On modelling burdens and standards of proof in structured argumentation

Henry PRAKKEN [a], Giovanni SARTOR [b]
[a] *Department of Information and Computing Sciences, Utrecht University and Faculty of Law, University of Groningen, The Netherlands*
[b] *CIRSFID, University of Bologna and European University Institute, Law Department, Florence, Italy*

Abstract. A formal model is proposed of argumentation with burdens and standards of proof, overcoming shortcomings of earlier work. The model is based on a distinction between default and inverted burdens of proof. This distinction is formalised by adapting the definition of defeat of the *ASPIC*$^+$ framework for structured argumentation. Since *ASPIC*$^+$ generates abstract argumentation frameworks, the model is thus given a Dungean semantics. It is shown to adequately capture shifting proof burdens as well as Carneades' definitions of proof standards.

Keywords. Argumentation, burden of proof, proof standards.

1. Introduction

In AI & law research several formal models of burdens and standards of proof have been proposed [4, 9, 5, 1, 13, 6, 12, 7]. In this paper we focus on the burden of persuasion, which is the burden to prove a fact to a specified degree, on the penalty of failing to establish that fact in the proceedings. Proof standards for the burden of persuasion differ depending on the type of case. For example, in Anglo-American jurisdictions, in civil cases the proof standard usually is 'preponderance of evidence' while in criminal cases it is 'beyond reasonable doubt'. The burden of persuasion is verified at the end of a proceeding, after all evidence has been provided, so its verification is purely a matter of inference, since no further evidence can be adduced. Accordingly, in [12] we logically characterised the burden of persuasion as the task to make sure that in the final stage of the proceedings there exists a justified argument for the claim. We claimed that to model this, a Dung-style argumentation logic can be used, that is, a logic that generates a set of arguments with a binary relation of defeat [3], such as our logic of [11]. We also proposed that proof standards for the burden of persuasion can be incorporated in the defeat relation between arguments for conflicting conclusions. However, we did not formalise the latter proposal. A first aim of this paper is to provide such a formalisation. A second aim of this paper is to reconsider [9]'s claims that shifts in the burden of persuasion cannot be modelled in a Dungean semantics. To model such shifts, [11]'s argument game for Dung's so-called grounded semantics was in [9] adapted to allow for role switches between a plaintiff and a defendant, depending on who has the burden of

persuasion for a claim. To date no Dungean semantics for the adapted game has been found, which is a theoretical drawback.

An alternative model of reasoning with burdens and standards of proof is the Carneades argumentation framework [5, 6]. In Carneades, arguments for and against statements can be constructed and to each statement a separate standard of proof can be assigned. A statement is acceptable if it satisfies its standard of proof. Until recently, it was an open question whether Carneades could be given a Dungean semantics but Bas van Gijzel [14] translated Carneades into Dung's frameworks via the *ASPIC*$^+$ framework of [10]. In light of these results, it would seem at first sight that simply adopting Carneades solves the semantic problem noted in [9] and in addition provides a suitable formalisation of proof standards. However, we will argue that Carneades' treatment of proof burdens and proof standards is not fully satisfactory. Accordingly, we will present a way to model shifts of the burden of persuasion within *ASPIC*$^+$ by changing its defeat relation, thus retaining Dungean semantics. A crucial element in our analysis is a distinction between *default* and *inverted* burdens of persuasion.

2. Background

In this section we review Dung's [3] abstract argumentation frameworks and the *ASPIC*$^+$ framework of [10]. An *abstract argumentation framework* (AF) is a pair $\langle Args, Def \rangle$, where $Args$ is a set of *arguments* and $Def \subseteq Args \times Args$ is a binary relation of *defeat*. A semantics for AFs returns sets of arguments called *extensions*, which are internally coherent and defend themselves against attack. A key notion is that of an admissible set of arguments: $S \subseteq Args$ is *admissible* if it is conflict-free (no argument in S defeats an argument in S) and for all $A \in S$: if $B \in Args$ defeats A, then some $C \in S$ defeats B. Each extension in maximal in some sense; different semantics define different notions of maximality. Since their differences do not matter for our purposes, we will not present Dung's semantics here. Relative to a semantics, an argument is *justified* on the basis of an AF if it is in all extensions of the AF returned by the semantics, it is *overruled* if it is defeated by a justified argument, and it is *defensible* if it is neither justified nor overruled.

