A labelling framework for probabilistic argumentation



*Published Version:*
A labelling framework for probabilistic argumentation / Riveret, Régis*; Baroni, Pietro; Gao, Yang; Governatori, Guido; Rotolo, Antonino; Sartor, Giovanni. - In: ANNALS OF MATHEMATICS AND OF ARTIFICIAL INTELLIGENCE. - ISSN 1012-2443. - STAMPA. - 83:1(2018), pp. 21-71. [10.1007/s10472-018-9574-1]

*Availability:*
This version is available at: https://hdl.handle.net/11585/644602 since: 2019-11-26

*Published:*
DOI: http://doi.org/10.1007/s10472-018-9574-1

Print ISSN 1012-2443; Online ISSN 1573-7470

Riveret, R., Baroni, P., Gao, Y. et al. *Ann Math Artif Intell* May 2018, Volume 83, Issue 1, pp 21-71.

The final authenticated version is available online at: https://dx.doi.org/10.1007/s10472-018-9574-1

# A Labelling Framework for Probabilistic Argumentation

Régis Riveret, *CSIRO, Australia*
Pietro Baroni, *University of Brescia, Italy*
Yang Gao, *Chinese Academy of Sciences, China*
Guido Governatori, *CSIRO, Australia*
Antonino Rotolo, *University of Bologna, Italy*
Giovanni Sartor, *European University Institute, Italy*

**Abstract**

The combination of argumentation and probability paves the way to new accounts of qualitative and quantitative uncertainty, thereby offering new theoretical and applicative opportunities. Due to a variety of interests, probabilistic argumentation is approached in the literature with different frameworks, pertaining to structured and abstract argumentation, and with respect to diverse types of uncertainty, in particular the uncertainty on the credibility of the premises, the uncertainty about which arguments to consider, and the uncertainty on the acceptance status of arguments or statements. Towards a general framework for probabilistic argumentation, we investigate a labelling-oriented framework encompassing a basic setting for rule-based argumentation and its (semi-) abstract account, along with diverse types of uncertainty. Our framework provides a systematic treatment of various kinds of uncertainty and of their relationships and allows us to back or question assertions from the literature.

## 1 Introduction

Argumentation aims at supporting rational persuasion and deliberation in domains where no conclusive logical proofs are available. It addresses defeasible claims raised on the basis of partial, uncertain and possibly conflicting pieces of information. Argumentation has been a traditional concern for philosophy and legal theory since Aristotle and Cicero, but only recently has become a focus for research in artificial intelligence [1, 6].

There exists a variety of formal models capturing different aspects of the complexity of argumentation activities. For instance, argument construction models (either rule-based or logic-based) [40, 63, 21] need to be integrated with formal approaches for the evaluation of the acceptance status of arguments, possibly at a more abstract level [12, 2]. Moreover, argument acceptance needs to be projected on the statements corresponding to their conclusions, in order to assess their acceptance in turn. In general, different aspects and their models need to be combined together to get a satisfactory coverage of the full picture.

While argumentation-based reasoning enhances a system's ability to reach defeasible determinations under conditions of uncertainty and conflict, most formal approaches to argumentation suffer a substantial limitation: they are unable to provide a comparative quantitative account of the persuasiveness of alternative conclusions. Typically, argumentation systems either identify a single skeptical outcome, or propose a set of credulous alternatives, without specifying corresponding degrees of credibility.

It has been convincingly argued that probability theory can help to fill the gap between qualitative and quantitative accounts of uncertainty. For example, probabilistic methods may propagate quantified uncertainties over (clusters of) premises into quantified uncertainties concerning the dialectical acceptability of arguments and conclusions. Hence, the combination of formal argumentation and probability theory has received increasing attention in recent years.

Different types of uncertainty, which can be the subject of probabilistic evaluation, can be identified in argumentation. For instance, there may be uncertainty about which premises to believe to construct arguments. There can also be uncertainty about which arguments and relationships to consider when an abstract representation of arguments and their relations is adopted. Then there can be further uncertainty about the outcomes of the assessment, namely the acceptance status of arguments and statements. This list is not exhaustive and all kinds of uncertainty are potentially appreciable and meaningful.

Since different types of probabilistic uncertainty can be discerned and argumentation can be dealt with at different level of abstraction, different families of approaches to probabilistic argumentation can be found in the literature. First probabilistic notions have been investigated at different abstraction levels, ranging from structured argumentation formalisms [47, 52, 48, 13] to abstract argumentation frameworks [36, 59, 27, 30, 49, 50, 51]. Further, one can distinguish approaches based on the uncertainty concerning the arguments to actually include in the argumentation process, as in the so-called 'constellations' approach [36, 27], and approaches focusing on a probabilistic notion of the acceptance status of arguments, as in the 'epistemic' approach [59, 27, 30]. Existing approaches are thus heterogeneous and complementary, leading to the goal of setting up a general and unified framework able to cover multiple types of probabilistic uncertainty in structured and abstract argumentation.

**Contribution.** Towards a general framework for probabilistic argumentation, we present a labelling-oriented framework that we call *probabilistic labellings* and which covers uncertainty on inclusion of argumentative pieces as well as uncertainty regarding acceptance of arguments or statements, even in the case all the argumentative pieces are included in the reasoning activity.

To establish the framework of probabilistic labellings, we approach probabilistic argumentation at various levels of abstraction. We start from a basic rule-based argumentation setting, then give a semi-abstract account of it, and finally devise probabilistic labellings of arguments, and of statements (an issue which has received limited attention in previous literature). Along this journey, the main contributions are as follows:

- we introduce some non-traditional argument labellings, going beyond the three labels IN, OUT and UN commonly used in the literature, to capture different kinds of uncertainty in a unified representation;
- we extend this unified labelling-based representation to statements corresponding to argument conclusions;
- we define a comprehensive set of probabilistic frames corresponding to different kinds of uncertainty at different abstraction levels and analyse the relationships betwen them;
- we analyse the enhanced expressiveness of probabilistic labellings, with particular reference to some influential probabilistic argumentation approaches in the literature.

We will refrain from discussing possible interpretations of probability values (such as classical, frequentist or Bayesian views on these matters), and we will not investigate algorithms to compute the probability of the status of arguments and statements, but redirect the readers to some relevant literature.

**Outline.** The paper is organised as follows. In Section 2, we introduce a minimalist rule-based argumentation framework and a (semi-) abstract account featuring argumentation graphs. On this basis we then introduce our generalised labelling approach for arguments and statements. In Section 3, we present a set of probabilistic frames for argumentation, capturing uncertainty at different stages of the argumentation process and leading to the main notion of probabilistic labellings of arguments and statements. Section 4 analyses our proposal and its technical properties in relation with two influential approaches to probabilistic argumentation, namely the 'constellations' approach and the 'epistemic' approach. We discuss relevant literature in Section 5, before concluding in Section 6. To help the reader, key notations are summarised at the end of the paper.

# 2 Labelling of Arguments and Statements

We assume a generic argumentation process which can be summarised as follows: arguments are built out of a rule base; their relationships induce an argumentation graph; argument acceptance is assessed on the basis of the argumentation graph through labelling semantics; and statement acceptance is assessed on the basis of argument acceptance labellings. This section introduces a formalisation of the various steps of the above process, which represents a necessary basis for our approach to probabilistic argumentation. In particular, Subsection 2.1 introduces a simple rule-based formalism and the relevant graph representation, while subsections 2.2 and 2.3 deal with argument and statement labellings respectively. The overall schematisation of the argumentation process rests on the existing literature and is not novel *per se*, while labellings to handle at the same time argument or statement inclusion and (if included) argument or statement acceptance, as proposed in subsections 2.2 and 2.3, provide an original generalisation of the traditional usage of labellings in the literature.

## 2.1 Argument construction and argumentation graphs

We first present a rule-based argumentation framework and its (semi-) abstract account, specifying the structures on which we will develop our proposal. This rule-based argumentation framework is 'minimalist', that is, we avoid overloading it with features which are not relevant for our approach to probabilistic argumentation. As for reference, this rule-based setting is akin to an argumentative interpretation of Defeasible Logic [22, 34], with some definitions inspired by the ASPIC$^+$ framework [40]. Eventually, Dung's abstract argumentation graphs [12] are slightly adapted to capture this rule-based setting in an abstract way and to characterise different probabilistic argumentation frameworks (see also [35] for an approach mapping ASPIC$^+$ frameworks to a special kind of abstract frameworks, called Extended Evidential Argumentation Frameworks).

The building blocks of the argumentation framework are atomic formulas (or atoms, i.e. formulas with no propositional structures), from which literals are considered.

**Definition 2.1 (Literal)** *A* literal *is an atomic formula or the negation of an atomic formula.*

**Definition 2.2 (Complementary literal)** *Given a literal $\varphi$, its* complementary literal *is a literal, denoted as $-\varphi$, such that if $\varphi$ is an atom $p$ then $-\varphi$ is its negation $\neg p$, and if $\varphi$ is $\neg q$ then $-\varphi$ is $q$.*

We distinguish the negation $\neg$ from the negation as failure, denoted $\sim$, according to which $\sim \varphi$ indicates the failure to derive $\varphi$.

The next construct regards rules that relate literals. For the sake of simplicity, we have only defeasible rules, i.e. in our context, rules that can be defeated by other rules.

**Definition 2.3 (Defeasible rule)** *A* defeasible rule *is a construct of the form:*

$$r : \varphi_1, \ldots, \varphi_n, \sim \varphi'_1, \ldots, \sim \varphi'_m \Rightarrow \varphi$$

*where*

- *$r$ is the unique identifier of the rule;*
- *$\varphi_1, \ldots, \varphi_n, \sim \varphi'_1, \ldots, \sim \varphi'_m$ ($0 \leq n$ and $0 \leq m$) is the body of the rule, and all the $\varphi_i$ and $\varphi'_j$ are literals;*
- *$\varphi$ is the consequent (or head) of the rule, which is a single literal.*

**Notation 2.1** *The body of a rule $r$ is denoted $Body(r)$. So given a rule as in Definition 2.3, we have $Body(r) = \{\varphi_1, \ldots, \varphi_n, \sim \varphi'_1, \ldots, \sim \varphi'_m\}$.*

Rules may lead to conflicting literals (we ensure later that two conflicting literals cannot be both accepted). For this reason, we assume that a conflict relation is defined over the set of literals to express conflicts in addition to those corresponding to negation.

**Definition 2.4 (Conflict relation)** *Given a set of literals $\Phi$, a* conflict relation *'$conflict$' is a binary relation over $\Phi$, i.e. $conflict \subseteq \Phi \times \Phi$, such that for any $\varphi_1, \varphi_2 \in \Phi$, if $\varphi_1$ and $\varphi_2$ are complementary, i.e. $\varphi_1 = -\varphi_2$, then $\varphi_1$ and $\varphi_2$ are in conflict, i.e. $(\varphi_1, \varphi_2) \in conflict$.*

The conflict relation may be further specified. For example, the relation may be refined with asymmetric and symmetric conflicts to deal with contrary or contradictory literals as in [40]. However, treating in detail such sophistications is not necessary for our purposes.

When two rules have heads in conflict, one rule may prevail over another one. Informally, a rule superiority relation $r_1 \succ r_2$ states that the rule $r_1$ prevails over the rule $r_2$.

**Definition 2.5 (Superiority relation)** *Let $Rules$ be a set of rules. A* superiority relation *$\succ$ is a binary relation over $Rules$, i.e. $\succ \subseteq Rules \times Rules$, with $r \succ r'$ denoting that $r$ is superior to $r'$.*

The superiority relation may enjoy some particular properties. For example, it may be antireflexive and antisymmetric, so that for a rule $r$ it does not hold that $r \succ r$ and for two distinct rules $r$ and $r'$, we cannot have both $r \succ r'$ and $r' \succ r$. However, as for the conflict relation, treating in detail such sophistications is not necessary for our purposes.

From a set of rules, a conflict relation and a superiority relation, we can define defeasible theories, cf. [22, 34].

**Definition 2.6 (Defeasible theory)** *A* defeasible theory *is a tuple $\langle Rules, conflict, \succ\rangle$ where $Rules$ is a set of rules, $conflict$ is a conflict relation over literals, and $\succ$ is a superiority relation over rules.*

By chaining rules of a defeasible theory, we can build arguments as defined below, cf. [40].

**Definition 2.7 (Argument)** *An* argument *$A$ constructed from a defeasible theory $\langle Rules, conflict, \succ\rangle$ is a finite construct of the form:*

$$A:\ A_1, \ldots A_n, \sim\varphi_1, \ldots, \sim\varphi_m \Rightarrow_r \varphi$$

*where*

- *$A$ is the unique identifier of the argument;*
- *$0 \leq n$ and $0 \leq m$, and $A_1, \ldots, A_n$ are arguments constructed from the defeasible theory $\langle Rules, conflict, \succ\rangle$;*
- *$\varphi$ is the conclusion of the argument $A$. The conclusion of an argument $A$ is denoted $conc(A)$, i.e. $conc(A) = \varphi$;*
- *$\exists r \in Rules$ such that $r : conc(A_1), \ldots, conc(A_n), \sim\varphi_1, \ldots, \sim\varphi_m \Rightarrow \varphi$.*

**Definition 2.8 (Subarguments and rules)** *Given an argument $A:\ A_1, \ldots A_n, \sim\varphi_1, \ldots, \sim\varphi_m \Rightarrow_r \varphi$, the set of its subarguments $Sub(A)$, the set of its direct subarguments $DirectSub(A)$, the last inference rule $TopRule(A)$, and the set of all the rules in the argument $Rules(A)$ are defined as follows:*

- *$Sub(A) = Sub(A_1) \cup \ldots \cup Sub(A_n) \cup \{A\}$;*
- *$DirectSub(A) = \{A_1, \ldots, A_n\}$;*
- *$TopRule(A) = (r : conc(A_1), \ldots, conc(A_n), \sim\varphi_1, \ldots, \sim\varphi_m \Rightarrow \varphi)$;*
- *$Rules(A) = Rules(A_1) \cup \ldots \cup Rules(A_n) \cup \{TopRule(A)\}$.*

**Notation 2.2** *The set of arguments constructed on the basis of a theory $T$ is denoted $Args(T)$.*

According to Definition 2.7, an argument without subarguments has thus the form $A : \sim\varphi_1, \ldots, \sim\varphi_m \Rightarrow_r \varphi$ with $(0 \leq m)$. 'Infinite' arguments are not considered, but we may have an infinite set of finite arguments constructed from a defeasible theory.

