we have both $-\delta \mathit{guilty}$ and $-\delta \neg \mathit{guilty}$.

When we look at the example above is it appropriate to say that we have reached a not guilty verdict without any reasonable doubt? The evidence supporting that the defendant was responsible has not been refuted.

EXAMPLE 7. Let us extend the previous example. Suppose that the legal system allows for compensation for wrongly accused people. A person (defendant) has been wrongly accused if the defendant is found innocent, where innocent is defined as $\neg guilty$. In addition, by default, people are not entitled to compensation. The additional elements of this scenario are modelled by the rules:

$$r_5 : \neg guilty \Rightarrow innocent$$
$$r_6 : innocent \Rightarrow compensation$$
$$r_7 : \Rightarrow \neg compensation$$

where $r_6 > r_7$.

Continuing the discussion from Example 6, if we adopt ambiguity blocking propagation, then we have that despite there is some doubt about responsibility and, consequently, we cannot rule out that the defendant was wrongly accused, the conclusion is that the defendant is entitled to be compensated for having been wrongly accused. ($+\partial compensation$). Ambiguity propagation does not allow us to draw the same conclusion; in fact we have $-\delta compensation$.

In the last two examples there was no need to use the superiority relation. This poses some doubts about the definition of the *beyond reasonable doubt* proof standard given in Carneades. Indeed to obtain a positive conclusion using ambiguity propagation we have to use the superiority relation to resolve conflicts/ambiguities, but we need is that the audience establishes that one of the two sides of the conflict overrides the other, but there is no need that they quantify how much it overrides it.

To conclude this section we would like to propose alternative definitions of the proof standards proposed in Carneades, based on the discussion we had so far.

DEFINITION 18. *Given a defeasible theory $D = (F, R, >)$*

- *p is proved with proof standard* scintilla of evidence *iff $D \vdash +\sigma^- p$;*
- *p is proved with proof standard* substantial evidence *iff $D \vdash +\sigma p$;*
- *p is proved with proof standard* preponderance of evidence *iff $D \vdash +\partial p$;*
- *p is proved with proof standard* beyond reasonable doubt *iff $D \vdash +\delta p$.*
- *p is proved with proof standard* dialectical validity *iff $D' \vdash +\delta p$, where $D' = (F, R, \emptyset)$.*[10]

The above classification leaves out *clear and convincing evidence*. It is unclear to us how to define proof conditions whose strength is in between that of *preponderance of evidence* and *beyond reasonable doubt*. A possible solution would be to use a technique similar to what we did to reconstruct Carneades, and to use a theory $D = (F, R, >^1, >^2)$ with two superiority relations, such that $>^2 \subseteq >^1$, and then *preponderance of evidence* corresponds to ∂ using $>^1$, *clear and convincing evidence* to ∂ using $>^2$, and *beyond reasonable doubt* to δ using $>^2$.

6. CONCLUSION

In the previous sections we have provided a reconstruction of the current Carneades proof standards in terms of Defeasible Logic, and we have seen that, with the exception of *scintilla of evidence* all these standards correspond to the ambiguity blocking no team defeat inference mechanism of Defeasible Logic. The result is twofold: (i) this means that it is possible to use an implementation of Defeasible Logic as the engine for computing acceptability in Carneades (ii) it was possible to use the theoretical results studied for Defeasible Logic to shed light on some Carneades features and to highlight some possible shortcomings of decision choices behind the design of the current proof standards, and to propose alternative proof standards to be used in conjunction with the other aspects of Carneades.

Carneades is also the name of a tool implementing (part of) the Carneades framework. Carneades, the tool, has the ability to invoke Semantic Web (OWL) reasoning services to handle monotonic aspects of (legal) reasoning. Defeasible Logic is equipped with strict rules to integrate some form of monotonic reasoning. In addition [38] shows how to combine Defeasible logic with an external Description Logic reasoning service. In addition. [15] proposes a methodology to create a Defeasible Description Logic; the resulting systems is able to use (some) construction typical of Description Logic adding defeasibility to them.

Acceptability of arguments is only one of the aspects of (legal) argumentation. Two other important aspects are about the burden of proofs associated with the claims in a (legal) proceedings and the dialectical nature of argumentation. In [24] we have shown how to adapt Defeasible Logic to cope with dynamic burden of proofs (dynamic, in the sense that the burden of proof can be determined during the proceedings depending on the facts of the case at hand), specifically with burden of production and burden of persuasion. For the dialectical nature of argumentation in [36] we proposed a reconstruction of ALIS [32] in terms of Defeasible Logic, and in [12] we used Defeasible Logic to model dialogue games where the parties involved have to follow asymmetric protocols (the protocols for the parties involved can be different, in particular the parties have to use different proof standards to support their claims).

