Burden of Persuasion in Argumentation

Roberta Calegari
CIRSFID – Alma AI,   *
University of Bologna, Italy
roberta.calegari@unibo.it

Giovanni Sartor
CIRSFID – Alma AI,
University of Bologna, Italy
European University Institute,
Florence, Italy
giovanni.sartor@unibo.it

This paper provides a formal model for the burden of persuasion in dialogues, and in particular, in legal proceedings. The model shows how an allocation of the burden of persuasion may induce single outcomes in dialectical contexts in which, without such an allocation, the status of conflicting arguments would remain undecided. Our approach is based on a two-stage labelling. The first-stage labelling determines what arguments are accepted, rejected or undecided, regardless of the allocation of burden. The second-stage labelling revises the dialectical status of first-stage undecided arguments, according to burdens of persuasion. The labelling is finally extended in such a way as to obtain a complete labelling. Our model combines two ideas that have emerged in the debate on the burden of persuasion: the idea that the burden of persuasion determines the solution of conflicts between arguments, and the idea that its satisfaction depends on the dialectical status of the arguments concerned. Our approach also addresses inversions of the burden of persuasion, namely, cases in which the burden of persuasion over an argument does not extend to its subarguments.

1 Introduction

The concept of a burden of proof plays a key role in argumentation and in law. It is a complex and controversial idea for which no agreed theory exists [12]. Generally, we can say that burdens of proof distribute dialectical responsibilities between the parties of argumentation: when a party has a burden of proof relative to a claim, that party will fail to establish the claim unless the party provides the kind of argument or evidence that is required to meet the burden. Burdens of proof can complement dialectical frameworks for argumentation systems. In particular, burdens of proof are important in adversarial contexts: they facilitate reaching a single outcome in contexts of doubt and lack of information.

In the legal domain, two burdens are distinguished: the burden of production (also called burden of providing evidence), and the burden of persuasion.

The burden of production concerns the necessity of providing evidence, in order to establish a claim. The evidence must be sufficient to establish the claim in the absence of counterarguments. When knowledge is represented through a set of rules and exceptions, the party who is interested in establishing the conclusion of a rule has the burden of production relative to the strong (by default non-assumable) elements in the antecedent condition of the rule. The other party – who is interested in preventing the derivation of the conclusion of the rule – has the burden of production relative to the exceptions to the rule. For instance, assume the law includes both the rule that there is murder if there are killing and intent, and the exception according to which there is no murder – or the murder rule does not apply – in case of self-defence. If evidence is not provided for both killing and intent, the prosecution will fail to substantiate the murder claim; on the other hand, if no evidence is provided for self-defence, the

defendant will fail to substantiate the claim that there is no murder, or that the murder rule is not to be applied.

The burden of persuasion concerns the necessity of providing a dialectically convincing argument, in order to establish a claim. In order to be convincing, the argument must prevail over all counterarguments – which are non-rejected on other grounds– to an extent that is determined by the applicable standard of proof. If this is not the case, then the argument burdened with persuasion must be rejected. Not only will that argument fail to justify its conclusion, but it will also be unable to successfully challenge any other arguments.

For instance, in criminal cases, the burden of persuasion for the non-existence of self-defence is on the prosecution, even if the burden of production is on the defendant. Assume that it has been established that the defendant has committed the criminal action intentionally, while it remains doubtful whether she acted in self-defence since there are arguments for and against self-defence. Under such circumstances, since prosecution has the burden of providing a convincing argument for the non-existence of self-defence, its doubtful argument against self-defence will be rejected. As consequence self-defence will be assumed, and the argument for murder will be rejected.

Our approach to the burden of persuasion expands the model proposed by Prakken and Sartor [9], according to which arguments for a claim burdened with persuasion will be strictly defeated unless they strictly prevail over arguments denying that claim. In our approach, to assess whether an argument for a claim ϕ meets the burden of persuasion, we not only consider whether the argument strictly prevails over arguments for $\bar{\phi}$. We also required that all (direct) subarguments are accepted. This idea corresponds to the model of burdens in argumentation proposed by Gordon et al. [4], who assess the satisfaction of burdens of proofs based on the dialectical status of the concerned argument.

Our model takes into account that a multistep argument can be attacked not only by contradicting its final conclusion but also by challenging its earlier steps. Thus, for an argument to meet a burden of persuasion, there must be no doubt – i.e., unsolved conflict – also relative to such steps. For instance, assume that the claim that the defendant did not act is self-defence is supported by testimony, and that it is doubtful whether the witness is reliable. In such a case too, the prosecution will not have met the burden on non-self-defence, and the argument against self defence will be rejected.