**Example 2.1 (Running program)** *Suppose a research scientist is managing a project critically depending on a program installed on a distant machine which is powered by a solar panel and a battery. This program runs intermittently. The scientist has no explicit information about whether the program is running, thus, a priori, its running status is undecided. However, if the scientist receives the notifications that there is no solar power and no power from the battery, then the machine does not have enough power, and therefore the program is not working. Such setting can give rise to different scenarios, and to capture such scenarios, we refer to the literals given in Table 1.*

Table 1: Literals and their meanings

| | | | |
|---|---|---|---|
| $\neg$b1 | No solar power. | $\neg$b | No power. |
| $\neg$b2 | No battery power. | c | The program is running. |

*Based on these literals, we may form the defeasible theory* $\mathsf{T} = \langle \mathsf{Rules}, \mathsf{conflict}, \succ \rangle$ *where:*

$$\begin{aligned} \mathsf{Rules} = \{\ & \mathsf{r}_{\neg \mathsf{b1}}: && \Rightarrow \neg\mathsf{b1}, \\ & \mathsf{r}_{\neg \mathsf{b2}}: && \Rightarrow \neg\mathsf{b2}, \\ & \mathsf{r}_{\neg \mathsf{b}}: \ \neg\mathsf{b1}, \neg\mathsf{b2} && \Rightarrow \neg\mathsf{b}, \\ & \mathsf{r}_{\mathsf{c}}: \ \sim \neg\mathsf{b} && \Rightarrow \mathsf{c}, \\ & \mathsf{r}_{\neg \mathsf{c}}: && \Rightarrow \neg\mathsf{c} \quad \}, \end{aligned}$$

$$\mathsf{conflict} = \{\ (\mathsf{c}, \neg\mathsf{c}), (\neg\mathsf{c}, \mathsf{c})\ \},$$

$$\succ \ = \ \emptyset.$$

*Based on the theory* $\mathsf{T}$*, we can build the following arguments:*

$$\begin{array}{llll} \mathsf{B1}: & \Rightarrow_{\mathsf{r}_{\neg \mathsf{b1}}} \neg\mathsf{b1} & \mathsf{C}: \ \sim\neg\mathsf{b} & \Rightarrow_{\mathsf{r}_{\mathsf{c}}} \mathsf{c} \\ \mathsf{B2}: & \Rightarrow_{\mathsf{r}_{\neg \mathsf{b2}}} \neg\mathsf{b2} & \mathsf{D}: & \Rightarrow_{\mathsf{r}_{\neg \mathsf{c}}} \neg\mathsf{c} \\ \mathsf{B}: \ \mathsf{B1}, \mathsf{B2} & \Rightarrow_{\mathsf{r}_{\neg \mathsf{b}}} \neg\mathsf{b} & & \end{array}$$

*This example illustrates the concept of defeasible theories and the construction of arguments. We will continue this example later in our probabilistic setting to determine a probability degree attached to the acceptance that the program is running.* □

Arguments may conflict and thus attacks between arguments may appear. We consider two types of attacks: rebuttals (clash of incompatible conclusions) and undercuttings[1] (attacks on negation as failure premises). In regard to rebuttals, we assume that there is a preference relation over arguments determining whether two rebutting arguments mutually attack each other or only one of them (being preferred) attacks the other. The preference relation over arguments can be defined in various ways on the basis of the preference over rules. For instance one may adopt the simple last-link ordering according to which an argument $A$ is preferred over another argument $B$, denoted as $A \succ B$, if, and only if, the rule $TopRule(A)$ is superior to the rule $TopRule(B)$, i.e. $TopRule(A) \succ TopRule(B)$. We make no assumptions over the specific argument preference relation which is adopted. This leads us to adopt the following definition of attack.

**Definition 2.9 (Attack relation)** *An* attack relation $\rightsquigarrow$ *over a set of arguments* $\mathcal{A}$ *is a binary relation over the set* $\mathcal{A}$*, i.e.* $\rightsquigarrow \subseteq \mathcal{A} \times \mathcal{A}$*. An argument* $B$ *attacks an argument* $A$*, i.e.* $B \rightsquigarrow A$*, if, and only if,* $B$ *rebuts or undercuts* $A$*, where*

- $B$ *rebuts* $A$ *(on* $A'$*) if, and only if,* $\exists A' \in Sub(A)$ *such that* $conc(B)$ *and* $conc(A')$ *are in conflict, i.e.* $conflict(conc(B), conc(A'))$*, and* $A' \not\succ B$*;*
- $B$ *undercuts* $A$ *(on* $A'$*) if, and only if,* $\exists A' \in Sub(A)$ *such that* $\sim conc(B)$ *belongs to the body of* $TopRule(A')$*, i.e.* $(\sim conc(B)) \in Body(TopRule(A'))$*.*

[1]The term undercutting is overloaded in argumentation literature and is used with different meanings in different contexts, cf. [40].

On the basis of arguments (and their subarguments) and attacks between arguments, we extend the common definition of an argumentation framework [12] by introducing argumentation graphs which comprise both attack and subargument relations. This is motivated by our upcoming probabilistic account where the subargument relation, encompassed by most structured argumentation formalisms, plays a key role. Intuitively, an argument cannot exist without its subarguments and can only be believed if also its subarguments are believed.

From the definition of the sets of direct subarguments of an argument (see Definition 2.8), we can straightforwardly define a direct subargument relation.

**Definition 2.10 (Direct subargument relation)** *A* direct subargument relation $\Rightarrow$ *over a set of arguments $\mathcal{A}$ is a binary relation over the set $\mathcal{A}$, i.e. $\Rightarrow \subseteq \mathcal{A} \times \mathcal{A}$. An argument $B$ is a direct subargument of $A$, i.e. $B \Rightarrow A$, if, and only if, $B$ belongs to the set of direct subarguments of $A$, i.e. $B \in DirectSub(A)$.*

By Definition 2.7, an argument is not a direct subargument of itself and cannot be a subargument of its direct subarguments. Therefore, the direct subargument relation is antireflexive and acyclic.

Arguments and attack relations can be then captured in Dung's abstract argumentation graphs, called abstract argumentation frameworks in [12].

**Definition 2.11 (Abstract argumentation graph)** *An* abstract argumentation graph *is a tuple $\langle \mathcal{A}, \rightsquigarrow \rangle$ where $\mathcal{A}$ is a set of arguments, and $\rightsquigarrow$ is an* attack *relation over $\mathcal{A}$.*

It is equally possible to refer to semi-abstract argumentation graphs, where both the attack and the subargument relations are encompassed, cf. [46].

**Definition 2.12 (Semi-abstract argumentation graph)** *A* semi-abstract argumentation graph *is a tuple $\langle \mathcal{A}, \rightsquigarrow, \Rightarrow \rangle$ where $\mathcal{A}$ is a set of arguments, $\rightsquigarrow$ is an* attack *relation over $\mathcal{A}$, and $\Rightarrow$ is a* direct subargument *relation over $\mathcal{A}$.*

**Notation 2.3** *Given a semi-abstract argumentation graph $G = \langle \mathcal{A}, \rightsquigarrow, \Rightarrow \rangle$, we denote $\mathcal{A}$ as $\mathcal{A}_G$, $\rightsquigarrow$ as $\rightsquigarrow_G$ and $\Rightarrow$ as $\Rightarrow_G$.*

Unconstrained semi-abstract argumentation graphs can be problematic. For instance cycles of the subargument relations are not allowable. Moreover, as usual in structured argumentation formalisms and reflecting the fundamental dependence of an argument on its subarguments, each attack against a subargument is meant to be an attack against all its superarguments. Accordingly, taking inspiration from [46, 14], we introduce the notion of *well-formed* argumentation graphs.

**Definition 2.13 (Well-formed semi-abstract argumentation graph)** *A semi-abstract argumentation graph $\langle \mathcal{A}, \rightsquigarrow, \Rightarrow \rangle$ is* well-formed *if, and only if:*

- *the relation $\Rightarrow$ is acyclic and antireflexive;*
- *if an argument $A$ attacks an argument $B$, and $B$ is a direct subargument of an argument $C$, then $A$ attacks $C$.*

In a well-formed semi-abstract argumentation graph it is easy to see that, by recursion, if an argument $A$ attacks an argument $B$ then $A$ attacks also all the superarguments of $B$.

**Definition 2.14 (Argumentation graph constructed from a theory)** *An (abstract or semi-abstract) argumentation graph $G$ is constructed from a defeasible theory $T$ if, and only if, $\mathcal{A}_G$ is the set of all arguments constructed from $T$.*

**Notation 2.4** *An argumentation graph constructed from a defeasible theory $T$ is denoted by $G_T$.*

From the facts that we exclude cyclic arguments and that the attack relation (Definition 2.9) refers to subarguments, we have the following proposition.

**Proposition 2.1** *A semi-abstract argumentation graph constructed from a defeasible theory is well-formed.*

**Proof 2.1** *The relation $\Rightarrow$ is acyclic and antireflexive. By Definition 2.9, if an argument $A$ attacks an argument $B$ then $A$ attacks all the superarguments of $B$, and thus $A$ attacks every argument $C$ such that $B$ is a direct subargument of $C$. Therefore, if an argument $A$ attacks an argument $B$, and $B$ is a direct subargument of an argument $C$, then $A$ attacks $C$.*

In the remainder, we assume that all semi-abstract argumentation graphs are well-formed, and accordingly, the qualification 'well-formed' is omitted.

**Example 2.1 (continuing from p. 5)** *Using the definition of attacks and direct subarguments, we can construct the semi-abstract argumentation graph as pictured in Figure 1.*

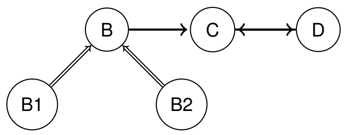


Figure 1: A semi-abstract argumentation graph. Argument B attacks C, arguments C and D attack each other. Arguments B1 and B2 are direct subarguments of argument B.

□

The definition of semi-abstract argumentation graphs allows us to give an account of probabilistic argumentation which is intermediate between Dung's abstract argumentation, encompassing attacks only, and structured argumentation formalisms, where both the attack and the subargument relations are defined in formalism-specific terms. Whilst this semi-abstract setting is built on the simple formalism based on definitions 2.1-2.10, it can be more generally seen as an abstraction basis to encompass more sophisticated rule-based argumentation systems (e.g. ASPIC$^+$ [40]) or related non-monotonic logic frameworks (e.g. Defeasible Logic [22]).

Semi-abstract argumentation graphs can be related to some existing proposals on different notions of supports. For example, the representation is similar to the notion of *inference graph* introduced by [44]. *Bipolar argumentation frameworks* [8] are an example of another semi-abstract settings in the literature encompassing two relations between arguments, namely attack and support. The generic notion of support in bipolar argumentation frameworks allows for different interpretations (namely deductive support, necessary support, and evidential support) with different properties and implications on the attack relation. A specific notion of *evidential support* is considered in *evidential argumentation systems* [42], where a special argument, denoted as $\eta$, is used to represent evidence (namely 'incontrovertible premises') and evidential support, rooted in $\eta$, is 'propagated' through a set-based support relation. Relationships and possibilities of translation between evidential argumentation systems and argumentation frameworks with necessities are investigated in [43]. Our proposal does resort neither to specific notions, like incontrovertible premises, nor to a specific evidence-based interpretation of the notion of support. Our approach directly relies on the subargument relation which has a univocal interpretation in our context and is sufficient to develop a general theory of probabilistic labellings. The inclusion of more articulated notions, like recursive attacks and supports [3, 9], is beyond the scope of this work and represents an interesting direction of future research.

Given an argumentation graph, some of the arguments may be omitted due to the use of some selection criterion: this leads to consider subgraphs (which will be useful to specify fundamental statuses of arguments in the next section.)

**Definition 2.15 (Subgraph)** *Let $G = (\mathcal{A}_G, \rightsquigarrow_G, \Rightarrow_G)$ denote an argumentation graph. The* subgraph *$H$ of $G$ induced by a set of arguments $\mathcal{A}_H \subseteq \mathcal{A}_G$ is an argumentation graph such that $H = (\mathcal{A}_H, \rightsquigarrow_G \cap(\mathcal{A}_H \times \mathcal{A}_H), \Rightarrow_G \cap(\mathcal{A}_H \times \mathcal{A}_H))$.*

**Notation 2.5** *The set of all subgraphs of an argumentation graph G is denoted $Sub(G)$, i.e. $Sub(G) = \{(\mathcal{A}_H, \rightsquigarrow_G \cap(\mathcal{A}_H \times \mathcal{A}_H), \Rightarrow_G \cap(\mathcal{A}_H \times \mathcal{A}_H)) \mid \mathcal{A}_H \subseteq \mathcal{A}_G\}$.*

Not all subgraphs may be deemed to be well-formed or interesting, since only some of them correspond to sensible constructions from a set of rules. For this reason we introduce subargument-complete and rule-complete sets of arguments.