The *ASPIC*$^+$ framework [10] gives structure to Dung's arguments and defeat relation, integrating and generalising work of [8, 15, 11] and others. It assumes an unspecified logical language $\mathcal{L}$ with a contrariness relation $-$ and defines arguments as inference trees formed by applying strict or defeasible inference rules of the form $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$ and $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$, interpreted as 'if the *antecedents* $\varphi_1, \ldots, \varphi_n$ hold, then *without exception*, respectively *presumably*, the *consequent* φ holds'. In order to focus on the essence, we present a simplified version, with a symmetric instead of arbitrary contrary relation and with just two types of premises instead of four.

In *ASPIC*$^+$ argumentation systems are applied to knowledge bases to generate arguments and counterarguments. Combining these with an argument ordering results in argumentation theories, which generate Dung-style AFs.

Definition 2.1. [Argumentation system] An *argumentation system* is a tuple $AS = (\mathcal{L}, -, \mathcal{R}, \leq)$ where

- $\mathcal{L}$ is a logical language closed under classical negation.
- $-$ is a symmetric contrariness relation on $\mathcal{L}$ (p and $-p$ are said to be each other's *contradictories*).

- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules such that $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$.
- $\leq$ is a partial preorder on $\mathcal{R}_d$.

Definition 2.2. [Knowledge bases] A *knowledge base* in an argumentation system $(\mathcal{L}, ^-, \mathcal{R}, \leq)$ is a pair $(\mathcal{K}, \leq')$ where $\mathcal{K} \subseteq \mathcal{L}$ and $\leq'$ is a partial preorder on K_p. $\mathcal{K}$ is partitioned into two subsets $\mathcal{K}_p$ (the ordinary premises) and $\mathcal{K}_a$ (the assumptions).

Arguments can be constructed step-by-step from knowledge bases by chaining inference rules into trees. In what follows, for a given argument the function `Prem` returns all its premises, `Conc` returns its conclusion and `Sub` returns all its sub-arguments.

Definition 2.3. [Argument] An *argument* A on the basis of a knowledge base $(\mathcal{K}, \leq')$ in an argumentation system $(\mathcal{L}, ^-, \mathcal{R}, \leq)$ is:

1. φ if $\varphi \in \mathcal{K}$ with: $\texttt{Prem}(A) = \{\varphi\}$; $\texttt{Conc}(A) = \varphi$; $\texttt{Sub}(A) = \{\varphi\}$;
2. $A_1, \ldots A_n \;\; \rightarrow\!\!\!\!/\Rightarrow \;\; \psi$ if $A_1, \ldots, A_n$ are arguments such that there exists a strict/defeasible rule $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow\!\!\!\!/\Rightarrow \psi$ in $\mathcal{R}_s/\mathcal{R}_d$. $\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$, $\texttt{Conc}(A) = \psi$, $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$.

Definition 2.4. [Argumentation theories] An *argumentation theory* is a triple $AT = (AS, KB, \preceq)$ where AS is an argumentation system, KB is a knowledge base in AS and $\preceq$ is a partial preorder on the set of all arguments that can be constructed from KB in AS (below denoted by $\mathcal{A}_{AT}$).

Arguments can be attacked in three ways: attacking a conclusion of a defeasible inference, attacking the defeasible inference itself, or attacking an assumption-type premise. Note that unlike in full *ASPIC*$^+$ we do not allow attacks on ordinary premises: for simplicity we assume that if in an earlier stage of a dispute an ordinary premise is attacked, further arguments for this premise will have been provided.

Definition 2.5. [Attacks] Let A and B be two arguments.

- A *undercuts* B (on B') iff $\texttt{Conc}(A) = -B'$ for some $B' \in \texttt{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \psi$.[1]
- A *rebuts* B on (B') iff $\texttt{Conc}(A) = -\varphi$ for some $B' \in \texttt{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \varphi$.
- A *undermines* B (on φ) iff $\texttt{Conc}(A) = -\varphi$ for some $\varphi \in \texttt{Prem}(B) \cap \mathcal{K}_a$.

Attacks combined with an argument ordering yield three kinds of defeat. For undercutting and undermining attack no preferences are needed to make it succeed, since undercutters provide implicit exceptions to the defeasible rule they undercut, while the contradictories of assumptions are explicit exceptions.

Definition 2.6. [Successful rebuttal and defeat]

- A *successfully rebuts* B if A rebuts B on B' and $A \not\prec B'$.
- A *defeats* B iff A undercuts, undermines or successfully rebuts B.