Our model originates from legal considerations and is applied to legal examples and is mostly relevant for the development of AI tools to support legal argumentation [2]. However, the issue of the burden of proof has a significance that goes beyond the law; it involves other domains as well – public discourse, risk management, etc. – in which multiple agents, having different tasks and responsibilities, provide evidence and arguments.

Formal models of argumentation are making significant and increasing contributions to AI—from defining the semantics of logic programs to implementing persuasive medical diagnostic systems, up to studying negotiation dialogues in multi-agent systems. The possibility to formalise burdens of proofs can be useful in multi-agent scenarios where agents have access to different pieces of information and may strategically withdraw information, and decisions are needed under situations of uncertainty.

Outline. The paper is organised as follows, Section 2 introduces a simple argumentation setting which is the basis of our model. Section 3 introduces two running examples. Section 4 presents the contribution of the paper, namely, the concept of burden of persuasion in the labelling stage. Final remarks are provided in Section 5.

2 Background notion: argumentation framework language

To illustrate our model of the burden of persuasion, we use a lightweight ASPIC+-like system for structured argumentation [7]. We do not include some key aspects of ASPIC+, since they are not needed to introduce our model and present our examples: we do not distinguish between premises, assumptions and rules, or between strict and defeasible rules, or between rebutting and undercutting attack, or between different criteria for priority between arguments. Our characterisation of burdens of persuasions can, however, be applied to the full ASPIC+, or to other similar argumentation systems.

In our model, arguments are produced from a set of defeasible rules, and attacks between arguments are captured by argumentation graphs, to which a labelling semantics is applied.

2.1 Defeasible theories and argumentation graphs

A literal is an atomic proposition or the negation of an atomic proposition.

Notation 2.1 *For any literal ϕ, its complement is written $\bar{\phi}$. That is, if ϕ is a proposition p then $\bar{\phi} = \neg p$, while if ϕ is $\neg p$ then $\bar{\phi}$ is p.*

Literals are put into relation through defeasible rules.

Definition 2.1 *A **defeasible rule** has the form:*

$$r: \quad \phi_1,...,\phi_n \Rightarrow \phi$$

with $0 \leq n$, and where

- *r is the unique identifier of the rule, and*
- *each $\phi_1,...\phi_n,\phi$ is a literal.*

A superiority relation $\succ$ is defined over rules: $s \succ r$ states that rule s prevails over rule r.

Definition 2.2 *A **superiority relation** $\succ$ over a set of rules Rules is an antireflexive and antisymmetric binary relation over Rules, i.e., $\succ \subseteq Rules \times Rules$.*

A defeasible theory consists in a set of rules and a superiority relation over the rules.

Definition 2.3 *A **defeasible theory** is a tuple $\langle Rules, \succ \rangle$ where Rules is a set of rules, and $\succ$ is a superiority relation over Rules.*

Given a defeasible theory and by chaining rules from the theory, we can construct arguments, as specified in the following definition, cf. [6, 3, 11].

Definition 2.4 *An **argument** A constructed from a defeasible theory $\langle Rules, \succ \rangle$ is a finite construct of the form:*

$$A: A_1,...A_n \Rightarrow_r \phi$$

with $0 \leq n$, where

- *A is the unique identifier of the argument;*
- *$A_1,...,A_n$ are arguments constructed from the defeasible theory $\langle Rules, \succ \rangle$;*
- *ϕ is the conclusion of the argument, denoted $Conc(A)$;*
- *$r: Conc(A_1),...,Conc(A_n) \Rightarrow \phi$ is the top rule of A, denoted $TopRule(A)$.*

Notation 2.2 *Given an argument* $A : A_1, \ldots A_n \Rightarrow_r \phi$ *as in definition 2.4,*

- *Sub*(A) *denotes the **set of subarguments** of* A, *i.e.,* $Sub(A) = Sub(A_1) \cup \ldots \cup Sub(A_n) \cup \{A\}$;

- *DirectSub*(A) *denotes the **direct subarguments** of* A, *i.e. DirectSub*(A) = $\{A_1, \ldots, A_n\}$.

We define the preference over arguments through a last-link ordering according to which an argument A is preferred over another argument B if the top rule of A is stronger than the top rule of B.

Definition 2.5 *A **preference relation** $\succ$ is a binary relation over a set of arguments $\mathscr{A}$: an argument* A *is preferred to argument B, denoted* $A \succ B$, *iff* $TopRule(A) \succ TopRule(B)$.

Argument A attacks argument B if A's conclusion is incompatible with the conclusion of a non-inferior subargument of B.

Definition 2.6 *An **attack relation** $\rightsquigarrow$ is a binary relation over a set of arguments $\mathscr{A}$: for any $A, B \in \mathscr{A}$,* A *attacks* B *iff* $\exists B' \in Sub(B)$ *such that* $Conc(A) = \overline{Conc(B')}$, *and* $B' \not\succ A$.