**Definition 2.16 (Subargument-complete set)** *Let $G = (\mathcal{A}_G, \rightsquigarrow_G, \Mapsto_G)$ denote an argumentation graph. A set of arguments $\mathcal{A}_H \subseteq \mathcal{A}_G$ is* subargument-complete *if for any argument $A$ in $\mathcal{A}_H$, all the direct subarguments of $A$ are in $\mathcal{A}_H$, i.e. $\forall A \in \mathcal{A}_H$, if $((B, A) \in \Mapsto_G)$ then $B \in \mathcal{A}_H$.*

**Definition 2.17 (Rule-complete set)** *Let $G = (\mathcal{A}_G, \rightsquigarrow_G, \Mapsto_G)$ denote an argumentation graph. A set of arguments $\mathcal{A}_H \subseteq \mathcal{A}_G$ is* rule-complete *if for all arguments $B$ in $\mathcal{A}_G$ such that all the direct subarguments of $B$ are in $\mathcal{A}_H$ and there exists an argument $A$ in $\mathcal{A}_H$, such that the top rule of $B$ belongs to the rules of $A$, then $B$ belongs to $\mathcal{A}_H$, i.e. $\forall B \in \mathcal{A}_G$, if $DirectSub(B) \subseteq \mathcal{A}_H$ and $\exists A \in \mathcal{A}_H$ such that $TopRule(B) \in Rules(A)$, then it holds that $B \in \mathcal{A}_H$.*

Notice that a set of arguments can be subargument-complete without being rule-complete and vice versa. Therefore these two notions are complementary.

Subargument-complete or rule-complete sets of arguments can be straightforwardly used to qualify subgraphs of an argumentation graph.

**Definition 2.18 (Subargument-complete subgraph)** *A subgraph $H = (\mathcal{A}_H, \rightsquigarrow_H, \Mapsto_H)$ of an argumentation graph $G$ induced by a set of arguments $\mathcal{A}_H$ is* subargument-complete *if, and only if, $\mathcal{A}_H$ is subargument-complete.*

In other words, $H$ is a subargument-complete subgraph of $G$ if any argument of $H$ appears with all its supporting arguments and $H$ has exactly the attacks and supports that appear in $G$ over the same set of arguments.

**Example 2.1 (continuing from p. 5)** *In Figure 2 the graph (a) is a subargument-complete subgraph of the graph in Figure 1, while the graph (b) is not.*


Figure 2: Subargument-completeness.

□

**Definition 2.19 (Rule-complete subgraph)** *A subgraph $H = \langle \mathcal{A}_H, \rightsquigarrow_H, \Mapsto_H \rangle$ of an argumentation graph $G_T$ built from a defeasible theory $T$, and induced by a set of arguments $\mathcal{A}_H$, is* rule-complete *if, and only if, $\mathcal{A}_H$ is rule-complete.*

**Example 2.2 (Abstract example)** *Consider a defeasible theory* $\mathsf{T} = \langle \mathsf{Rules}, \emptyset, \emptyset \rangle$ *such that:*

$$\begin{array}{lllllll} \mathsf{Rules} = \{ & \mathsf{r}_1 : & \Rightarrow \mathsf{a}, & \mathsf{r}_2 : & \Rightarrow \mathsf{b}, & \\ & \mathsf{r}_3 : \ \mathsf{a} & \Rightarrow \mathsf{b}, & \mathsf{r}_4 : \ \mathsf{b} & \Rightarrow \mathsf{c} & \}. \end{array}$$

*We can build the following arguments:*

$$\begin{array}{llllll} \mathsf{A}: & \Rightarrow_{\mathsf{r1}} \mathsf{a}, & \mathsf{AB}: & \Rightarrow_{\mathsf{r1}} \mathsf{a} \Rightarrow_{\mathsf{r3}} \mathsf{b}, & \mathsf{ABC}: & \Rightarrow_{\mathsf{r1}} \mathsf{a} \Rightarrow_{\mathsf{r3}} \mathsf{b} \Rightarrow_{\mathsf{r4}} \mathsf{c} \\ \mathsf{B}: & \Rightarrow_{\mathsf{r2}} \mathsf{b}, & \mathsf{BC}: & \Rightarrow_{\mathsf{r2}} \mathsf{b} \Rightarrow_{\mathsf{r4}} \mathsf{c}. & & \end{array}$$

*The direct subargument relation $\Mapsto$ is such that:*

$$\Mapsto = \{(\mathsf{A}, \mathsf{AB}), (\mathsf{AB}, \mathsf{ABC}), (\mathsf{B}, \mathsf{BC})\}.$$

*Consequently,*

- *the argumentation graph* $\mathsf{G_T} = \langle\{\mathsf{A}, \mathsf{B}, \mathsf{AB}, \mathsf{BC}, \mathsf{ABC}\}, \emptyset, \Mapsto\rangle$ *is rule-complete and subargument-complete;*
- *the argumentation subgraph* $\mathsf{H} = \langle\{\mathsf{A}, \mathsf{B}, \mathsf{AB}, \mathsf{BC}\}, \emptyset, \{(\mathsf{A}, \mathsf{AB}), (\mathsf{B}, \mathsf{BC})\}\rangle$ *is not rule-complete, but it is subargument-complete.*

*This example shows that a subgraph of an argumentation graph built from a defeasible theory can be subargument-complete without being rule-complete.* □

**Proposition 2.2** *If an argumentation graph is constructed from a defeasible theory then it is subargument-complete and rule-complete.*

**Proof 2.2** *From Definition 2.14, an* argumentation graph $G$ *constructed from a defeasible theory* $T$ *is a tuple* $\langle\mathcal{A}, \rightsquigarrow, \Mapsto\rangle$ *where* $\mathcal{A}$ *is the set of* all *arguments constructed from* $T$*. Thus,* $\mathcal{A}$ *is subargument-complete and rule-complete. Therefore,* $G$ *is subargument-complete and rule-complete.*

As we will see, one may be interested in building a set of subargument-complete and rule-complete subgraphs by considering 'subtheories' of a defeasible theory.

**Definition 2.20 (Defeasible subtheory)** *Let* $T = \langle Rules, conflict, \succ\rangle$ *denote a defeasible theory. Given a set* $Rules' \subseteq Rules$*, the* defeasible subtheory $U$ *of* $T$ *induced by* $Rules'$ *is the defeasible theory* $U = \langle Rules', conflict, \succ \cap (Rules' \times Rules')\rangle$*.*

We will use such subtheories later in our probabilistic investigations.

**Notation 2.6** *The set of all the subtheories of a theory* $T = \langle Rules, conflict, \succ\rangle$ *is denoted* $Sub(T)$*, i.e.* $Sub(T) = \{\langle Rules', conflict, \succ \cap (Rules' \times Rules')\rangle \mid Rules' \subseteq Rules\}$.

To recap, starting from a defeasible theory, we can build arguments and identify the attack and subargument relationships over arguments, thus creating an argumentation graph including its subgraphs. The acceptance and justification status of the arguments in a graph needs then to be evaluated. The next section discusses how these statuses can be determined by using labellings.

## 2.2 Labelling of arguments

Given an argumentation graph, the acceptance of arguments is evaluated on the basis of a formal specification traditionally called *argumentation semantics*. This evaluation can be carried out in terms of sets of arguments, called *extensions*, as originally proposed in [12], or in terms of labellings [2]. The idea underlying the extension-based approach is to identify sets of arguments, called extensions, which can collectively survive the conflict and therefore represent a reasonable position. In general, several alternative extensions exist. The idea underlying the labelling-based approach is to describe each reasonable position by a labelling, namely by assigning to each argument a label taken from a given set. Traditional extension-based semantics have an equivalent labelling-based characterization using the well-known set of three labels $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$ which we also adopt here. Briefly, for each extension $E$ the corresponding labelling is defined as follows: each argument belonging to $E$ is labelled IN, each argument attacked by the extension is labelled OUT, every other argument is labelled UN. Other approaches differing in the set of adopted labels and/or in the underlying semantics have been explored in the literature (see e.g. [45, 65, 31, 70]). Our proposal is not bounded to the choice of a specific set of labels or semantics and can be extended to other approaches too, a full development in this direction is left to future work.

As a first step, we introduce a pair of labels devoted to capture a situation of uncertainty about the presence of an argument within a framework. Traditionally the set of three labels $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$ is used to characterise the acceptance status of arguments on the basis of a given argumentation semantics. However, in our context of uncertainty, a labelling needs also to be able to express the fact that an argument is actually included and has some effect (let say it is 'ON') in the framework or that it is actually not included and hence has no effect (let say it is 'OFF'). Then, as we will see, 'classic' acceptance statuses 'IN', 'OUT' and 'UN' of arguments and the ON-OFF view can be seamlessly combined. This combination will allow us to clearly distinguish the probability of 'construction' or 'inclusion' of an argument and the probability of its acceptance.

Let us first introduce the basic formal notions and the relevant notation.

**Definition 2.21 (Labelling of a set of arguments)** *Let $G$ be an argumentation graph, and* $\mathsf{ArgLab}$ *a set of labels for arguments. An* $\mathsf{ArgLab}$-labelling *of a set of arguments $\mathcal{A} \subseteq \mathcal{A}_G$ is a total function* $\mathrm{L} : \mathcal{A} \to \mathsf{ArgLab}$.

**Notation 2.7** *Given an argumentation graph $G$, the universe of all possible* $\mathsf{ArgLab}$*-labelling assignments of a set $\mathcal{A} \subseteq \mathcal{A}_G$ is denoted $\mathcal{L}_{\mathsf{ArgLab}}(\mathcal{A})$.*

In general a labelling involves an arbitrary set of arguments. When this set is a singleton, with a little abuse of language we also speak of labelling of an argument. Labellings involving all the arguments of an argumentation graph play a special role and deserve a specific terminology and notation.

**Definition 2.22 (Labelling of an argumentation graph)** *Let $G$ be an argumentation graph, and* $\mathsf{ArgLab}$ *a set of labels for arguments. An* $\mathsf{ArgLab}$-labelling *of $G$ is a total function* $\mathrm{L} : \mathcal{A}_G \to \mathsf{ArgLab}$.

**Notation 2.8** *Given an argumentation graph $G$, the universe of all possible* $\mathsf{ArgLab}$*-labellings of $G$ is denoted $\mathcal{L}_{\mathsf{ArgLab}}(G)$.*

For instance, a $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$*-labelling* of $G$ is a total function $\mathrm{L} : \mathcal{A}_G \to \{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$.

**Notation 2.9** *The set of arguments labelled with a label $l$ by a labelling $\mathrm{L}$ is denoted $l(\mathrm{L})$, i.e. $l(\mathrm{L}) = \{A|\mathrm{L}(A) = l\}$. For instance, if $\mathsf{IN}$ is a label, then $\mathsf{IN}(\mathrm{L}) = \{A|\mathrm{L}(A) = \mathsf{IN}\}$.*

Not all labellings in $\mathcal{L}_{\mathsf{ArgLab}}(G)$ are meaningful or have satisfactory properties. Moreover, even among meaningful labellings, one may identify those satisfying some specific requirements. Hence typically, one needs to define some criteria to identify those labellings in $\mathcal{L}_{\mathsf{ArgLab}}(G)$ which are meaningful or interesting in some sense. Formally speaking, a specification identifies a subset of $\mathcal{L}_{\mathsf{ArgLab}}(G)$ for every possible argumentation graph $G$.

**Definition 2.23 (Labelling specification)** *Given a set of labels* $\mathsf{ArgLab}$*, an* $X$-$\mathsf{ArgLab}$-labelling specification *identifies for every argumentation graph $G$ a set of* $\mathsf{ArgLab}$*-labellings of $G$ denoted as $\mathcal{L}^X_{\mathsf{ArgLab}}(G) \subseteq \mathcal{L}_{\mathsf{ArgLab}}(G)$, where $X$ is called the criterion of the specification.*

If a labelling $\mathrm{L}$ belongs to the set of some specified labellings $\mathcal{L}^X_{\mathsf{ArgLab}}(G)$, i.e. $\mathrm{L} \in \mathcal{L}^X_{\mathsf{ArgLab}}(G) \subseteq \mathcal{L}_{\mathsf{ArgLab}}(G)$, we say that the labelling $\mathrm{L}$ is an $X$-$\mathsf{ArgLab}$-labelling of $G$, or an $X$ $\mathsf{ArgLab}$-labelling (without the hyphen).