[1] Here an unspecified method is assumed to name defeasible inferences in the object language.

ASPIC$^+$'s argumentation theories then generate Dung's abstract argumentation frameworks as follows:

Definition 2.7. [Argumentation framework] An *abstract argumentation framework (AF) corresponding to an* $AT = \langle AS, KB, \preceq \rangle$ is a pair $\langle Args, Def \rangle$ such that:

- $Args$ is the set $\mathcal{A}_{AT}$ as defined by Definition 2.3,
- Def is the relation on $Args$ given by Definition 2.6.

3. Conceptual analysis

In this section we give a conceptual analysis of the burden of persuasion, arguing for a distinction between *default* and *inverted* burdens of persuasion. This distinction will be the basis for our formal proposal in the following sections.

Any legal dispute has a main claim defended by one party and opposed by another party. Usually the burden of persuasion is on the main claim. We call this the default burden of persuasion. It is by default inherited by any proposition advanced in support of the main claim (either directly as a premise or conclusion of a subargument, or indirectly as an element of an argument that attacks a counterargument). However, this default inheritance is blocked if an explicit burden is assigned to an element of a counterargument of the opposing side. This we call an inverted burden of persuasion. Once a burden is inverted, then new burdens explicitly assigned to elements in the defending side's counter-counterarguments are also of the inverted type, and so on. As a matter of fact, in both civil- and common law jurisdictions the burden of persuasion is never inverted in criminal cases while it can be inverted in civil cases.

The difference between default and inverted burdens is only relevant when they are not met. If a default burden of persuasion for ϕ is not met, then ϕ is merely considered unprovable while if an inverted burden for ϕ is not met, the opposite, that is $-\phi$, is considered to hold, even if the arguments pro and con ϕ have equal weight. Thus, when the evidence is balanced, the two burdens lead to different outcomes: if there is a default burden on ϕ, a legal reasoner may well remain agnostic on the matter, while if there is an inverted burden on ϕ, he must accept $-\phi$ as legally justified. In other words, an inverted burden on ϕ generates a presumption of $-\phi$ in case the evidence is balanced.

Example 3.1. Consider a criminal case, with proof standard 'beyond reasonable doubt'.

r_1: *crime* $\Rightarrow$ *punishment*
r_2: *selfdefence* $\Rightarrow \neg r_1$

Since in criminal cases the burden of persuasion is never inverted, if arguments for *selfdefence* are provided, then the prosecution has to remove any reasonable doubt on this issue. Assume first that the prosecution attempts to do so with an argument for $\neg$ *selfdefence*. Then, on our analysis, if it outweighs any argument for *selfdefence* beyond reasonable doubt, then $\neg$ *selfdefence* is justified and *selfdefence* is overruled. If the evidence is balanced or the arguments for $\neg$ *selfdefence* are stronger but not beyond reasonable doubt, then both statements are defensible. Finally, if the evidence for *selfdefence* is stronger by any degree, then *selfdefence* is justified and $\neg$ *selfdefence* is overruled. Assume next that the prosecution does not provide an argument for $\neg$ *selfdefence* but

an argument attacking a premise, subargument or defeasible inference of the accused's argument for *selfdefence*. Then our analysis applies likewewise: if the prosecution's argument outweighs its target beyond reasonable doubt, then the argument for *selfdefence* is overruled and the argument for *pubishment* is justified.

Example 3.2. Consider a civil case with the preponderance of the evidence standard. Assume that in a medical malpractice case, a doctor is liable for compensation if the patient was injured because of the doctor's negligence:

r_1: *injury* $\wedge$ *negligence* $\Rightarrow$ *compensation*

Assume also that (as in Italian law) the burden of persuasion with respect to *negligence* is inverted, that is, the doctor has the inverted burden to prove $\neg$ *negligence*. If both arguments for and against *negligence* are provided, then our analysis implies the following. If the argument for $\neg$ *negligence* outweighs the argument for *negligence*, then $\neg$ *negligence* is justified. However, if the evidence is balanced, then since the burden is an inverted one, instead *negligence* is justified and $\neg$ *negligence* is overruled. A fortiori this also holds if an argument for *negligence* is stronger.

Assume next that the burden of persuasion concerning negligence is not inverted. Then the outcome should be different when the evidence on *negligence* is balanced: then all of *negligence*, $\neg$ *negligence* and *compensation* should be defensible.