Given the arguments and attacks obtained from a defeasible theory, we can create argumentation graphs.

Definition 2.7 *An **argumentation graph** constructed from a defeasible theory T is a tuple $\langle \mathscr{A}, \rightsquigarrow \rangle$ where $\mathscr{A}$ is the set of all arguments constructed from T, and $\rightsquigarrow$ is an* attack *relation over $\mathscr{A}$.*

In the following, we assume that all argumentation graphs are constructed from a defeasible theory.

Notation 2.3 *Given an argumentation graph $G = \langle \mathscr{A}, \rightsquigarrow \rangle$, we may write $\mathscr{A}_G$, and $\rightsquigarrow_G$ to denote the graph's arguments, and attacks respectively.*

Example 1 (Argumentation graph) *Let us consider the following rules:*

$$
\begin{array}{ll}
r0: \quad \Rightarrow p & r1: \quad \Rightarrow q \\
r2: \quad p \Rightarrow \neg r & r3: \quad q \Rightarrow r \\
r4: \quad r \Rightarrow s & r5: \quad \Rightarrow \neg s
\end{array}
$$

with $r2 \succ r3$. *Accordingly to the above definitions, we can then build the following arguments:*

$$
\begin{array}{lll}
A1: \quad \Rightarrow p & B1: \quad \Rightarrow q & C1: \quad \Rightarrow \neg s \\
A2: \quad A1 \Rightarrow \neg r & B2: \quad 1 \Rightarrow r & \\
& B3: \quad B2 \Rightarrow s &
\end{array}
$$

Arguments A1 *and* B1 *are subarguments of* A2 *and* B2 *respectively.* B2 *is a subargument of* B3. A2 *attacks* B2 *and* B3, *while* C1 *and* B3 *attack each other. The superiority relation between r2 and r3 makes it possible to solve the conflict between arguments* B2 *and* A2 *in favour of* A2. *The corresponding argumentation graph is shown in Figure 1.*

□

2.2 Labelling semantics

We use $\{\text{IN}, \text{OUT}, \text{UND}\}$-labellings where any argument is associated with one label which is either IN, OUT, UND, respectively meaning that the argument is accepted, rejected, or undecided.

Definition 2.8 *A $\{\text{IN}, \text{OUT}, \text{UND}\}$-**labelling** of an argumentation graph G is a total function*

$$
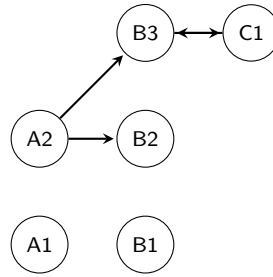\mathscr{A}_G \rightarrow \{\text{IN}, \text{OUT}, \text{UND}\}.
$$

Figure 1: Argumentation graph corresponding to the theory of Example 1.

Notation 2.4 *Given a labelling L, we write* IN(L) *for* $\{A \mid L(A) = $ IN$\}$, OUT(L) *for* $\{A \mid L(A) = $ OUT$\}$ *and* UND(L) *for* $\{A \mid L(A) = $ UND$\}$.

Definition 2.9 *An* $\{$IN, OUT, UND$\}$*-labelling of an argumentation graph G is a **complete** $\{$IN, OUT, UND$\}$-**labelling** iff for every argument* A *in* $\mathscr{A}_G$:

- A *is* IN *iff all attackers of* A *are* OUT,

- A *is* OUT *iff* A *has (at least) an attacker* IN.

Since $\{$IN, OUT, UND$\}$-labellings are total functions, if an argument is not IN or OUT, then it is UND.

An argumentation graph may have several complete $\{$IN, OUT, UND$\}$-labellings, we will focus on the unique complete labelling with the smallest set of labels IN, namely the grounded $\{$IN, OUT, UND$\}$-labelling.

Definition 2.10 *A **grounded** $\{$IN, OUT, UND$\}$-labelling L of an argumentation graph G is a complete* $\{$IN, OUT, UND$\}$*-labelling of G such that* IN(L) *is minimal.*

Example 2 (Grounded $\{$IN, OUT, UND$\}$- labelling) *The grounded $\{$IN, OUT, UND$\}$-labelling of the argumentation graph of Figure 1 is shown in Figure 2.*

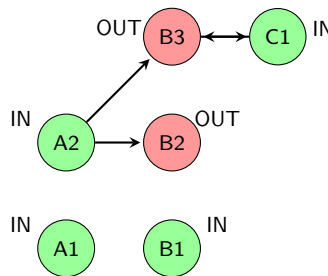
Figure 2: Grounded {IN,OUT,UND}-labelling of the argumentation graph of Figure 1

□

3   Running examples

In this section, we introduce two examples concerning legal cases, a civil law one and a criminal law one. We shall here consider grounded labelling in the absence of burdens of persuasion and then, in the next section, the labelling obtained according to the applicable burdens.