**Definition 2.24 (Mono- and multi-labelling specification)** *An $X$-*$\mathsf{ArgLab}$*-labelling specification is a* mono-labelling specification *if, and only if, $|\mathcal{L}^X_{\mathsf{ArgLab}}(G)| = 1$ for any argumentation graph $G$, and it is a* multi-labelling specification *if, and only if, $|\mathcal{L}^X_{\mathsf{ArgLab}}(G)| > 1$ for some $G$.*

As an anticipation of our probabilistic setting, we employ three types of labellings: $\{\mathsf{ON}, \mathsf{OFF}\}$-labellings, $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$-labellings and $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}$-labellings. In a $\{\mathsf{ON}, \mathsf{OFF}\}$-labelling, each argument is associated with one label which is either $\mathsf{ON}$ or $\mathsf{OFF}$ to indicate whether an argument occurs, that is, whether an argument is regarded as playing an active role or not in the graph. Intuitively, arguments labeled $\mathsf{OFF}$ can be ignored. In a $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$-labelling, each argument is associated with one label representing its status in the context of a semantics-based evaluation [2]. Intuitively, a label '$\mathsf{IN}$' means the argument is accepted, while a label '$\mathsf{OUT}$' indicates that it is rejected and '$\mathsf{UN}$' that it is undecided, i.e. neither accepted nor rejected. A $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}$-labelling extends a $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$-labelling with the $\mathsf{OFF}$ label to indicate that an argument does not occur. Dually, a $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}$-labelling can be regarded as a refinement of a $\{\mathsf{ON}, \mathsf{OFF}\}$-labelling where occurring (i.e. $\mathsf{ON}$) arguments are the subject of semantics-based evaluation.

Having introduced basic intuitions on some types of labellings, we need now to specify how actual labellings of the various types can be built and which constraints they should satisfy. The simplest specification for a given set of labels is the one without constraints, denoted as the all-$\mathsf{ArgLab}$-labelling specification.

**Definition 2.25 (all-$\mathsf{ArgLab}$-labelling specification)** *The* all-$\mathsf{ArgLab}$-labelling *of an argumentation graph $G$ is such that $\mathcal{L}^{\mathsf{all}}_{\mathsf{ArgLab}}(G) = \mathcal{L}_{\mathsf{ArgLab}}(G)$.*

Concerning $\{\text{ON}, \text{OFF}\}$-labellings, we will focus on specifications ensuring that the set of arguments labelled ON features the completeness properties previously discussed.

**Definition 2.26 (X-$\{\text{ON}, \text{OFF}\}$-labelling specification)** *Let $G$ denote an argumentation graph. A* $\{\text{ON}, \text{OFF}\}$-labelling $\text{L}$ *of $G$ is*

- subargument-complete *if, and only if,* $\text{ON}(\text{L})$ *is subargument-complete;*
- rule-complete *if, and only if,* $\text{ON}(\text{L})$ *is rule-complete;*
- legal *if, and only if,* $\text{ON}(\text{L})$ *is subargument-complete and rule-complete.*

**Example 2.2 (continuing from p. 9)** *A legal $\{\text{ON}, \text{OFF}\}$-labelling and a subargument-complete $\{\text{ON}, \text{OFF}\}$-labelling (which is not legal because it is not rule-complete) are pictured in Figure 3.*

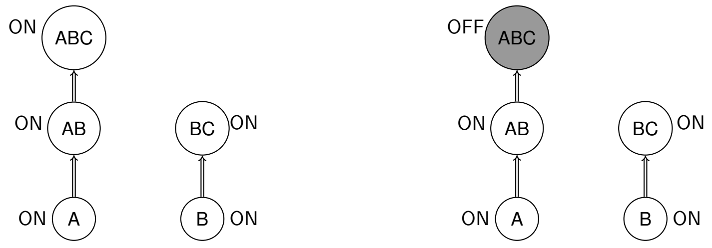


Figure 3: A legal $\{\text{ON}, \text{OFF}\}$-labelling (a), and a subargument-complete $\{\text{ON}, \text{OFF}\}$-labelling which is not legal (b) .

□

We can note that $\{\text{ON}, \text{OFF}\}$-labellings can be mapped to subgraphs: intuitively they correspond to 'switching off' arguments outside a given subgraph.

**Definition 2.27 ($\{\text{ON}, \text{OFF}\}$-labelling with respect to a subgraph)** *Let $G$ denote an argumentation graph, and $H$ a subgraph of $G$. The* $\{\text{ON}, \text{OFF}\}$-labelling of $G$ with respect to $H$, *denoted as $\text{L}_{G,H}$, is such that for every argument $A$ in $\mathcal{A}_G$, it holds that:*

- *$A$ is labelled ON if, and only if, $A$ is in $\mathcal{A}_H$, and*
- *$A$ is labelled OFF otherwise.*

**Proposition 2.3** *Let $G$ denote an argumentation graph constructed from a defeasible theory $T$, and $H$ any subgraph constructed from a subtheory $U$ of $T$, i.e. $U \in Sub(T)$. The $\{\text{ON}, \text{OFF}\}$-labelling of $G$ with respect to $H$ is* legal.

**Proof 2.3** *If an argumentation (sub)graph $H$ is constructed from a defeasible (sub)theory then $H$ is subargument-complete and rule-complete (Proposition 2.2). Thus, $\mathcal{A}_H$ is subargument and rule-complete (Definition 2.18 and 2.19). Consequently, every $\{\text{ON}, \text{OFF}\}$-labelling $\text{L}$ of $G$ with respect to $H$ is such that $\text{ON}(\text{L})$ is subargument and rule-complete. Therefore, every $\{\text{ON}, \text{OFF}\}$-labelling $\text{L}$ of $G$ with respect to $H$ is legal (Definition 2.26) .*

Turning to $\{\text{IN}, \text{OUT}, \text{UN}\}$-labellings, we follow the well-known labelling-based approach to argumentation semantics reviewed in [2]. In a nutshell, given an argumentation graph $G$ and a set of labels, an argumentation semantics associates with the graph $G$ a subset of the set of all possible $\{\text{IN}, \text{OUT}, \text{UN}\}$-labellings of the graph $G$. Note that, in our generalised setting, argumentation semantics can be introduced as an instance of labelling specification (Definition 2.23).

**Definition 2.28 (Argumentation semantics)** *An* argumentation semantics *is an $X$-$\{\text{IN}, \text{OUT}, \text{UN}\}$-labelling specification.*

Every argumentation semantics $X$ in the literature satisfies the property of being conflict-free (abbreviated cf), that is, for every argumentation graph $G$ $\mathcal{L}^{X}_{\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}}(G) \subseteq \mathcal{L}^{\mathsf{cf}}_{\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}}(G)$ as defined below.

**Definition 2.29 (Conflict-free $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling specification)** *A* conflict-free $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling *(or* cf-$\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$*-labelling) of an argumentation graph $G$ is a* $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$*-labelling such that for every argument $A$ in $\mathcal{A}_G$ it holds that $A$ is labelled* IN *if, and only if, all attackers of $A$ are not labelled* IN.

Several literature semantics (and in particular all the semantics introduced in [12]) are based on *complete* semantics: its definition as a specification for complete labellings is recalled below.

**Definition 2.30 (Complete $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling specification)** *A* complete $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling *of an argumentation graph $G$ is a $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling such that for every argument $A$ in $\mathcal{A}_G$ it holds that:*

- *$A$ is labelled* IN *if, and only if, all attackers of $A$ are* OUT*, and*
- *$A$ is labelled* OUT *if, and only if, $A$ has an attacker* IN.

Since a complete $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling is total, if an argument is neither labelled IN nor OUT, then this argument is labelled UN.

In general an argumentation graph may have several complete $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labellings: complete-based semantics, like the grounded, preferred and stable semantics, recalled below, provide additional specifications among the complete labellings.

**Definition 2.31 (Grounded $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling specification)**

*A* grounded $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling L *of an argumentation graph $G$ is a complete $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling of $G$ such that* $\mathsf{IN}(\mathrm{L})$ *is minimal (w.r.t. set inclusion) among all complete $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labellings of $G$.*

**Definition 2.32 (Preferred $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling specification)**

*A* preferred $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling L *of an argumentation graph $G$ is a complete $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling of $G$ such that* $\mathsf{IN}(\mathrm{L})$ *is maximal (w.r.t. set inclusion) among all complete $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labellings of $G$.*

**Definition 2.33 (Stable $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling specification)**

*A* stable $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling L *of an argumentation graph $G$ is a complete $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling of $G$ such that* $\mathsf{UN}(\mathrm{L})$ *is empty.*

**Example 2.1 (continuing from p. 5)** *A grounded $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling which is also a preferred and stable $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling is pictured in Figure 4.*

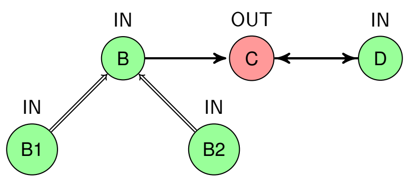


Figure 4: A grounded $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling which is also a preferred and stable $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling.

□

The traditional literature distinction between *single status* and *multiple status* argumentation semantics corresponds to our notions of *mono-labelling* and *multi-labelling* respectively (see Definition 2.24). For example, the grounded $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling semantics is mono-labelling, while the preferred $\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}$-labelling semantics is multi-labelling.

A few other semantics may be considered, see [2], but this is beyond the scope of the present paper.

We are now ready to introduce the novel $\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$-labellings as a way to combine a legal $\{\text{ON}, \text{OFF}\}$-labelling with $\{\text{IN}, \text{OUT}, \text{UN}\}$-labellings of the subgraph induced by the set of arguments labelled ON. Intuitively, this corresponds to combine together legality and semantics criteria.

**Definition 2.34 ($X$-$\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$-labelling specification)** *Let $G$ denote an argumentation graph, and $H$ a subargument-complete subgraph of $G$. A $\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$-labelling of $G$ is an $X$-$\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$-*labelling *with respect to $H$ if, and only if:*

- *every argument in $\mathcal{A}_H$ is labelled according to an $X$-$\{\text{IN}, \text{OUT}, \text{UN}\}$-labelling of $H$, and*
- *every argument in $\mathcal{A}_G \backslash \mathcal{A}_H$ is labelled* OFF.

*As for terminology, the $X$-$\{\text{IN}, \text{OUT}, \text{UN}\}$-labelling is called the sublabelling of the $X$-$\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$-labelling specification.*

The above definitions can be instantiated. For example, Definition 2.34 can be used to define the grounded $\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$-labelling of an argumentation graph.

**Definition 2.35 (Grounded $\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$-labelling specification)** *Let $H$ be a subargument-complete subgraph of an argumentation graph $G$. A* grounded $\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$-labelling *of $G$ with respect to $H$ is a $\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$-labelling such that:*

- *every argument in $\mathcal{A}_H$ is labelled according to the grounded $\{\text{IN}, \text{OUT}, \text{UN}\}$-labelling of $H$, and*
- *every argument in $\mathcal{A}_G \backslash \mathcal{A}_H$ is labelled* OFF.

**Example 2.1 (continuing from p. 5)** *A grounded $\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$-labelling and a preferred (and stable) $\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$-labelling are pictured in Figure 5.*

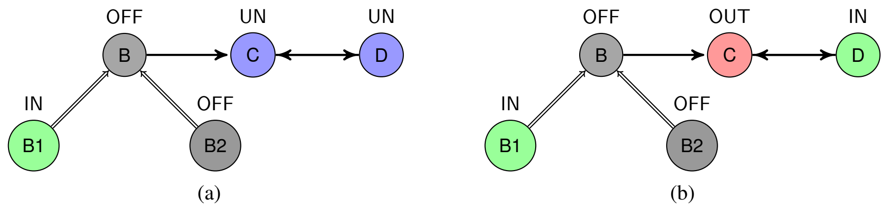


Figure 5: A grounded $\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$-labelling (a), and a preferred $\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$-labelling (b).

□

An argumentation graph $G$ has a unique grounded $\{\text{IN}, \text{OUT}, \text{UN}\}$-labelling, but it has as many grounded $\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$-labellings as subgraphs of $G$. Therefore, the grounded-$\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$-labelling specification is in general multi-labelling. More generally, the $X$-$\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$-labelling specification is multi-labelling, even though its sublabelling semantics is mono-labelling.

Argument labellings produced by the semantics evaluation are called argument *acceptance* labellings. On the basis of argument acceptance labellings, it is possible to aggregate the acceptance statuses into a justification status for every considered argument, resulting into argument *justification* labellings.

In non-probabilistic settings, the justification status of an argument is usually determined with respect to all the labellings in the set of labellings $\mathcal{L}^X_{\text{ArgLab}}(G)$ identified by an $X$-ArgLab-labelling specification, by using the traditional notions of skeptical and credulous justification. In particular, an argument is *skeptically justified* if it is labelled IN by all labellings in $\mathcal{L}^X_{\text{ArgLab}}(G)$, while it is *credulously justified* if it is labelled IN by at least one labelling in $\mathcal{L}^X_{\text{ArgLab}}(G)$, and it is *not justified* otherwise.

This simple notion of argument justification can be generalised considering a generic set of justification states, represented by a set of labels JArglabels, and a function producing a JArglabels-labelling of $G$ on the basis of the labellings $\mathcal{L}^{X}_{\mathsf{ArgLab}}(G)$. The traditional approach can be regarded as an instance of this general scheme where $\mathsf{JArglabels} = \{\text{SKJ}, \text{CRJ}, \text{NOJ}\}$ and the function is as specified above. The reader may refer to [5] for a wider formal treatment of this generalised labelling approach. We focus here only on the aspects which are necessary for the development of the present paper. In particular, the extension of the set of argument labels to $\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$ calls for an extended set of argument justification labels too. As a minimum, since an argument can be always ignored (labelled OFF by all $\mathcal{L}^{X}_{\mathsf{ArgLab}}(G)$) an additional justification label is needed to cover this case. On this basis, we introduce the 'semi-skeptical' $\{\text{OFJ}, \text{SKJ}, \text{CRJ}, \text{NOJ}\}$-labelling.