4. Burdens and standards of proof in Carneades

In Carneades [5, 6] an argument has a set of premises P, a set of exceptions E and a conclusion c, which is either pro or con a statement. Unlike in *ASPIC*$^+$, all arguments are elementary, that is, they contain a single inference step; they are combined in recursive definitions of *applicability* of an argument and *acceptability* of its conclusion. In essence, an *argument* is *applicable* if (1) all its premises are given as a fact or are else an acceptable conclusion of another argument and (2), none of its exceptions is given as a fact or is an acceptable conclusion of another argument. A *statement* is *acceptable* if it satisfies its proof standard. Facts are stated by an *audience*, which also provides numerical *weights* of each argument plus *thresholds* for argument weights and differences in argument weights. Three of Carneades' proof standards are then defined as follows:

Definition 4.1. [Three proof standards in Carneades [6]] Statement p satisfies:

- *preponderance of the evidence* iff there exists at least one applicable argument pro p for which the weight is greater than the weight of any applicable argument con p.
- *clear-and-convincing evidence* iff there is an applicable argument A pro p for which:
  - p satisfies *preponderance of the evidence* because of A; and
  - the weight for A exceeds the threshold α, and
  - the difference between the weight of A and the maximum weight of the applicable con arguments exceeds the threshold β.
- *beyond-reasonable-doubt* iff p satisfies *clear-and-convincing evidence* and the maximum weight of the applicable con arguments is less than the threshold γ.

Let us consider Carneades in light of our above conceptual analysis. We first argue that its notions of burdens and standards of proof do not adequately model their legal counterparts. To start with, since Carneades models proof burdens and proof standards as orthogonal concepts, it does not model that in the law proof standards are relative to a proof burden in that first a proof burden is assigned to either ϕ or $\neg\phi$ and then the proof standard is only assigned to the statement that has the burden of proof. Among other things, this leads to the following problem. In [6] the burden of persuasion for a statement ϕ is said to be met if ϕ is acceptable in the final stage of a dispute. However in our Example 3.1 this implies that if *selfdefence* is acceptable in the final stage, then the accused has fulfilled his burden of persuasion for *selfdefence*, while the law instead assigns the burden of persuasion for $\neg$ *selfdefence* to the prosecution.

Next, in order to respect our conceptual analysis of Example 3.1, a Carneades user has to represent the example differently depending on the stage of a dispute (as also noted by [6]). To model that initially the accused has the burden of production for *selfdefence*, this statement has to be modelled as an exception. Consider the following arguments:

- A_1 with $P = \{crime\}$, $E = \{selfdefence\}$ and a conclusion pro *punishment*
- A_2 with $P =$ any, $E = \emptyset$ and a conclusion pro *selfdefence*
- A_3 with $P =$ any, $E = \emptyset$ and a conclusion con *selfdefence* ($=$ pro $\neg$ *selfdefence*)

Let us assume that with A_2 (and before A_3 has been moved) the accused has fulfilled its burden of production. Then A_1 has to be modified by making $\neg$ *selfdefence* an ordinary premise:

- A_1' with $P = \{crime, \neg selfdefence\}$, $E = \emptyset$ and a conclusion pro *punishment*

If not, then if neither for *selfdefence* nor for $\neg$ *selfdefence* the proof standard is satisfied (so there remains reasonable doubt on this issue), then A_1 is acceptable since its exception *selfdefence* is not acceptable. Clearly, this is counterintuitive, so A_1 has to be modified to A_1': then if there remains reasonable doubt on *selfdefence*, A_1' is not acceptable. However, then the problem is that attempts by the prosecution to meet its burden of persuasion for *punishment* by attacking a premise of A_2 without constructing an argument for $\neg$ *selfdefence* cannot be modelled. So [6] representation method does not fully respect our analysis of Example 3.1.

5. A new proposal to represent burdens and standards of persuasion in *ASPIC*$^+$

We now formalise our idea in [12] that proof standards for the burden of persuasion can be incorporated in the defeat relation. A stronger argument should strictly defeat a weaker rebuttal only if the degree to which it is stronger satisfies the applicable proof standard; otherwise both arguments should defeat each other. For example, if the standard is 'preponderance of evidence', then A already strictly defeats B if A is just a little bit stronger than B, while if the standard is 'beyond reasonable doubt', A strictly defeats B only if A is very much stronger than B. We combine this with a formal account of the difference between default and inverted burdens of persuasion. For simplicity we only consider the proof standards preponderance and beyond reasonable doubt.