Example 3 (Civil law example: medical malpractice) *Assume the applicable law according to which doctors are liable for the harm suffered by a patient, if they were negligent in treating the patient. However, there is an exception: doctors can avoid liability if they show they were not negligent. Assume also that doctors are considered to be non-negligent – i.e., diligent – if they followed the medical guidelines that govern the case. Assume that in the case at hand it is uncertain whether the doctor has followed the guidelines. Accordingly, let us consider the following rules:*

r1:	$\Rightarrow \neg guidelines$	*r2:*	$\Rightarrow guidelines$
r3:	$\neg guidelines \Rightarrow negligent$	*r4:*	$guidelines \Rightarrow \neg negligent$
r5:	$\Rightarrow harm$	*r6:*	$\neg negligent \Rightarrow \neg liable$
r7:	$harm \Rightarrow liable$		

with $r6 \succ r7$. We can then build the following arguments:

A1:	$\Rightarrow \neg guidelines$	B1:	$\Rightarrow guidelines$	C1:	$\Rightarrow harm$
A2:	$A1 \Rightarrow negligent$	B2:	$B1 \Rightarrow \neg negligent$	C2:	$C1 \Rightarrow liable$
		B3:	$B2 \Rightarrow \neg liable$		

The argumentation graph and its grounded {IN, OUT, UND}*-labelling are depicted in Figure 3, in which all arguments are* UND, *except argument* C1.
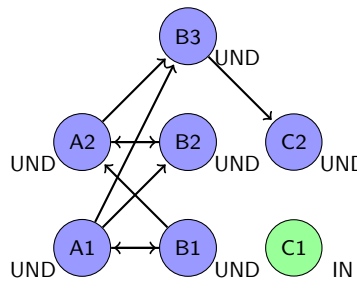

Figure 3: Grounded {IN, OUT, UND}-labelling of Example 3 in the absence of burdens of persuasion.

The result shown in Figure 3 is not satisfactory, according to the law, since it does not take into account the applicable burdens of persuasion. The doctor should have lost the case – i.e., be declared liable – since she failed to discharge her burden of proving she was non-negligent—namely, she didn't satisfy her burden of persuasion for her non-negligence. The doctor's failure results from the fact that it remains uncertain whether she followed the guidelines. To capture this aspect of the argument, we need to provide a model for the burden of persuasion.

□

Example 4 (Criminal law example: self-defence in murder case) *Let us consider a case in which a woman shot a robber in her house, killing him. The applicable law includes the rule according to which intentional killing constitutes murder, as well as the exception according to which there is no murder if the victim was killed in self-defence. Assume that it has been established that the woman shot the robber and that she did so intentionally, i.e, the prosecution has met her burden of persuasion with regard to both killing and intent. However, it remains uncertain, due to conflicting evidence, whether the robber was threatening her with a gun, as claimed by the defence, or was running away after having been discovered, as claimed by the prosecution. Accordingly, let us consider the following rules:*

r1:	$\Rightarrow killed$	*r2:*	$\Rightarrow intention$
r3:	$\Rightarrow threatWithWeapon$	*r4:*	$\Rightarrow \neg threatWithWeapon$
r5:	$threatWithWeapon \Rightarrow selfDefence$	*r6:*	$\neg threatWithWeapon \Rightarrow \neg selfDefence$
r7:	$selfDefence \Rightarrow \neg murder$	*r8:*	$killed, intention \Rightarrow murder$

with r7 ≻ r8. We can build the following arguments:

A1 :	$\Rightarrow killed$	B1 :	$\Rightarrow threatWithWeapon$	C1 :	$\Rightarrow \neg threatWithWeapon$
A2 :	$\Rightarrow intention$	B2 :	$B1 \Rightarrow selfDefence$	C2 :	$C1 \Rightarrow \neg selfDefence$
A3 :	$A1, A2 \Rightarrow murder$	B3 :	$B2 \Rightarrow \neg murder$		

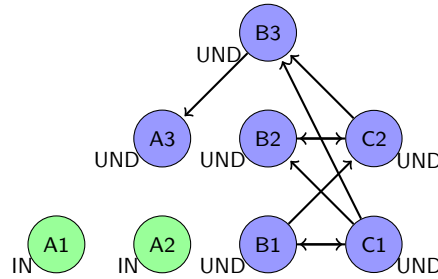*In the grounded* {IN, OUT, UND}*-labelling of Figure 4, all arguments are* UND*, except for the undisputed facts.*

Figure 4: Grounded {IN, OUT, UND}-labelling of Example 4 in the absence of burdens of persuasion.