**Definition 2.36 (Semi-skeptical$\{\text{OFJ}, \text{SKJ}, \text{CRJ}, \text{NOJ}\}$-labelling)** *Let $\mathcal{L}$ denote a non-empty set of $\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$-labellings of an argumentation graph $G$. The* semi-skeptical $\{\text{OFJ}, \text{SKJ}, \text{CRJ}, \text{NOJ}\}$-labelling $\mathrm{L}_{\mathsf{J}}$ of $G$ with respect to $\mathcal{L}$ *is a $\{\text{OFJ}, \text{SKJ}, \text{CRJ}, \text{NOJ}\}$-labelling such that for every argument $A$ in $\mathcal{A}_G$, it holds that:*

- *$A$ is labelled OFJ, i.e. $\mathrm{L}_{\mathsf{J}}(A) = \text{OFJ}$, if, and only if, $\forall \mathrm{L} \in \mathcal{L}$, $\mathrm{L}(A) = \text{OFF}$;*
- *$A$ is labelled SKJ, i.e. $\mathrm{L}_{\mathsf{J}}(A) = \text{SKJ}$, if, and only if, $\forall \mathrm{L} \in \mathcal{L}$, $\mathrm{L}(A) = \text{IN}$;*
- *$A$ is labelled CRJ, i.e. $\mathrm{L}_{\mathsf{J}}(A) = \text{CRJ}$ if, and only if, $\exists \mathrm{L} \in \mathcal{L} : \mathrm{L}(A) = \text{IN}$ and $\mathrm{L}_{\mathsf{J}}(A) \neq \text{SKJ}$;*
- *otherwise $A$ is labelled NOJ, i.e. $\mathrm{L}_{\mathsf{J}}(A) = \text{NOJ}$.*

The 'semi-skeptical' $\{\text{OFJ}, \text{SKJ}, \text{CRJ}, \text{NOJ}\}$-labelling is only one of the many options in the design space of argument justification labellings. Its application is exemplified below, while further investigations about alternative notions of argument justification in the extended context we are exploring are left to future work. In this respect, we remark that introducing uncertainty in argumentation requires revising the formalisation in all the phases of the process, some of which have received relatively limited attention until now.

**Example 2.1 (continuing from p. 5)** *Assume the set of $\{\text{IN}, \text{OUT}, \text{UN}, \text{OFF}\}$-labellings $\mathcal{L}$ as previously illustrated in Figure 5 (two labellings). The semi-skeptical $\{\text{OFJ}, \text{SKJ}, \text{CRJ}, \text{NOJ}\}$-labelling with respect to $\mathcal{L}$ is illustrated in Figure 6.*

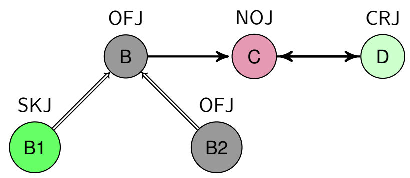


Figure 6: A semi-skeptical $\{\text{OFFJ}, \text{SKJ}, \text{CRJ}, \text{NOJ}\}$-labelling.

□

The semi-skeptical $\{\text{OFJ}, \text{SKJ}, \text{CRJ}, \text{NOJ}\}$-labelling is obviously total because the label NOJ covers all cases not covered by the other labels, but we can nevertheless define NOJ in more specific terms, as follows.

**Proposition 2.4** *Let $\mathcal{L}$ denote a non-empty set of observed argument labellings. It holds that $\mathrm{L}_{\mathsf{J}}(A) = \text{NOJ}$ if, and only if, $\exists \mathrm{L} \in \mathcal{L}, \mathrm{L}(A) = \text{OUT}$ or $\mathrm{L}(A) = \text{UN}$ and $\forall \mathrm{L} \in \mathcal{L}, \mathrm{L}(A) \neq \text{IN}$.*

To recap, given an argumentation graph, the acceptance and justification statuses of arguments can be expressed according to a variety of labelling specifications. To encompass uncertainty about the inclusion/omission of arguments, we introduced two additional labels indicating inclusion ('ON') or omission ('OFF'). We showed how existing labelling specifications can be gently adapted in that regard, and we introduced an upgraded notion of argument justification. On this basis, statements can be labelled in turn, as we will see next.

## 2.3 Labelling of statements

The final phase of the process concerns the labelling of statements, and in particular the labelling of the conclusions supported by arguments. As a matter of fact, assessing conclusions is the ultimate goal of the argumentation process. For instance, in Example 2.1 the research scientist wants to form an opinion about whether the program is running or not. In the remainder of this section, we focus on the labelling of statements which are, in our case, literals.

Labellings of statements can be performed in different manners, see e.g. [5]. From an abstract point of view, given a set of statements − where a statement may not be the conclusion of any argument − a labelling of this set is a function associating every statement with a label.

**Definition 2.37 (Labelling of literals)** *Let* $\Phi$ *denote a set of literals, and* LitLabels *denote a set of labels on literals. A* LitLabels-labelling of $\Phi$ *is a function* $\mathrm{K} : \Phi \rightarrow$ LitLabels.

Per se, a labelling of literals is just a function mapping a set of literals to a set of labels, but such a labelling may rely on an acceptance labelling of arguments. For this reason, we introduce an acceptance labelling function as follows.

**Definition 2.38 (Acceptance labelling of literals)** *Let* $G$ *denote an argumentation graph,* $\mathcal{L}^X_{\mathsf{ArgLab}}(G)$ *the set of* $X$-ArgLab-*labellings of the argumentation graph* $G$*,* $\Phi$ *a set of literals, and* LitLabels *a set of labels on literals. An* acceptance LitLabels-labelling of $\Phi$ *is a function* $\mathrm{K} : \mathcal{L}^X_{\mathsf{ArgLab}}(G) \times \Phi \rightarrow$ LitLabels.

**Example 2.3** *If we write* $\mathrm{K}(\mathrm{L}, \varphi) = \mathsf{in}$*, then it means that, given the argument labelling* L*, the literal* $\varphi$ *is labelled* in. ☐

It is often desirable that the considered labelling function is total, and we assume total labelling functions in the remainder.

Different specifications for labellings of statements are possible. A simple labelling is the so-called *bivalent labelling*, according to which a statement is either justified or not, without further sophistication. If a statement is accepted then it is labelled in, otherwise it is labelled no. In this case, the necessary and sufficient condition for a statement to be labelled in is to have at least one argument labelled IN supporting this statement.

**Definition 2.39 (Bivalent labelling of literals)** *Let* $G$ *denote an argumentation graph,* $\mathcal{L}^X_{\mathsf{ArgLab}}(G)$ *denote the set of* $X$-ArgLab-*labellings of the argumentation graph* $G$*, and* $\Phi$ *denote a set of literals. A* bivalent $\{\mathsf{in}, \mathsf{no}\}$-labelling of $\Phi$ *is a total function* $\mathrm{K} : \mathcal{L}^X_{\mathsf{ArgLab}}(G), \Phi \rightarrow \{\mathsf{in}, \mathsf{no}\}$*, such that* $\forall \varphi \in \Phi$*:*

- $\varphi$ *is labelled* in*, i.e.* $\mathrm{K}(\mathrm{L}, \varphi) = \mathsf{in}$*, if and only if* $\exists A \in \mathsf{IN}(\mathrm{L}) : conc(A) = \varphi$*, i.e. there is at least one argument* $A$ *such that*
  - *this argument* $A$ *is labelled* IN *in* L*, and*
  - *the literal* $\varphi$ *is the conclusion of argument* $A$*;*
- $\varphi$ *is labelled* no *otherwise, i.e* $\mathrm{K}(\mathcal{L}, \varphi) = \mathsf{no}$.

**Example 2.1 (continuing from p. 5)** *Consider the grounded* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}$*-labelling* L *as given in Figure 5, the bivalent labelling of the set of literals supported by the arguments is such that* $\mathrm{K}(\mathrm{L}, \neg\mathsf{b1}) = \mathsf{in}$*,* $\mathrm{K}(\mathrm{L}, \neg\mathsf{b2}) = \mathsf{no}$*,* $\mathrm{K}(\mathrm{L}, \neg\mathsf{b}) = \mathsf{no}$*,* $\mathrm{K}(\mathrm{L}, \mathsf{c}) = \mathsf{no}$*, and* $\mathrm{K}(\mathrm{L}, \neg\mathsf{c}) = \mathsf{no}$. ☐

The bivalent labelling is simple, but it takes no advantage of the fine-grained labelling of arguments. For this reason, we may consider more sophisticated labellings concerning the statuses of statements supported by no arguments labelled IN. So, following the same spirit underlying $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}$-labelling for arguments, we propose the novel *worst-case sensitive labellings* where the labels out, un, off and unp are considered together with in.

**Definition 2.40 (Worst-case sensitive labelling of literals)** *Let* $G$ *denote an argumentation graph,* $\mathcal{L}^X_{\mathsf{ArgLab}}(G)$ *the set of* $X$-ArgLab-*labellings of the argumentation graph* $G$*, and* $\Phi$ *a set of literals. A* worst-case sensitive $\{\mathsf{in}, \mathsf{out}, \mathsf{un}, \mathsf{off}, \mathsf{unp}\}$-labelling of $\Phi$ *is a total function* $\mathrm{K} : \mathcal{L}^X_{\mathsf{ArgLab}}(G), \Phi \rightarrow \{\mathsf{in}, \mathsf{out}, \mathsf{un}, \mathsf{off}, \mathsf{unp}\}$*, such that* $\forall \varphi \in \Phi$*:*

- *$\varphi$ is labelled* in*, i.e.* $\mathrm{K}(\mathrm{L}, \varphi) = \mathsf{in}$*, if, and only if,* $\exists A \in \mathsf{IN}(\mathrm{L}) : conc(A) = \varphi$*, i.e. there is at least one argument $A$ such that*
  - *this argument $A$ is labelled* IN *in* L*, and*
  - *the literal $\varphi$ is the conclusion of argument $A$.*
- *$\varphi$ is labelled* out*, i.e.* $\mathrm{K}(\mathrm{L}, \varphi) = \mathsf{out}$*, if, and only if,* $\exists A \in \mathsf{OUT}(\mathrm{L}) : conc(A) = \varphi$ *and* $\forall A : conc(A) = \varphi$*,* $A \in \mathsf{OUT}(\mathrm{L}) \cup \mathsf{OFF}(\mathrm{L})$*, i.e.*
  - *there is at least one argument whose conclusion is $\varphi$, which is not labelled* OFF *in* L*, and*
  - *all arguments whose conclusion is $\varphi$ are labelled* OUT *or* OFF *in* L.
- *$\varphi$ is labelled* un*, i.e.* $\mathrm{K}(\mathrm{L}, \varphi) = \mathsf{un}$*, if, and only if,* $\nexists A \in \mathsf{IN}(\mathrm{L}) : conc(A) = \varphi$ *and* $\exists A \in \mathsf{UN}(\mathrm{L}) : conc(A) = \varphi$*, i.e.*
  - *there is no argument labelled* IN *in* L *whose conclusion is $\varphi$, and*
  - *there is at least one argument labelled* UN *in* L *whose conclusion is $\varphi$.*
- *$\varphi$ is labelled* off*, i.e.* $\mathrm{K}(\mathrm{L}, \varphi) = \mathsf{off}$*, if, and only if,* $\exists A : conc(A) = \varphi$ *and* $\forall A : conc(A) = \varphi$ $A \in \mathsf{OFF}(\mathrm{L})$*, i.e.*
  - *there is at least one argument in* L *whose conclusion is $\varphi$, and*
  - *all arguments whose conclusion is $\varphi$ are labelled* OFF *in* L.
- *$\varphi$ is labelled* unp*, i.e.* $\mathrm{K}(\mathrm{L}, \varphi) = \mathsf{unp}$*, if, and only if,* $\nexists A : conc(A) = \varphi$*, i.e. there is no argument labelled in* L *whose conclusion is $\varphi$.*

The labels out, un, off and unp can be seen as a finer classification of the cases corresponding to the no label: with the label out all arguments concluding $\varphi$ are either switched off or plainly rejected, with un at least one of them is undecided, with off they are all switched off. Finally, the label unp caters for queries about a statement for which there is no argument (we use the label unp as an abbreviation for 'unprovable'). The name of this labelling reflects a distance from acceptance, so out is worse than un which is worse than off, which is worse than unp.

**Example 2.1 (continuing from p. 5)** *Consider the grounded* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}$*-labelling* L *as given in Figure 5, the worst-case sensitive labelling of the set of literals supported by the arguments is such that* $\mathrm{K}(\mathrm{L}, \neg \mathsf{b1}) = \mathsf{in}$, $\mathrm{K}(\mathrm{L}, \neg \mathsf{b2}) = \mathsf{off}$, $\mathrm{K}(\mathrm{L}, \neg \mathsf{b}) = \mathsf{off}$, $\mathrm{K}(\mathrm{L}, \mathsf{c}) = \mathsf{un}$*, and* $\mathrm{K}(\mathrm{L}, \neg \mathsf{c}) = \mathsf{un}$. □

Other labels and types of labellings for statements can be considered, in particular a notion of statement justification can also be investigated, but the labellings introduced here are sufficient to illustrate our probabilistic framework, as we will see in the next section.