Acknowledgements

I would like to thank Thomas Gordon, Henry Prakken, Antonino Rotolo and Giovanni Sartor for discussions on the relationships between Carneades and Defeasible Logic. Thanks are also due to the anonymous ICAIL 2011 referees for their valuable comments.

NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

7. REFERENCES

[1] G. Antoniou and A. Bikakis. DR-Prolog: A system for defeasible reasoning with rules and ontologies on the semantic web. *IEEE Transaction on Knowledge and Data Engineering*, 19(2):233–245, 2007.

[2] G. Antoniou, D. Billington, G. Governatori, and M. J. Maher. A flexible framework for defeasible logics. In *Proc. American National Conference on Artificial Intelligence (AAAI-2000)*, pages 401–405. AAAI/MIT Press, 2000.

[3] G. Antoniou, D. Billington, G. Governatori, and M. J. Maher. Representation results for defeasible logic. *ACM Transactions on Computational Logic*, 2(2):255–287, 2001.

[4] G. Antoniou, D. Billington, G. Governatori, and M. J. Maher. Embedding defeasible logic into logic programming. *Theory and Practice of Logic Programming*, 6(6):703–735, 2006.

[5] G. Antoniou, D. Billington, G. Governatori, M. J. Maher, and A. Rock. A family of defeasible reasoning logics and its implementation. In W. Horn, editor, *ECAI 2000. Proceedings of the 14th European Conference on Artificial Intelligence*, pages 459–463. IOS Press, 2000.

[10] Notice that removing the superiority relation without adding extra rules is different from the technique presented in [3] to show that the superiority relation does not add to the expressive power of the ambiguity blocking variant of Defeasible Logic. However, we have some concerns whether the resulting variant is meaningful or nor.

[6] G. Antoniou, M. J. Maher, and D. Billington. Defeasible logic versus logic programming without negation as failure. *Journal of Logic Programming*, 42(1):47–57, 2000.

[7] N. Bassiliades, G. Antoniou, and I. Vlahavas. A defeasible logic reasoner for the semantic web. *International Journal of Semantic Web and Information Systems*, 2(1):1–41, 2006.

[8] D. Billington, G. Antoniou, G. Governatori, and M. J. Maher. An inclusion theorem for defeasible logic. *ACM Transactions in Computational Logic*, 12(1):article 6.

[9] A. B. Brann, editor. *The Law Dictionary*. Anderson Publishing, 7th edition, 2002.

[10] G. Brewka and T. F. Gordon. Carneades and abstract dialectical frameworks: A reconstruction. In P. Baroni, F. Cerutti, M. Giacomin, and G. R. Simari, editors, *Computational Models of Argument. Proceedings of COMMA 2010*, pages 3–12. IOS Press, 2010.

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

[12] J. Eriksson Lundström, G. Governatori, S. Thakur, and V. Padmanabhan. An asymmetric protocol for argumentation games in defeasible logic. In A. Ghose and G. Governatori, editors, *10 Pacific Rim International Workshop on Multi-Agents*, LNAI 5044, pages 219–231. Springer, 2008.

[13] T. F. Gordon, H. Prakken, and D. Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171(10-11):875–896, 2007.

[14] T. F. Gordon and D. Walton. Proof burdens and standards. In I. Rahwan and G. Simari, editors, *Argumentation in Artificial Intelligence*, pages 239–260. Springer, 2009.

[15] G. Governatori. Defeasible description logic. In G. Antoniou and H. Boley, editors, *Rules and Rule Markup Languages for the Semantic Web: Third International Workshop, RuleML 2004*, LNCS 3323, pages 98–112. Springer, 2004.

[16] G. Governatori. Representing business contracts in RuleML. *International Journal of Cooperative Information Systems*, 14(2-3):181–216, 2005.

[17] G. Governatori and M. J. Maher. An argumentation-theoretic characterization of defeasible logic. In W. Horn, editor, *ECAI 2000. Proceedings of the 14th European Conference on Artificial Intelligence*, pages 469–474. IOS Press, 2000.

[18] G. Governatori, M. J. Maher, D. Billington, and G. Antoniou. Argumentation semantics for defeasible logics. *Journal of Logic and Computation*, 14(5):675–702, 2004.

[19] G. Governatori and A. Rotolo. BIO logical agents: Norms, beliefs, intentions in defeasible logic. *Journal of Autonomous Agents and Multi Agent Systems*, 17(1):36–69, 2008.