Also in this case, the labelling does not provide the legally correct answers, namely, acquittal. Acquittal results from the fact that the burden of persuasion falls on the non-existence of self-defence. Prosecution failed to provide a convincing argument for that, since it remains dubious whether the woman who shot the robber was threatened with a weapon by the latter.

4   Labelling and burden of persuasion

Now, let us introduce the concept of burden of persuasion in a second labelling stage.

Notation 4.1 *To indicate that there is a burden of persuasion on every literal in a set of literals* Φ*, we write BP(*Φ*). If* Φ *is a singleton* $\{\phi\}$*, we may simply write BP(*ϕ*).*

The burden of persuasion may be captured in subsequent labelling stages, cf. multi-labelling systems [1]. A burden of persuasion labelling (BP-labelling) based on an {IN, OUT, UND}-labelling assigns OUT* and IN* labels to UND arguments depending on the allocation of the burden of persuasion. IN and OUT arguments maintain their status, i.e., are labelled IN* and OUT* respectively.

4.1   BP-labelling definition

Definition 4.1 *A* {IN*, OUT*, UND*}*-labelling of an argumentation graph G is a total function* $\mathscr{A}_G \rightarrow$ {IN*, OUT*, UND*}.

Definition 4.2 *Let G be an argumentation graph, L the grounded* {IN, OUT, UND}*-labelling of G, and* Φ *a consistent set of literals. A* **BP-labelling** *of G, relative to burdens of persuasion BP(*Φ*), is a* {IN*, OUT*, UND*}*-labelling such that* $\forall (A : A_1, \ldots A_n \Rightarrow_r \phi) \in \mathscr{A}_G$

1. *if* $A \in \text{IN}(L)$ *then* $A \in \text{IN}^*$;

2. *if* $A \in \text{OUT}(L)$ *then* $A \in \text{OUT}^*$;

3. *if* $A \in \text{UND}(L)$ *then*

   (a) $A \in \text{IN}^*$ *iff*

      i. $\bar{\phi} \in \Phi$ *and*
- $\nexists B \in \mathscr{A}_G$ *such that* $Conc(B) = \bar{\phi}$, $B \succ A$ *and* B *is* IN^*
  and
- $\nexists A' \in DirectSub(A) : A'$ *is* OUT^*

      or

      ii. $\bar{\phi} \notin \Phi$ *and*
- $\forall B \in \mathscr{A}_G$ *such that* $Conc(B) = \bar{\phi}$, $A \not\succ B : B$ *is* OUT^*
  and
- $\forall A' \in DirectSub(A) : A'$ *is* IN^*;

   (b) $A \in \text{OUT}^*$ *iff*

      i. $\phi \in \Phi$ *and*
- $\exists B \in \mathscr{A}_G$ *such that* $Conc(B) = \bar{\phi}$, $A \not\succ B$ *and* B *is not* OUT^*
  or
- $\exists A' \in DirectSub(A) : A'$ *is not* IN^*

      or

      ii. $\phi \notin \Phi$ *and*
- $\exists B \in \mathscr{A}_G$ *such that* $Conc(B) = \bar{\phi}$, $B \succ A$ *and* B *is* IN^*
  or
- $\exists A' \in DirectSub(A) : A'$ *is* OUT^*;

   (c) $A \in \text{UND}^*$ *otherwise.*

In Definition 4.2, items *3.(a)* and *3.(b)* concern respectively conditions for acceptance and rejection based on burdens of persuasion.

Condition for acceptance.

- Item *3.(a)(i)* concerns the case in which the burden of persuasion in on the complement $\bar{\phi}$ of the conclusion ϕ of argument A. A counterargument B for $\bar{\phi}$ is disfavoured by the burden of persuasion while A is favoured, i.e., A is to be accepted under conditions of uncertainty. Thus, acceptance of A is not affected by B unless B is strongly superior to A. and acceptance of A is also not affected by the fact that a strict subarguments of A is UND^*, rather than IN^*.

- Item *3.(a)(ii)* concerns the case in which the conclusion of argument A is contradicted by a counterargument B upon which there is no burden of persuasion. Then there is no favour for A. Acceptance of A may be affected by B whenever A is not strictly superior to *B*, and acceptance of A is also affected by the fact that one A's strict subarguments is UND^*.

Condition for rejection.

- Item *3.(b)(i)* concerns the case in which the burden of persuasion is on the conclusion of argument A, so that A is disfavoured by the burden of persuasion. Then the rejection of A may be determined by a counterargument B that is uncertain (UND^*) and also by any uncertainty on one of A's strict subarguments.

- Item *3.(b)(i)* concerns the case in which there is no burden of persuasion on the conclusion of argument A. Then the rejection of A is only determined by a counterargument B which A which is IN* or by a strict subargument of A that is OUT*.