To summarise the section, we have considered a minimalist but articulated formal framework covering the various phases of a rule-based argumentation process. In this framework, arguments are constructed from defeasible theories by chaining defeasible rules. Some arguments can attack or be subarguments of other arguments, and these relations amongst arguments are captured into abstract argumentation graphs. Then, the acceptance status of arguments can be expressed in terms of labellings specified by an argumentation semantics. The acceptance labellings are in turn the basis for a labelling-based definition of argument justification and statement acceptance. In addition to existing labelling specifications, we considered two novel labels indicating the inclusion ('ON') or omission ('OFF') of arguments. We showed how existing labelling specifications can be gently adapted in that regard. As we will see next, this framework will allow us to investigate and compare different probabilistic argumentation settings, leading us to a labelling-oriented approach called probabilistic labellings.

# 3 Probabilistic Labelling of Arguments and Statements

Building on the framework introduced in Section 2, we are now ready to present, as the main contribution of the paper, a systematic approach to encompass probabilistic uncertainty in this context. In particular, we consider several possible answers to the question 'which are the elements we are uncertain about?', which, adopting Kolmogorov axioms, amounts to identify various alternatives for the definition of the probability space. To this purpose, we will first introduce probabilistic argumentation frames (Subsection 3.1), and then we will consider their use for probabilistic labellings of arguments (Subsection 3.2) and statements (Subsection 3.3).

## 3.1 Probabilistic argumentation frames

Different probability spaces can provide a basis for probabilistic argumentation, each featuring different types of uncertainty. In that regard, we can distinguish probability spaces featuring some uncertainty on the argumentative structures, namely defeasible theories or argumentation graphs, and probability spaces focusing on the uncertainty on the acceptance of arguments. We investigate these alternatives in the remainder.

### 3.1.1 Structural uncertainty: probabilistic theory and graph frames

Since arguments are built on the basis of a defeasible theory, one may consider uncertainty at the level of the theory itself, that is, given a reference set of defeasible rules, uncertainty about which subset of rules to adopt. In this view, the sample space corresponds to the set of subtheories of a defeasible theory, leading to the definition of probabilistic theory frames (PTFs).

**Definition 3.1 (Probabilistic theory frame)** *A* probabilistic theory frame *(PTF) based on a defeasible theory $T$ is a tuple $\langle T, \langle \Omega_{\mathrm{PTF}}, F_{\mathrm{PTF}}, P_{\mathrm{PTF}} \rangle\rangle$ where $\langle \Omega_{\mathrm{PTF}}, F_{\mathrm{PTF}}, P_{\mathrm{PTF}} \rangle$ is a probability space such that:*

- *the sample space $\Omega_{\mathrm{PTF}}$ is the set of subtheories of $T$, i.e. $\Omega_{\mathrm{PTF}} = Sub(T)$;*
- *the $\sigma$-algebra[2] $F_{\mathrm{PTF}}$ is the power set of $\Omega_{\mathrm{PTF}}$, i.e. $F_{\mathrm{PTF}} = 2^{\Omega_{\mathrm{PTF}}}$;*
- *the function $P_{\mathrm{PTF}}$ from $F_{\mathrm{PTF}}$ to $[0,1]$ is a probability distribution satisfying Kolmogorov axioms.*

**Notation 3.1** *The (infinite) set of the PTFs based on a defeasible theory $T$ (differing on the probability function $P$) is denoted $PTF(T)$.*

Every subtheory $U$ of a theory $T$ generates exactly one argumentation graph $G_U$, thus uncertainty about the set of rules also corresponds to being uncertain about the induced argumentation graph among the subgraphs of $G_T$. For this reason, one may decide to adopt a representation at a more abstract level, where, given a reference argumentation graph, there is uncertainty about which of its subgraphs should be actually taken into account, leading to the definition of probabilistic graph frames (PGFs).

**Definition 3.2 (Probabilistic graph frame)** *A* probabilistic graph frame *(PGF) based on an argumentation graph $G$ is a tuple $\langle G, \langle \Omega_{\mathrm{PGF}}, F_{\mathrm{PGF}}, P_{\mathrm{PGF}} \rangle\rangle$ where $\langle \Omega_{\mathrm{PGF}}, F_{\mathrm{PGF}}, P_{\mathrm{PGF}} \rangle$ is a probability space such that:*

- *the sample space $\Omega_{\mathrm{PGF}}$ is the set of subgraphs of $G$, i.e. $\Omega_{\mathrm{PGF}} = Sub(G)$;*
- *the $\sigma$-algebra $F_{\mathrm{PGF}}$ is the power set of $\Omega_{\mathrm{PGF}}$, i.e. $F = 2^{\Omega_{\mathrm{PGF}}}$;*
- *the function $P_{\mathrm{PGF}}$ from $F_{\mathrm{PGF}}$ to $[0,1]$ is a probability distribution satisfying Kolmogorov axioms.*

**Notation 3.2** *The (infinite) set of the PGFs based on an argumentation graph $G$ (differing on the probability function $P$) is denoted $PGF(G)$.*

The relation from the subtheories of a theory $T$ to the subgraphs of $G_T$ is neither injective nor surjective. Non injectivity is due to the fact that it may be the case $G_U = G_{U'}$ with $U \neq U'$ i.e. different subtheories may generate the same graph. This is due to the fact that subtheories may differ due to the presence of rules which are actually not used in the construction of any argument. Non suriectivity

[2] An algebra of sets is a collection $\mathcal{C}$ of subsets of a given set $S$ such that (i) $S \in \mathcal{C}$, (ii) if $X \in \mathcal{C}$ and $Y \in \mathcal{C}$ then $X \cup Y \in \mathcal{C}$, (iii) if $X \in \mathcal{C}$ then $S \backslash X \in \mathcal{C}$. The collection $\mathcal{C}$ is also closed under intersections. A $\sigma$-algebra is additionally closed under countable unions (and intersections): (iv) If $X_n \in \mathcal{C}$ for all $n$, then $\bigcup_{n=0}^{\infty} X_n \in \mathcal{C}$ [32]. Note that, while we specify the $\sigma$-algebra as a power set, it may be specified differently. We adopt here the most common specification, which appears to us as the most 'natural' for our purposes.

is due to the fact that for every subtheory $U$, $G_U$ is subargument-complete and rule-complete, while $Sub(G_T)$ typically includes also subgraphs not satisfying these properties. This has an effect on the correspondence from PTFs to PGFs defined as follows.

**Definition 3.3 (PGF corresponding to a PTF)** *Given a PTF $\langle T, \langle \Omega_{PTF}, F_{\mathrm{PTF}}, P_{\mathrm{PTF}} \rangle\rangle$, its corresponding PGF is a PGF tuple $\langle G_T, \langle \Omega_{\mathrm{PGF}}, F_{\mathrm{PGF}}, P_{\mathrm{PGF}} \rangle\rangle$ where the probability distribution $P_{\mathrm{PGF}}$ is such that $\forall S \in F_{\mathrm{PGF}}$:*[3]

$$P_{\mathrm{PGF}}(S) = \sum_{\{U \in Sub(T) | G_U \in S\}} P_{\mathrm{PTF}}(\{U\}).$$

**Notation 3.3** *The PGF corresponding to a PTF $\mathfrak{F}$ is denoted $PGF(\mathfrak{F})$.*

Applying Definition 3.3 to the case of singletons, we get that the distribution $P_{\mathrm{PGF}}$ is such that for every subgraph $H$ in $Sub(G)$,

$$P_{\mathrm{PGF}}(\{H\}) = \sum_{\{U \in Sub(T) | G_U = H\}} P_{\mathrm{PTF}}(\{U\}). \tag{1}$$

Given that for every subtheory $U \in Sub(T)$ the graph $G_U$ is subargument-complete and rule-complete, it follows also that for every subgraph $H$ in $Sub(G)$ which is not subargument-complete or rule-complete it holds that $P_{\mathrm{PGF}}(\{H\}) = 0$. Hence, the mapping between the sample space of a PTF and its corresponding PGF may be neither injective nor surjective.

**Example 2.2 (continuing from p. 9)** *Consider the PTF $\langle \mathsf{T}, \langle \Omega_{\mathrm{PTF}}, F_{\mathrm{PTF}}, P_{\mathrm{PTF}} \rangle\rangle$ and its corresponding PGF $\langle \mathsf{G_T}, \langle \Omega_{\mathrm{PGF}}, F_{\mathrm{PGF}}, P_{\mathrm{PGF}} \rangle\rangle$. Figure 7 shows the mapping from $\Omega_{\mathrm{PTF}}$ to $\Omega_{\mathrm{PGF}}$, where each graph is represented by the arguments belonging to it. The mapping is clearly neither injective nor surjective. An arbitrary probability distribution on the subtheories of $\mathsf{T}$ is given and the corresponding distribution on the subgraphs of $\mathsf{G_T}$ is shown.* □

Given a defeasible theory $T$, the set of PGFs based on the subgraphs of $G_T$ is a strict superset of the set of PGFs that can be generated starting from the PTFs based on $T$. Formally:

$$PGF(G_T) \supsetneq \{PGF(\mathfrak{F}) \mid \mathfrak{F} \in PTF(T)\}. \tag{2}$$

This is because, for every PTF $\mathfrak{F}$ in $PTF(T)$, $PGF(\mathfrak{F})$ is such that the subgraphs of $G_T$ which are not subargument-complete and rule-complete have probability zero, while $PGF(G_T)$ includes also PGFs not satisfying this constraint.

Hence, PGFs are stricty more expressive than PTFs as far uncertainty about the actual graph to be considered is concerned. When moving from a PTF to a corresponding PGF, some information is also lost due to non-injectivity: in general many different PTFs may generate the same PGF. However, this loss of information has no practical effects, since if two PTFs generate the same PGF, the difference between them has no influence on the subsequent argumentation steps.

PTFs and PGFs capture two related forms of structural uncertainty. PTFs refer to uncertainty about the structure of the theory to be used, while PGFs refer to the (consequent) uncertainty about which arguments are present in an argumentation graph. Now, a distinct but possibly coexisting kind of uncertainty not involving structural aspects can be considered too, as investigated next.

#### 3.1.2 Acceptance uncertainty: probabilistic labelling frames

Acceptance uncertainty refers to the fact that, given an argumentation graph, a semantics may prescribe multiple acceptance labellings for it, each representing individually a reasonable outcome of the conflict resolution process. The agent carrying out argument-based reasoning is then confronted with the choice of one among these outcomes. Indeed, choosing one of the reasonable options is mandatory to avoid standstill, at least in practical reasoning contexts, as exemplified by the famous Buridan's

[3]It is assumed that a sum over the empty set is 0, i.e. $P_{\mathrm{PGF}}(S) = 0$ if $\{U \in Sub(T) \mid G_U \in S\} = \emptyset$.

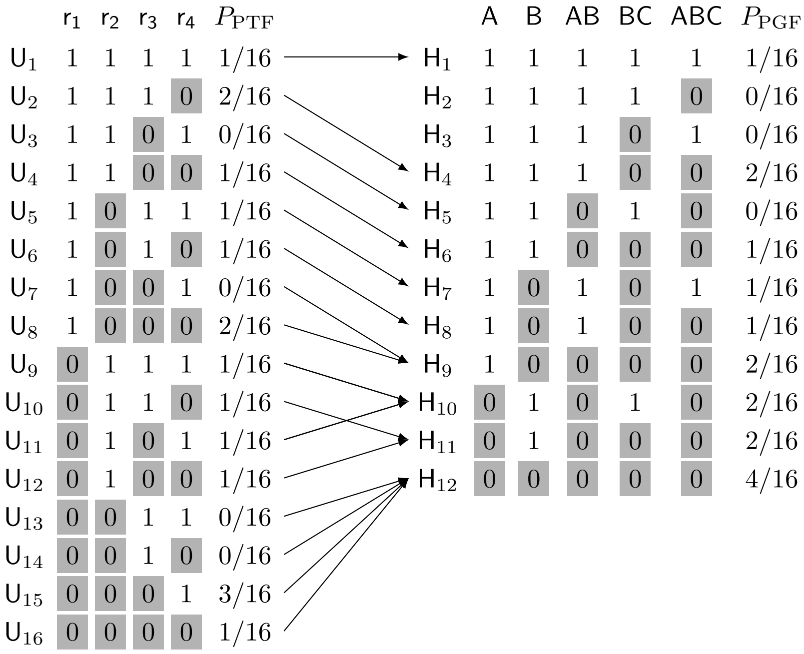


Figure 7: Non-injective non-surjective mapping from the sample space (of subtheories $\mathsf{U}_i$) of the considered PTF (on the left) to the sample space of its corresponding PGF (on the right).

donkey example. One may therefore want to express quantitative uncertainty about which option, i.e. which acceptance labelling, to choose. Clearly this kind of acceptance uncertainty is 'orthogonal' to structural uncertainty and is outside the scope of PTFs and PGFs.

To address acceptance uncertainty, we devise a sample space which is the set of possible labellings of an argumentation graph, as previously employed in [49, 50, 51], and resulting into what we call *probabilistic labelling frames* (PLFs) that we investigate in the remainder of this section.

PLFs are based on the assumption that, given an $X$-ArgLab-labelling specification for an argumentation graph $G$, an agent can assign a probability to every labelling of the set of labellings $\mathcal{L}^X_{\mathsf{ArgLab}}(G)$ which represents the sample space. Intuitively, this can be interpreted as follows: the $X$-ArgLab-labelling specification gives rise to a first, uncertainty free, selection of the set of labellings $\mathcal{L}^X_{\mathsf{ArgLab}}(G)$ that are worth considering within the universe $\mathcal{L}_{\mathsf{ArgLab}}(G)$. Then the labellings in $\mathcal{L}^X_{\mathsf{ArgLab}}(G)$ may not be regarded as equally probable and hence associated with different probability values. Note that, at this abstract level, we do not commit to any specific meaning of the probability of a labelling belonging to $\mathcal{L}^X_{\mathsf{ArgLab}}(G)$. For instance, in an epistemic reasoning scenario, different probability values may correspond to the fact that an agent, according to some domain-specific information, views the various labellings as differently credible, while in a social argumentation context, the probability values may correspond to the expectation that a specific labelling is approved by the members of a community. Our framework, being application agnostic, provides a uniform formal model to investigate general properties valid for a large family of probabilistic argumentation scenarios, each of which can be characterised by the adopted labelling specification and, possibly, by additional specific properties.