For ease of comparison we retain Carneades' weights and thresholds. They are used here just as a language to express natural-language judgements about argument strength.

For example, in practice such judgements are often not just about whether one argument is stronger than another but about whether the difference in strength is sufficient to let one argument defeat another. Such a comparison can be expressed in our proposal as $w(A) > w(B) + \beta$, without having to specify or calculate with numbers. In the examples we use numbers for ease of illustration only, as shorthand for such comparative judgements; no further meaning should be read into the numbers.

Our method still produces a set of arguments plus a binary defeat relation, so it still generates an abstract argumentation framework according to Definition 2.7, which has Dungean semantics. We now assume various other input elements besides a knowledge base, namely, a given explicit allocation of inverted proof burdens to statements, plus a given assignment of a proof standard to each statement that has a proof burden. The new definition of successful rebuttal considers whether the burden of persuasion is on ϕ or on $-\phi$, and it replaces the reference to the argument ordering $\preceq$ with an expression in terms of the weights of the rebutting arguments. Since these expressions assume numerical argument weights and thresholds for these weights, we now also assume these as given.

Definition 5.1. [Bop-argumentation theories] A *bop argumentation theory* is a tuple $AT = (AS, KB, t, B, w, \alpha, \beta)$ where AS is an argumentation system KB a knowledge base in AS as before, and

- $t \in \mathcal{L}$ (the main topic of the AT);
- $B \subseteq \mathcal{L}$ such that for no ϕ, both ϕ and $-\phi$ are in B (we write $\texttt{ebop}(\phi)$ iff $\phi \in B$);
- $w : \mathcal{A}_{AT} \longrightarrow \mathbb{R}^+ \cup \{0\}$;
- $\alpha, \beta, \gamma \in \mathbb{R}^+ \cup \{0\}$.

For any $A \in \mathcal{A}_{AT}$ such that $\texttt{conc}(A)$ has the proof standard beyond reasonable doubt $\mathcal{R}_s$ is assumed to contain rules $\rightarrow \neg A$ if $w(A) < \alpha$, and rules of the form $B_1, \ldots, B_n \rightarrow \neg A$ for any $B = B_1, \ldots, B_n \rightarrow \neg \texttt{Conc}(A)$ such that $w(B) \geq \gamma$.[2]

Next we assume that a bop-argumentation theory determines a set I of (explicit or implicit) inverted proof burdens, with the constraint that $B \subseteq I$. We will discuss the definition of I below. We write $\texttt{ibop}(\phi)$ iff $\phi \in I$ and $\texttt{dbop}(\phi)$ iff ϕ has a default burden (see also below). Our new definition of successful rebuttal is then as follows:

Definition 5.2. [successful rebuttal under burden of persuasion]. Argument A *successfully rebuts* argument B if A rebuts B on B' and

1. $\texttt{ibop}(\texttt{Conc}(A))$ and $w(A) > w(B') + \beta$; or else
2. $\texttt{dbop}(\texttt{Conc}(A))$ and $w(A) \not< w(B')$; or else
3. $w(A) + \beta \not< w(B')$.

The idea behind the definition is twofold (i) Shifts in the burden of persuasion are captured since arguments that attempt to fulfill an inverted burden defeat their target only if they are stronger than their target. (ii) Proof standards are modelled by incorporating [6]'s β threshold in our definition of defeat. Note that with the preponderance standard (2) and (3) are equivalent, since in that case $\beta = 0$. Note also that if there are no other relevant arguments and assuming for beyond reasonable doubt that A exceeds the thresh-

[2]This idea is adapted from [14]. To avoid circularity, $w(A)$ is assumed to depend only on the content of A.

old α and B does not meet the threshold γ, then A for p strictly defeats B for $-p$ just in case A meets the proof standard for p according to Definition 4.1.

Let us illustrate the definition with our two running examples. For simplicity we assume that weights range from 0 to 1, and that beyond reasonable doubt is satisfied iff the weight of an argument exceeds 0.9. Then for this standard $\beta = 0.8$ (recall that for preponderance it is 0).