4.2 Completion of a BP-labellings

A BP-labelling may not be complete, in the sense that an argument favoured by the burden of persuasion may be IN* even is some of its subarguments are UND*. As shown in Example 5 (Figure 5 left), an argument (e.g. A2) may be IN* because the burden of persuasion is on the complement of its conclusion, even though some of its subarguments are UND*. This is admitted by Definition 4.2, 3.(a)(i). The intuition is the following: when we have accepted a conclusion ϕ since the burden of persuasion is on $\bar{\phi}$, we may still remain uncertain on the premises for that conclusion. For instance, we may accept that the doctor was negligent since she failed to provide a convincing argument about why she was not negligent, and still be uncertain whether the doctor did not comply with the guidelines.

The opposite approach is also possible: if we accept an argument based on the burden of persuasion, we are also bound to accept all of its subarguments. This approach leads to the concepts of completion and grounding of BP-labellings.

Definition 4.3 *A **completion** L_c **of a BP-labelling** L of an argumentation graph G, relative to burdens BP(Φ), is an* {IN*, OUT*, UND*}*-labelling such that*

1. IN*$(L) \subseteq$ IN*(L_c), *and*

2. OUT*$(L) \subseteq$ OUT*(L_c), *and*

3. $\forall A \in \mathscr{A}_G$
    - *(a)* A $\in$ IN*(L_c) *iff all attackers of* A *are in* OUT*(L_c), *and*
    - *(b)* A $\in$ OUT*(L_c) *iff at least one attacker of* A *is* IN*(L_c).

Definition 4.4 *The **grounding** L_g **of a BP-labelling** L of an argumentation graph G, relative to burdens BP(Φ), is a completion L_c of L such that* IN*(L_c) *is minimal.*

4.3 Discussion of the examples

Let us now apply these definitions to our running examples, and to some further cases.

Example 5 (Burden of persuasion on civil law example: medical malpractice) *Continuing Example 3 presented in Section 3, assume – accordingly to Italian Law – to have BP(¬negligent) (i.e., the doctor has to provide a convincing argument that she was non-negligent, i.e., that she was diligent).*

By Definition 4.2 argument A2 *for the doctor's negligence must be* IN* *and the argument for non-negligence is* OUT*, *but it remains uncertain whether the doctor followed the guidelines (the corresponding arguments* A1 *and* B1 *are* UND*).*

A different result can be obtained by using the grounding of the BP-labelling. Then argument A1 *for ¬guidelines is* IN* *and argument* B1 *for guidelines is* OUT*. *The BP-labelling is shown in Figure 5 (left) and the grounded labelling in Figure 5 (right).*

□

Example 6 (Burden of persuasion on criminal law example: self-defence in murder case) *In the criminal law example, the burden of persuasion is on the prosecution, both for the constitutive elements of the crime and for the non-existence of justifications (like self-defence). Thus, we specify that*
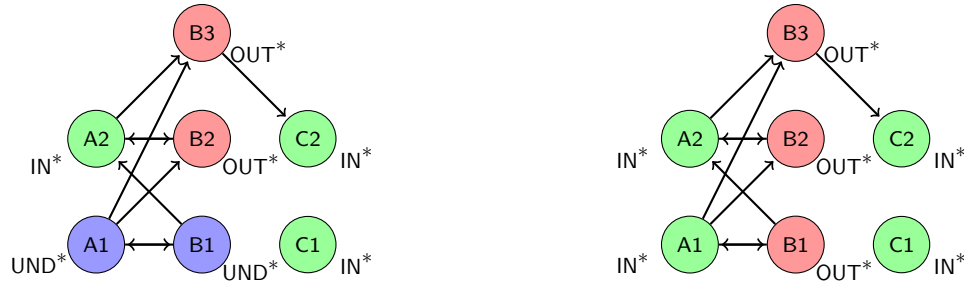
Figure 5: BP-labelling (left) its grounding (right), with BP($\neg negligent$), where $\neg negligent = Conc(\text{B2})$.

there is a burden of persuasion on killing and intent as well as on the absence of self-defence, i.e., BP(\{killed, intention, $\neg$selfDefence\}). Hence, the prosecution has to persuade the judge that there was killing, that it was intentional, and that the killer did not act in self-defence. The grounding is depicted in Figure 6.