[20] G. Governatori and A. Rotolo. Changing legal systems: legal abrogations and annulments in defeasible logic. *Logic Journal of IGPL*, 18(1):157–194, 2010.

[21] G. Governatori, A. Rotolo, R. Riveret, M. Palmirani, and G. Sartor. Variants of temporal defeasible logic for modelling norm modifications. In R. Winkels, editor, *Proceedings of 11th International Conference on Artificial Intelligence and Law*, pages 155–159. ACM Press, 2007.

[22] G. Governatori, A. Rotolo, and G. Sartor. Temporalised normative positions in defeasible logic. In A. Gardner, editor, *10th International Conference on Artificial Intelligence and Law (ICAIL05)*, pages 25–34. ACM Press, 2005.

[23] G. Governatori and S. Sadiq. The journey to business process compliance. In J. Cardoso and W. van der Aalst, editors, *Handbook of Research on BPM*, chapter 20, pages 426–454. IGI Global, 2009.

[24] G. Governatori and G. Sartor. Burdens of proof in monological argumentation. In R. Winkels, editor, *Legal Knowledge and Information Systems JURIX 2010: The Twenty-Third Annual Conference*, Frontiers in Artificial Intelligence and Applications. IOS Press, 2010.

[25] J. F. Horty. Skepticism and floating conclusions. *Artificial Intelligence*, 135(1-2):55–72, 2002.

[26] H.-P. Lam and G. Governatori. The making of SPINdle. In G. Governatori, J. Hall, and A. Paschke, editors, *Rule Representation, Interchange and Reasoning on the Web*, LNCS 5858, pages 315–322. Springer, 2009.

[27] H.-P. Lam and G. Governatori. On the problem of computing ambiguity propagation and well-founded semantics in defeasible logic. In M. Dean, J. Hall, A. Rotolo, and S. Tabet, editors, *RuleML 2010: 4th International Web Rule Symposium*, LNCS 6403, pages 119–127. Springer, 2010.

[28] M. J. Maher. Propositional defeasible logic has linear complexity. *Theory and Practice of Logic Programming*, 1(6):691–711, 2001.

[29] M. J. Maher and G. Governatori. A semantic decomposition of defeasible logic. In *Proc. American National Conference on Artificial Intelligence (AAAI-99)*, pages 299–305. AAAI Press, 1999.

[30] M. J. Maher, A. Rock, G. Antoniou, D. Billington, and T. Miller. Efficient defeasible reasoning systems. *International Journal of Artificial Intelligence Tools*, 10:483–501, 2001.

[31] D. Nute. Defeasible logic. In *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 3, pages 353–395. Oxford University Press, 1994.

[32] H. Prakken and G. Sartor. Formalising arguments about the burden of persuation. In R. Winkels, editor, *Proceedings of 11th International Conference on Artificial Intelligence and Law*, pages 97–106. ACM Press, 2007.

[33] H. Prakken and G. Sartor. A logical analysis of burdens of proof. In H. Kaptein, H. Prakken, and B. Veheij, editors, *Legal Evidence and Proof: Statistics, Stories, Logic*, pages 223–253. Ashgate Publishing, 2009.

[34] S. Sadiq, G. Governatori, and K. Naimiri. Modelling of control objectives for business process compliance. In G. Alonso, P. Dadam, and M. Rosemann, editors, *BPM 2007*, LNCS 4714, pages 149–164. Springer, 2007.

[35] L. A. Stein. Resolving ambiguity in nonmonotonic inheritance hierarchies. *Artificial Intelligence*, 55(2):259–310, 1992.

[36] S. Thakur, G. Governatori, V. Padmanabhan, and J. Eriksson Lundström. Dialogue games in defeasible logic. In M. A. Orgun and J. Thornton, editors, *20th Australian Joint Conference on Artificial Intelligence, AI 2007*, LNAI 4830, pages 497–506. Springer, 2007.

[37] F. Toni. Assumption-based argumentation for epistemic and practical reasoning. In P. Casanovas, G. Sartor, N. Casellas, and R. Rubino, editors, *Computable Models of the Law, Languages, Dialogues, Games, Ontologies*, LNCS 4884, pages 185–202. Springer, 2008.

[38] K. Wang, D. Billington, J. Blee, and G. Antoniou. Combining description logic and defeasible logic for the semantic web. In G. Antoniou and H. Boley, editors, *Rules and Rule Markup Languages for the Semantic Web: Third International Workshop, RuleML 2004*, LNCS 3323, pages 170–181. Springer, 2004.