**Notation 3.4** *Whilst our notational nomenclature refers to $X$-ArgLab-labellings (e.g.* legal-$\{\mathsf{ON}, \mathsf{OFF}\}$-*labellings,* complete-$\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$-*labellings,* complete-$\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}$-*labellings and so on) where $X$ and* ArgLab *refers to an $X$-*ArgLab*-labelling specification, for the sake of notational conciseness we will sometimes speak of $\mathcal{S}$-labellings, using a single symbol $\mathcal{S}$ to synthesise the pair of symbols $X$-*ArgLab.

**Definition 3.4 (Probabilistic labelling frame)** *A* probabilistic labelling frame *(PLF) based on an argumentation graph $G$ is a tuple*[4] $\langle G, \mathcal{S}, \langle \Omega_{\mathrm{PLF}}, F_{\mathrm{PLF}}, P_{\mathrm{PLF}} \rangle\rangle$ *where $\mathcal{S}$ denotes an $X$-$\mathsf{ArgLab}$-labelling specification, and $\langle \Omega_{\mathrm{PLF}}, F_{\mathrm{PLF}}, P_{\mathrm{PLF}} \rangle$ is a probability space such that:*

- *the sample space $\Omega_{\mathrm{PLF}}$ is the set of $X$-$\mathsf{ArgLab}$-labellings of $G$, i.e. $\Omega_{\mathrm{PLF}} = \mathcal{L}^{X}_{\mathsf{ArgLab}}(G)$;*
- *the $\sigma$-algebra $F_{\mathrm{PLF}}$ is the power set of $\Omega_{\mathrm{PLF}}$, i.e. $F_{\mathrm{PLF}} = 2^{\Omega_{\mathrm{PLF}}}$;*
- *the function $P_{\mathrm{PLF}}$ from $F_{\mathrm{PLF}}$ to $[0, 1]$ is a probability distribution satisfying Kolmogorov axioms.*

Thanks to the versatility of labellings, PLFs feature enhanced expressiveness with respect to PGFs. In fact PGFs can be regarded as a special case of PLFs with $\mathsf{ArgLab} = \{\mathsf{ON}, \mathsf{OFF}\}$, as illustrated by the following definition.

**Definition 3.5 (PLF corresponding to a PGF)** *Given a PGF $\langle G, \langle \Omega_{\mathrm{PGF}}, F_{\mathrm{PGF}}, P_{\mathrm{PGF}} \rangle\rangle$, its corresponding PLF is a PLF tuple $\langle G, \mathsf{all}\text{-}\{\mathsf{ON}, \mathsf{OFF}\}, \langle \Omega_{\mathrm{PLF}}, F_{\mathrm{PLF}}, P_{\mathrm{PLF}} \rangle\rangle$ where the probability distribution $P_{\mathrm{PLF}}$ is such that $\forall S \in F_{\mathrm{PLF}}$:*

$$P_{\mathrm{PLF}}(S) = \sum_{\{H \in Sub(G) \,|\, \mathrm{L}_{G,H} \in S\}} P_{\mathrm{PGF}}(\{H\}).$$

**Notation 3.5** *The PLF corresponding to a PGF $\mathfrak{F}$ is denoted as $PLF(\mathfrak{F})$.*

Given a PGF based on an argumentation graph $G$, one can draw different mappings from the elements of the sample space $\Omega_{\mathrm{PGF}} = Sub(G)$ to the elements of the sample space of a PLF based on $G$, depending on the adopted labelling specification. As mentioned above, if one adopts the $\mathsf{all}$-$\{\mathsf{ON}, \mathsf{OFF}\}$ specification every element of $\Omega_{\mathrm{PGF}}$, i.e. every subgraph of $G$, corresponds exactly to an element of $\Omega_{\mathrm{PLF}}$, i.e. to a $\{\mathsf{ON}, \mathsf{OFF}\}$-labelling and vice versa. If instead one adopts an $X$-$\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}$ specification where $X$ is a multiple status argumentation semantics, in general for each subgraph $H$ of $G$ there are many corresponding elements in $\Omega_{\mathrm{PLF}}$, namely the set of $X$-$\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}$-labellings of $G$ corresponding to the set of $X$-$\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$-labellings of $H$. However, a reverse correspondence can be drawn as follows.

**Definition 3.6 (PGF corresponding to a PLF)**
*Given a PLF $\langle G, \mathcal{S}, \langle \Omega_{\mathrm{PLF}}, F_{\mathrm{PLF}}, P_{\mathrm{PLF}} \rangle\rangle$, its corresponding PGF is a PGF tuple $\langle G, \langle \Omega_{\mathrm{PGF}}, F_{\mathrm{PGF}}, P_{\mathrm{PGF}} \rangle\rangle$ where the probability distribution $P_{\mathrm{PGF}}$ is such that $\forall S \in F_{\mathrm{PGF}}$:*

$$P_{\mathrm{PGF}}(S) = \sum_{\{\mathrm{L} \in \Omega_{\mathrm{PLF}} \,|\, G(\mathrm{L}) \in S\}} P_{\mathrm{PLF}}(\{\mathrm{L}\})$$

*where $G(\mathrm{L})$ is the subgraph of $G$ induced by the set of arguments which are not labelled $\mathsf{OFF}$ in $\mathrm{L}$, namely, letting $\mathcal{X} = \mathcal{A} \setminus \mathsf{OFF}(\mathrm{L})$, $G(\mathrm{L}) = (\mathcal{X}, \leadsto_G \cap (\mathcal{X} \times \mathcal{X}), \Rrightarrow_G \cap (\mathcal{X} \times \mathcal{X}))$.*

The correspondence is illustrated in Example 3.1. If $X$ is single status the sets mentioned above are singletons and the mapping between $\Omega_{\mathrm{PGF}}$ and $\Omega_{\mathrm{PLF}}$ is bijective.

**Example 3.1 (Abstract example II)** *Let us consider a PGF $\langle \mathsf{G}, \langle \Omega_{\mathrm{PGF}}, F_{\mathrm{PGF}}, P_{\mathrm{PGF}} \rangle\rangle$ and a PLF $\langle \mathsf{G}, \mathsf{preferred}\text{-}\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}, \langle \Omega_{\mathrm{PLF}}, F_{\mathrm{PLF}}, P_{\mathrm{PLF}} \rangle\rangle$ with the argumentation graph $\mathsf{G} = \langle \{\mathsf{B}, \mathsf{C}\}, \{(\mathsf{B}, \mathsf{C}), (\mathsf{C}, \mathsf{B})\}\rangle$ as pictured in Figure 8 .*

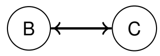


Figure 8: A simple argumentation graph.

*Figure 9 shows the mapping from the sample space $\Omega_{\mathrm{PLF}}$ to $\Omega_{\mathrm{PGF}}$, a given probability distribution on $\Omega_{\mathrm{PLF}}$, and the corresponding distribution on $\Omega_{\mathrm{PGF}}$.*

□

[4]Later we will often omit the subscript PLF for the sake of conciseness.

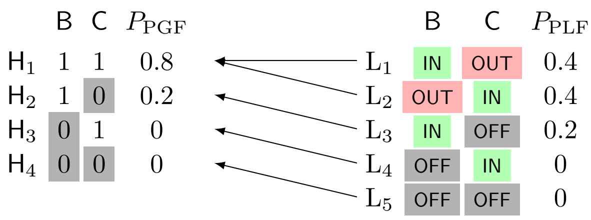


Figure 9: Mapping from the sample space of the considered PLF (on the right) to the sample space of the relevant PGF (on the left) for the argumentation graph of Figure 8.

Since PGFs are more expressive than PTFs, as far as uncertainty about the actual argumentation graph is concerned, this correspondence shows that also PLFs are strictly more expressive than PTFs in this respect. Further, one can also extend the non-injective non-surjective relationship from PTFs to PGFs (see Definition 3.3) to the corresponding PLFs.

While PLFs using $\{$ON, OFF$\}$-labellings are able to encompass the uncertainty about argumentation graphs expressed by PTFs and PGFs, PLFs using $\{$IN, OUT, UN$\}$-labellings can express uncertainty about the actual labelling to choose in a multiple status semantics with a fixed argumentation graph, thus covering the acceptance uncertainty dimension alone. More interestingly, PLFs using $\{$IN, OUT, UN, OFF$\}$-labellings are able to combine both kinds of uncertainty in a single representation. As discussed later, this allows to capture some existing literature approaches as special cases of PLFs, providing at the same time a comprehensive formalism to combine them.

To recap, we have laid down different kinds of probabilistic frames, namely PTFs, PGFs and (the novel concept of) PLFs, each kind featuring a different probability space. We showed that PLFs subsume PTFs and PGFs, and for this reason, we will focus on PLFs in the remainder. Before studying the relationships between PLFs with other works, we will see next that such PLFs are convenient to attach probabilistic measures to different statuses of arguments and to investigate their relationships.

## 3.2 Probabilistic labelling of arguments

The generalised framework introduced in the previous sections provides the basis for introducing further notions related to the reasoning tasks an agent can be interested in. For instance, an agent may focus attention on a few arguments which are most significant for his/her purposes, and be interested in the labels that can be assigned to these arguments and in the relevant probabilities both at the level of argument acceptance and of argument justification.

### 3.2.1 Probabilistic argument acceptance labellings

In order to define probabilistic argument acceptance labellings, we introduce suitable random variables on the basis of the generic probabilistic space introduced for PLFs.

Recall that a random variable is a function (traditionally denoted by an upper case letter as $X$, $Y$ or $Z$ for example) from $\Omega$ into another set $E$ of elements. An assignment of a random variable $X$ is denoted as $X = e$ where $e \in E$. An assignment for a set of random variables $\{X_1, \ldots, X_n\}$ is simply a set $\{X_1 = e_1, \ldots, X_n = e_n\}$ of assignments to all variables in the set.

For every argument $A$, we use a categorical random variable called a *random labelling* denoted $L_A$ from $\Omega$ into a set ArgLab of labels such as $\{$ON, OFF$\}$, or $\{$IN, OUT, UN$\}$ or $\{$IN, OUT, UN, OFF$\}$. So, for instance, if the sample space $\Omega$ is a set of $\{$ON, OFF$\}$-labellings, the event $L_A =$ ON is a shorthand for the outcomes $\{\mathrm{L} | \mathrm{L} \in \Omega, \mathrm{L}(A) =$ ON$\}$. If the sample space $\Omega$ is a set of $\{$IN, OUT, UN$\}$-labellings or $\{$IN, OUT, UN, OFF$\}$-labellings, then $L_A =$ ON is a shorthand for the outcomes $\{\mathrm{L} | \mathrm{L} \in \Omega, \mathrm{L}(A) =$ IN or $\mathrm{L}(A) =$ OUT or $\mathrm{L}(A) =$ UN$\}$.

**Notation 3.6**

- *Sets of random labellings are denoted using upper boldface type. So* $\boldsymbol{L}$ *denotes a set of random labellings* $\{L_{A1}, \ldots, L_{An}\}$.

- *Assignments to (sets of) random labellings are denoted using boldface type. For instance, given a set of random labellings* $\boldsymbol{L} = \{L_{A1}, \ldots, L_{An}\}$*, a possible assignment is* $\mathbf{l} = \{L_{A1} = \mathsf{ON}, \ldots, L_{An} = \mathsf{OFF}\}$.
- *An assignment to a set of random labellings can be straightforwardly mapped to a labelling, and for this reason, an 'assignment' may be also called a 'labelling' in the remainder. When the distinction is made, and for the sake of compactness, the labelling* $\mathrm{L}$ *corresponding to an assignment* $\mathbf{l}$ *is denoted* $\mathrm{L}_\mathbf{l}$. *To make a bridge with our notation for labellings and to avoid any ambiguity, we write* $\mathbf{l}_A = l$ *to say that the random labelling of argument* $A$ *is assigned the value* $l$ *according to the assignment* $\mathbf{l}$.
- *The joint distribution over a set* $\boldsymbol{L} = \{L_{A1}, L_{A2}, \ldots, L_{An}\}$ *of random labellings is formally denoted* $P(\{L_{A1}, L_{A2}, \ldots, L_{An}\})$*, but following standard notation, we will write it* $P(L_{A1}, L_{A2}, \ldots, L_{An})$.