Note that while arguments B1 *and* C1 *are* UND* *in the BP-labelling, in the grounding (Figure 6 right),* B1 *is now labelled* IN* *while* C1 *is labelled* OUT*.
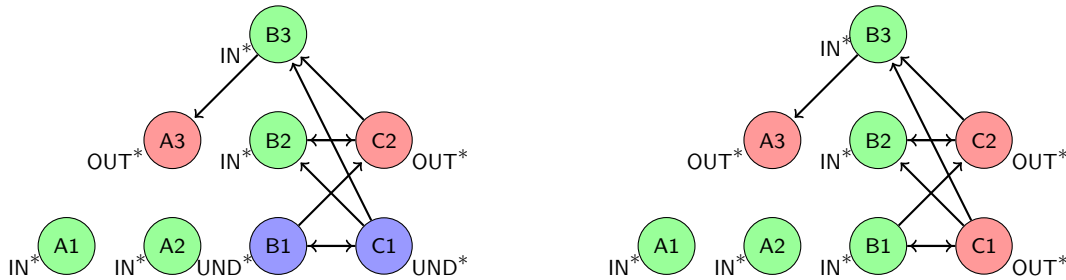

Figure 6: BP-labelling (left) and its grounding (right), with the burden of persuasion BP(\{killed, intention, $\neg$selfDefence\}).

The BP-labelling provides the legally correct answer. Prosecution failed to meet its burden of proving non-self-defence (its argument C2 *is not convincing). Therefore, the argument for self-defence (*B2*) is* IN* *and so is the argument for not-murder (*B3*), which means that the killer is to be acquitted.*

Example 7 (Conflicting burdens of persuasion) *When there are multiple conflicts amongst arguments and subarguments, burdens disfavouring the conclusions of subarguments are evaluated first. As we shall see in the next example, this allows us to capture inversions in the burden of persuasion.*

Let us assume the following premises and rules with BP(\{$\neg a, b$\}):

$$r1: \quad \Rightarrow a \qquad r2: \quad a \Rightarrow b$$
$$r3: \quad \Rightarrow \neg a \qquad r4: \quad \neg a \Rightarrow \neg b$$

We can then build the following arguments:

$$\text{A1}: \quad \Rightarrow a \qquad \text{B1}: \quad \Rightarrow \neg a$$
$$\text{A2}: \quad \text{A1} \Rightarrow b \qquad \text{B2}: \quad \text{B1} \Rightarrow \neg b$$

The corresponding grounding is depicted in Figure 7. By Definition 4.2, A1 *and* A2 *are* IN*, *while* B1 *and* B2 *are* OUT*.
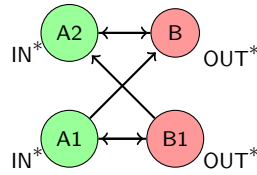
Figure 7: BP-labelling with $BP(\{\neg a, b\})$, where $\neg a = Conc(\text{B1})$ and $b = Conc(\text{A2})$.

Example 8 (Inversion of burden of persuasion: medical malpractice) *In legal and other contexts, often there is an inversion of the burden of persuasion: a burden of persuasion on the conclusion of an argument* A *coexists with a burden of persuasion on the complement of the conclusion of a sub-argument of* A. *Developing Example 5, let us now assume the following: the patient has the burden of persuasion on liability, while the doctor has the burden of persuasion on being non-negligent, i.e., $BP(\{liable, \neg negligent\})$ (negligence being a conclusion of a subargument of the doctor). Accordingly, the patient will lose if he fails to provide a convincing argument for the doctor's liability, which includes providing arguments for all preconditions of such liability. On the latter there is an inversion: it is up to the doctor to provide a convincing argument for her non-negligence.*

From Definition 4.2 it follows that:

- A2 *and* C2 *(concluding negligent and liable respectively) are* IN*, and*
- B2 *and* B3 *(concluding ¬negligent and ¬liable respectively) are* OUT*.*

Completing the labelling we also obtain that A1 *is* IN* *and* B1 *is* OUT*.*

The example shows that, according to Definition 4.2 the burdens of production are applied step by step: the burden are applied first to the assessment of subarguments, and then to the assessment of arguments including them. Since the doctor could not provide a convincing argument for ¬negligent (where she has the burden of persuasion), negligence is established. Therefore the doctor is unable to challenge the patient's liability through an argument based on her non-negligence.
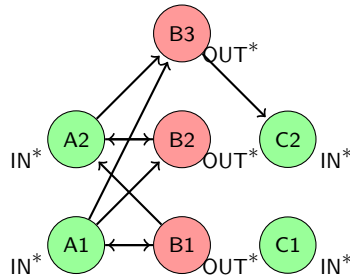

Figure 8: Grounding of the BP-labelling with $BP(\{liable, \neg negligent\})$.