In our criminal case, we have $\texttt{dbop}(punishment)$ and $\texttt{dbop}(\neg selfdefence)$. Consider again the arguments A_1, A_2, A_3 but now in their obvious *ASPIC*$^+$ format. Assume first $w(A_2) = 0.05, w(A_3) = 0.95$. Then $w(A_2) + \beta = 0.85$, and $0.85 < 0.95$ so A_3 defeats A_2 while A_2 does not defeat A_3. So $\neg selfdefence$ and $punishment$ are justified, as we want. Assume next $w(A_2) = 0.3, w(A_3) = 0.7$. Then $w(A_2) + \beta = 1.1$, so while A_3 defeats A_2, now A_2 also defeats A_3. So all of $\neg selfdefence$, $\neg selfdefence$ and $punishment$ are justified, as we want. Assume finally $w(A_2) = 0.6, w(A_3) = 0.4$. Then $w(A_2) + \beta = 1.4$, so A_2 now defeats A_2 while A_3 does not defeat A_2. So $selfdefence$ is justified while $\neg selfdefence$ and $punishment$ are overruled, as we want.

In our civil case, we have $\texttt{ibop}(\neg negligence)$. Furthermore, our idea is that we have $\texttt{dbop}(compensation)$ since this is plaintiff's main claim, while we don't have $\texttt{dbop}(negligence)$ since the inverted burden is on the opposite. Consider the following arguments (leaving implicit what are the facts and rules):

A_1:	*injury*	B_1:	*goodPastRecord*
A_2:	*Appendicitis*	B_2:	$B_1 \Rightarrow \neg$ *negligence*
A_3:	$A_2 \Rightarrow \neg$ *riskyOperation*		
A_4:	$A_3 \Rightarrow$ *negligence*		
A_5:	$A_1, A_4 \Rightarrow$ *compensation*		

arguments A_4 and B_2 rebut each other on whether the doctor was negligent. Assume first $w(A_4) = 0.4, w(B_2) = 0.6$. Then B_2 defeats A_4 by clause (1) of definition 5.2, while A_4 does not defeat B_2 by clause (3). So $\neg negligence$ is justified while $negligence$ and $compensation$ are overruled, as we want. Assume next $w(A_4) = 0.6, w(B_2) = 0.4$. Then A_4 defeats B_2 by clause (3) while B_2 does not defeat A_4 by clause (1). So $negligence$ and $compensation$ are justified while $\neg negligence$ is overruled, as we want. Assume finally $w(A_4) = w(B_2) = 0.5$. Then A_4 defeats B_2 by clause (3) while B_2 does not defeat A_4 by clause (1). So again $negligence$ and $compensation$ are justified while $\neg negligence$ is overruled, as we want.

Assume next that the burden of persuasion with respect to negligence is not inverted. Then the plaintiff has the default burden of persuasion for both *compensation* and *negligence*. Then the outcome only differs in the third case, that is, when $w(A_4) = w(B_2) = 0.5$. Then A_4 defeats B_2 by clause (2) while B_2 defeats A_4 by clause (3). So all statements are defensible, as we want.

In sum, our formal proposal adequately models our conceptual analysis of Section 3, as illustrated by our two running examples.

One issue is left to be discussed, namely, how an explicit allocation of inverted burdens and proof standards determines burdens and standards for other statements. Let us add the following arguments to the ones above in our civil example (see Figure 1):

B_3:	*medicalTests1*	C_1:	*medicalTests2*
B_4:	$B_3 \Rightarrow$ *badCirculation*	C_2:	$C_1 \Rightarrow \neg$ *badCirculation*
B_5:	$B_4 \Rightarrow$ *riskyOperation*		

As we said above, we want that plaintiff's main claim *compensation* has a default burden

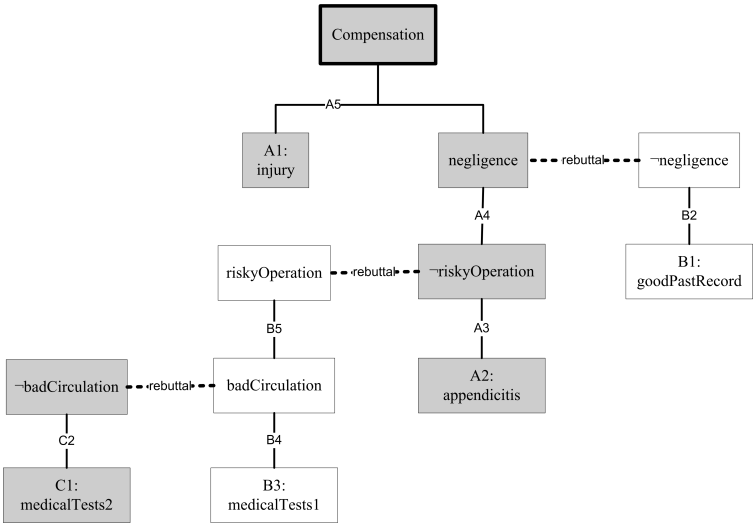

Figure 1. The civil case

of persuasion. Then, for instance, since A_1 is a subargument simply supporting plaintiff's main claim and no explicit inverted burden was assigned to *injury*, that statement should also have a default burden. On the other hand, *negligence* should not have a default burden even though A_4 also is a subargument supporting plaintiff's main claim. The reason is that A_4 is rebutted by B_2 and the conclusion of B_2 has an inverted burden: then *negligence* should have no burden. Let us next look at A_3 and B_5. The latter argument rebuts a subargument for *negligence* and therefore indirectly supports B_2 for ¬ *negligence*, which has an inverted burden. Therefore, *riskyOperation* should also have an inverted burden. This in turn implies that ¬*riskyOperation* should have no burden. Now we can also say more about, for instance, B_4 and C_2. Since B_4 supports an argument for an inverted burden, *badCirculation* should also have an inverted burden, so ¬*badCirculation* should have no burden. All these default inferences can be overridden by explicit inverted burden, but in our example these were not given.

In [9] similar ideas were formalised in terms of an argument game, but in this paper we are focussing on Dung-style semantics, so this route is not open to us. Since a formalisation is not trivial and our space is limited, we leave this issue for future research.

6. Conclusion

In this paper we have proposed a formalisation of reasoning with burdens and standards of proof, overcoming some shortcomings of earlier work of ourselves and the Carneades system of [5, 6]. Our main contribution is a Dungean semantics for argumentation with shifting burdens of persuasion, while we have also shown how Carneades' proof standards can be modelled within our semantics. A key ingredient in our model was a distinction between default and inverted burdens of persuasion.

In future research we wish to formalise our sketch in Section 5 of how explicit proof burdens determine implicit burdens. We also want to incorporate [7]'s idea to make proof burdens relative to legal rules. Finally, we aim to investigate whether our proposal respects the consistency and closure postulates of [2]. In [10] these postulates were proven for *ASPIC*$^+$ as summarised above in Section 2 but since we changed its definition of defeat, satisfaction of the postulates has to be investigated anew.

References

[1] K. Atkinson and T.J.M. Bench-Capon. Argumentation and standards of proof. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Law*, pages 107–116, New York, 2007. ACM Press.

[2] M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171:286–310, 2007.

[3] P.M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and *n*–person games. *Artificial Intelligence*, 77:321–357, 1995.

[4] K. Freeman and A.M. Farley. A model of argumentation and its application to legal reasoning. *Artificial Intelligence and Law*, 4:163–197, 1996.

[5] T.F. Gordon, H. Prakken, and D.N. Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171:875–896, 2007.

[6] T.F. Gordon and D.N. Walton. Proof burdens and standards. In I. Rahwan and G.R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 239–258. Springer, Berlin, 2009.

[7] G. Governatori and G. Sartor. Burdens of proof in monological argumentation. In R.G.F. Winkels, editor, *Legal Knowledge and Information Systems. JURIX 2010: The Twenty-Third Annual Conference*, pages 37–46. IOS Press, Amsterdam etc., 2010.

[8] J.L. Pollock. *Cognitive Carpentry. A Blueprint for How to Build a Person*. MIT Press, Cambridge, MA, 1995.

[9] H. Prakken. Modelling defeasibility in law: logic or procedure? *Fundamenta Informaticae*, 48:253–271, 2001.

[10] H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1:93–124, 2010.

[11] H. Prakken and G. Sartor. A dialectical model of assessing conflicting arguments in legal reasoning. *Artificial Intelligence and Law*, 4:331–368, 1996.

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

[13] K. Satoh, S. Tojo, and Y. Suzuki. Formalizing a switch of burden of proof by logic programming. In *Proceedings of the 1st International Workshop on Jurisinformatics (JURISIN 2007)*, pages 76–85, Miyazaki, Japan, 2007.

[14] B. van Gijzel and H. Prakken. Relating Carneades with abstract argumentation. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)*, pages 1113–1119, 2011.

[15] G.A.W. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90:225–279, 1997.