Example 9 (Self-defeating argument) *Now let us consider a self-defeating argument. Let us assume the following premises with $BP(\neg c)$:*

$$\Rightarrow \neg c \qquad \neg c \Rightarrow p \qquad p \Rightarrow c$$

We can build the following arguments:

$$
\begin{aligned}
\text{A1}: &\quad \Rightarrow \neg c \\
\text{A2}: &\quad \text{A1} \Rightarrow p \\
\text{A3}: &\quad \text{A2} \Rightarrow c
\end{aligned}
$$

The BP-labelling is depicted in Figure 9. By Definition 4.2 and item 3.(b)(i), argument A1 *is labelled* OUT*, *if its attacker* A3 *is not* OUT*. *By item 3.(b)(ii), argument* A2 *is labelled* OUT*. *By Definition 4.2, argument* A3 *cannot be labelled* IN*, *and thus it is labelled* UND*. *It turns out that in this case there is no complete BP-labelling, and thus no grounding.*
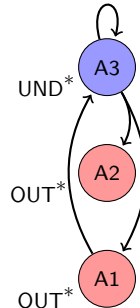

Figure 9: BP-labelling with $BP(\neg c)$ where $\neg c = Conc(A1)$.

5   Conclusion

A formal model of the burden of persuasion has been provided. The model shows how an allocation of the burden of persuasion may induce single outcomes (IN* arguments), in contexts in which the assessment of conflicting arguments would otherwise remain undecided.

Our model is a preliminary exploration of the intersection between the burden of persuasion and argumentation labelling frameworks, and it can provide a starting point for further research. In particular, it combines the insight of [8, 10], where the burden of persuasion provides a criterion for adjudicating conflicts of arguments, and the insight of [4, 5], where the satisfaction of burdens of argumentation depends on the dialectical status of the arguments concerned. The proposed model also deals with situations in which we have to combine a general burden of persuasion for one party, with exceptional propositions for which the burden is shifted onto the other party (so-called inversion of the burden of persuasion, see [10].

The model can be extended to take into account different standards of proof that are required for meeting the burden of persuasion, such as the standard "preponderance of the evidence", which applies in private law, and the more rigorous "beyond reasonable doubt" standard of criminal law (see [10]).

References

[1] Pietro Baroni & Régis Riveret (2019): *Enhancing Statement Evaluation in Argumentation via Multi-labelling Systems*. Journal Artificial Intelligence Research 66, pp. 793–860, doi:10.1613/jair.

[2] Roberta Calegari, Giovanni Ciatto, Enrico Denti & Andrea Omicini (2020): *Logic-based Technologies for Intelligent Systems: State of the Art and Perspectives*. Information 11(3), pp. 1–29, doi:10.3390/info11030167. Available at `http://www.mdpi.com/2078-2489/11/3/167`. Special Issue "10th Anniversary of Information—Emerging Research Challenges".

[3] Martin Caminada & Leila Amgoud (2007): *On the Evaluation of Argumentation Formalisms*. Artificial Intelligence 171(5—6), pp. 286–310, doi:10.1016/j.artint.2007.02.003.

[4] Thomas F. Gordon, Henry Prakken & Douglas Walton (2007): *The Carneades model of argument and burden of proof*. Artificial Intelligence 171(10), pp. 875–896, doi:10.1016/j.artint.2007.04.010.

[5] Thomas F. Gordon & Douglas N. Walton (2009): *Proof Burdens and Standards*. In Iyad Rahwan & Guillermo R. Simari, editors: *Argumentation in Artificial Intelligence*, Springer, Boston, MA, pp. 239–60, doi:10.1007/978-0-387-98197-0_12.

[6] Sanjay Modgil & Henry Prakken (2014): *The* ASPIC$^+$ *framework for structured argumentation: a tutorial*. Argument & Computation 5(1), pp. 31–62, doi:10.1080/19462166.2013.869766.

[7] Henry Prakken (2010): *An Abstract Framework for Argumentation with Structured Arguments*. Argument and Computation 1, pp. 93–124, doi:10.1080/19462160903564592.

[8] Henry Prakken & Giovanni Sartor (2008): *More on Presumptions and Burdens of Proof*. In: *21th Annual Conference on Legal Knowledge and Information Systems*, IOS, Groningen, The Netherlands, pp. 176–85, doi:10.2139/ssrn.1317348.

[9] Henry Prakken & Giovanni Sartor (2010): *A Logical Analysis of Burdens of Proof*. Legal Evidence and Proof: Statistics, Stories, Logic 1, pp. 223–253. Available at http://hdl.handle.net/11370/fe1fb134-a25b-4d5b-b7b7-bd7d408e80c9.

[10] Henry Prakken & Giovanni Sartor (2011): *On Modelling Burdens and Standards of Proof in Structured Argumentation*. In: *24th Annual Conference on Legal Knowledge and Information Systems*, IOS, Vienna, Austria, pp. 83–92, doi:10.3233/978-1-60750-981-3-83.

[11] Gerard Vreeswijk (1997): *Abstract Argumentation Systems*. Artificial Intelligence 90(1–2), pp. 225–279, doi:10.1016/S0004-3702(96)00041-0.

[12] Douglas Walton (2014): *Burden of proof, presumption and argumentation*. Cambridge University Press, USA, doi:10.1017/CBO9781107110311.