From the definition of random variables, the probability of an assignment $\mathbf{l}$ for a set of random labellings is the sum of the probabilities of the labellings $\mathrm{L}_\mathbf{l}$ in the sample space where this assignment occurs:

$$P(\mathbf{l}) = \sum_{\mathrm{L}_\mathbf{l} \in \Omega} P(\{\mathrm{L}_\mathbf{l}\}). \tag{3}$$

Consequently, if the assignment concerns only one argument $A$, then the marginal probability of the status of this argument $A$ is the sum of the probabilities of the labellings in the sample space where this assignment occurs:

$$P(L_A = l) = \sum_{\mathrm{L} \in \Omega : \mathrm{L}(A) = l} P(\{\mathrm{L}\}). \tag{4}$$

**Example 3.1 (continuing from p. 20)**

*Suppose the PLF* $\langle \mathsf{G}, \text{preferred-}\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}, \langle \Omega_{\mathrm{PLF}}, F_{\mathrm{PLF}}, P_{\mathrm{PLF}} \rangle \rangle$ *has a distribution as indicated in Figure 9. By Equation 4, we can compute the marginal probability that arguments* B *or* C *obtain a certain acceptance status, as follows.*

$$\begin{array}{ll}
P_{\mathrm{PLF}}(L_\mathsf{B} = \mathsf{IN}) = P_{\mathrm{PLF}}(\{\mathrm{L}_1\}) + P_{\mathrm{PLF}}(\{\mathrm{L}_3\}) & (= 0.6) \\
P_{\mathrm{PLF}}(L_\mathsf{B} = \mathsf{OUT}) = P_{\mathrm{PLF}}(\{\mathrm{L}_2\}) & (= 0.4) \\
P_{\mathrm{PLF}}(L_\mathsf{B} = \mathsf{UN}) = 0 & \\
P_{\mathrm{PLF}}(L_\mathsf{B} = \mathsf{OFF}) = 0 & \\
P_{\mathrm{PLF}}(L_\mathsf{C} = \mathsf{IN}) = P_{\mathrm{PLF}}(\{\mathrm{L}_2\}) & (= 0.4) \\
P_{\mathrm{PLF}}(L_\mathsf{C} = \mathsf{OUT}) = P_{\mathrm{PLF}}(\{\mathrm{L}_1\}) & (= 0.4) \\
P_{\mathrm{PLF}}(L_\mathsf{C} = \mathsf{UN}) = 0 & \\
P_{\mathrm{PLF}}(L_\mathsf{C} = \mathsf{OFF}) = P_{\mathrm{PLF}}(\{\mathrm{L}_3\}) & (= 0.2)
\end{array}$$

□

Since a random labelling is a random variable, we get directly the desirable property that the probabilities of the different labels for an argument sum up to 1.

**Proposition 3.1** *Let* $\langle G, X\text{-}\mathsf{ArgLab}, \langle \Omega, F, P \rangle \rangle$ *be a PLF.*

$$\sum_{l \in \mathsf{ArgLab}} P(L_A = l) = 1.$$

From Proposition 3.1, for any $X$-$\{\mathsf{ON}, \mathsf{OFF}\}$-labelling specification, we obtain:

$$P(L_A = \mathsf{ON}) + P(L_A = \mathsf{OFF}) = 1. \tag{5}$$

For any $X$-$\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$-labelling specification, we have:

$$P(L_A = \mathsf{IN}) + P(L_A = \mathsf{OUT}) + P(L_A = \mathsf{UN}) = 1. \tag{6}$$

For any $X$-$\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}$-labelling specification:

$$P(L_A = \mathsf{IN}) + P(L_A = \mathsf{OUT}) + P(L_A = \mathsf{UN}) + P(L_A = \mathsf{OFF}) = 1. \quad (7)$$

Stable labellings enjoy more specific properties given that no argument can be labelled UN, and therefore, for the stable-$\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$-labelling specification, we have:

$$P(L_A = \mathsf{IN}) + P(L_A = \mathsf{OUT}) = 1. \quad (8)$$

For the stable-$\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}$-labelling specification:

$$P(L_A = \mathsf{IN}) + P(L_A = \mathsf{OUT}) + P(L_A = \mathsf{OFF}) = 1. \quad (9)$$

Other results on the probability of argument acceptance statuses can be derived. We will review some of them in the remainder.

### 3.2.2 Probabilistic argument justification labellings

We now move to explore how uncertainty on argument acceptance may in turn affect argument justification: in this respect, different considerations can be drawn depending on the set of labels adopted in the underlying PLF. PLFs based on $\{\mathsf{ON}, \mathsf{OFF}\}$-labellings express uncertainty about the structure of the framework and this simple set of labels does not carry significant information about argument justification. The situation is different for PLFs based on $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$ or $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}$-labellings.

Assuming a (possibly random) selection of a unique acceptance labelling amongst a set of $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$ or $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}$-labellings, we can then distinguish argument justification status *before* this selection (which in the present proposal is specified by Definition 2.36) and argument justification status *after* the selection. The latter is directly determined by the unique selected labelling, and, by Definition 2.36 in this restricted case, only three justification labels are possible, namely OFJ, SKJ, and NOJ: after selection of a unique acceptance labelling, an argument $A$ gets the justification label OFJ if and only if $A$ is labelled OFF in the selected labelling, SKJ if and only if it is labelled IN in the selected labelling, and NOJ if and only if it is labelled OUT or UN in the selected labelling.

So, if one assumes, as we do, that an agent at some point makes a definite choice of a unique acceptance labelling among the alternative acceptance labellings produced by a semantics, the traditional notion of argument justification is somehow redundant, since it is related to argument acceptance by direct and simple relationships and some justification notions, like credulous acceptance, actually are not relevant. Argument justification keeps its role if one assumes instead that an agent does not have to choose among the alternative labellings but rather needs to draw a sort of synthetic view about them. Both scenarios make sense in different contexts, the former appearing more suitable for practical reasoning, i.e. reasoning about what to do, the latter for epistemic reasoning, i.e. reasoning about what to believe. Further discussions of these aspects are beyond the scope of the present paper and are left to future work. We just observe that, in the latter scenario, it is possible to identify a relationship between the justification status of an argument and PLFs based on $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}$-labellings, provided that one assumes that each labelling has non zero probability.

**Proposition 3.2** *Let $\langle G, X\text{-}\mathsf{ArgLab}, \langle \Omega, F, P \rangle \rangle$ be a PLF, $\mathcal{L} \subseteq \mathcal{L}^X_{\mathsf{ArgLab}}(G)$ a non-empty set of argument labellings such that for every labelling $\mathrm{L} \in \mathcal{L}$ $P(\{\mathrm{L}\}) > 0$, and $\mathrm{L}_\mathsf{J}$ the semi-skeptical $\{\mathsf{OFJ}, \mathsf{SKJ}, \mathsf{CRJ}, \mathsf{NOJ}\}$-labelling of $G$. For every argument $A$ in $\mathcal{A}_G$, it holds that:*

- $\mathrm{L}_\mathsf{J}(A) = \mathsf{OFJ}$ *if, and only if,* $P(L_A = \mathsf{OFF}) = 1$*;*
- $\mathrm{L}_\mathsf{J}(A) = \mathsf{SKJ}$ *if, and only if,* $P(L_A = \mathsf{IN}) = 1$*;*
- $\mathrm{L}_\mathsf{J}(A) = \mathsf{CRJ}$ *if, and only if,* $0 < P(L_A = \mathsf{IN}) < 1$*;*
- $\mathrm{L}_\mathsf{J}(A) = \mathsf{NOJ}$ *if, and only if,* $P(L_A = \mathsf{OUT}) > 0$ *or* $P(L_A = \mathsf{UN}) > 0$*, and* $P(L_A = \mathsf{IN}) = 0$.

**Proof 3.1** *Let us make a proof for each justification label.*

OFJ. *From Definition 2.36,* $\mathrm{L_J}(A) = \textsf{OFJ}$ *if, and only if,* $\forall \mathrm{L} \in \mathcal{L}$, $\mathrm{L}(A) = \textsf{OFF}$.

*We have '*$\forall \mathrm{L} \in \mathcal{L}$, $\mathrm{L}(A) = \textsf{OFF}$*' if, and only if,* $P(L_A = \textsf{OFF}) = 1$.

*Therefore,* $\mathrm{L_J}(A) = \textsf{OFJ}$ *if, and only if,* $P(L_A = \textsf{OFF}) = 1$.

SKJ. *From Definition 2.36,* $\mathrm{L_J}(A) = \textsf{SKJ}$ *if, and only if,* $\forall \mathrm{L} \in \mathcal{L}$, $\mathrm{L}(A) = \textsf{IN}$. *We have '*$\forall \mathrm{L} \in \mathcal{L}$, $\mathrm{L}(A) = \textsf{IN}$*' if, and only if,* $P(L_A = \textsf{IN}) = 1$. *Therefore,* $\mathrm{L_J}(A) = \textsf{SKJ}$ *if, and only if,* $P(L_A = \textsf{IN}) = 1$.

CRJ. *From Definition 2.36,* $\mathrm{L_J}(A) = \textsf{CRJ}$ *if, and only if,* $\exists \mathrm{L} \in \mathcal{L}$*:* $\mathrm{L}(A) = \textsf{IN}$ *and* $\mathrm{L_J}(A) \neq \textsf{SKJ}$*; and*

1. $\exists L \in \mathcal{L}, \mathrm{L}(A) = \textsf{IN}$*, if, and only if,* $P(L_A = \textsf{IN}) > 0$.
2. $\mathrm{L_J}(A) \neq \textsf{SKJ}$ *if, and only if,* $\exists \mathrm{L} \in \mathcal{L}$*:* $\mathrm{L}(A) \neq \textsf{IN}$. *That is,* $\mathrm{L_J}(A) \neq \textsf{SKJ}$ *if, and only if,* $P(L_A = \textsf{IN}) < 1$.

*Therefore,* $\mathrm{L_J}(A) = \textsf{CRJ}$ *if, and only if,* $0 < P(L_A = \textsf{IN}) < 1$.

NOJ. *From Proposition 2.4,* $\mathrm{L_J}(A) = \textsf{NOJ}$ *if, and only if,* $\exists \mathrm{L} \in \mathcal{L}$, $\mathrm{L}(A) = \textsf{OUT}$ *or* $\mathrm{L}(A) = \textsf{UN}$*, and* $\forall \mathrm{L} \in \mathcal{L}, \mathrm{L}(A) \neq \textsf{IN}$.

*Therefore,* $\mathrm{L_J}(A) = \textsf{NOJ}$ *with respect to* $\mathcal{L}$ *if, and only if,* $P(L_A = \textsf{OUT}) > 0$ *or* $P(L_A = \textsf{UN}) > 0$*, and* $P(L_A = \textsf{IN}) = 0$.

**Example 3.1 (continuing from p. 20)** *Arguments* B *and* C *are both credulously justified, i.e.,* $\mathrm{L_J}(\mathsf{B}) = \textsf{CRJ}$ *and* $\mathrm{L_J}(\mathsf{C}) = \textsf{CRJ}$. □

To recap, PLFs are a convenient basis to define probabilistic measures about the acceptance status of arguments and to derive some essential properties. Turning to argument justification labelling, we showed that, under suitable hypotheses, it is possible to identify a relationship between the justification status of an argument and PLFs based on $\{\textsf{IN}, \textsf{OUT}, \textsf{UN}, \textsf{OFF}\}$-labellings. Probabilistic measures concerning statement labellings and their relationships can be also derived on the basis of PLFs, as we will see in next section.

## 3.3 Probabilistic labelling of statements

The probabilistic labelling of statements directly follows from the probabilistic labelling of arguments. We introduce, as we did for random labellings of arguments, random variables concerning the labelling of statements. For any literal $\varphi$, we introduce a categorical random variable which is denoted $K_\varphi$ and which can take value in the set of labels $\mathsf{LitLabels}$ of a considered $\mathsf{LitLabels}$-labelling of literals. These random variables are also called random labellings.

The marginal probability of a literal labelled $k$ is the sum of labellings in the sample space where the literal is labelled as such:

$$P(K_\varphi = k) = \sum_{\{\mathrm{L} \in \Omega \,|\, \mathrm{K}(\mathrm{L}, \varphi) = k\}} P(\{\mathrm{L}\}). \tag{10}$$

Since $K_\varphi$ is a random variable, the sum of marginal probabilities over its possible assignments equals 1.

**Proposition 3.3** *Let* $\langle G, \mathcal{S}, \langle \Omega, F, P \rangle \rangle$ *be a PLF and* $\varphi$ *a literal. For any* $\mathsf{LitLabels}$*-labelling of literals,*

$$\sum_{k \in \mathsf{LitLabels}} P(K_\varphi = k) = 1.$$

Proposition 3.3 can be instantiated with different statement labellings. For example, given a PLF $\langle G, \mathcal{S}, \langle \Omega, F, P \rangle \rangle$ and the bivalent $\{\mathsf{in}, \mathsf{no}\}$-labellings, we have:

$$P(K_\varphi = \mathsf{in}) + P(K_\varphi = \mathsf{no}) = 1 \tag{11}$$

while in case of worst-case sensitive $\{\mathsf{in}, \mathsf{out}, \mathsf{un}, \mathsf{off}, \mathsf{unp}\}$-labellings we obtain:

$$P(K_\varphi = \mathsf{in}) + P(K_\varphi = \mathsf{out}) + P(K_\varphi = \mathsf{un}) + P(K_\varphi = \mathsf{off}) + P(K_\varphi = \mathsf{unp}) = 1. \qquad (12)$$

Note that either $P(K_\varphi = \mathsf{unp}) = 0$ or $P(K_\varphi = \mathsf{unp}) = 1$, since, given a set of arguments, there is no uncertainty on the fact that a literal $\varphi$ is the conclusion of at least one argument or none in the set.

**Proposition 3.4** *Let $\langle G, \mathcal{S}, \langle \Omega, F, P\rangle\rangle$ be a PLF, and $\varphi_1$ and $\varphi_2$ literals in conflict, i.e. $conflict(\varphi_1, \varphi_2)$ and let us consider the bivalent $\{\mathsf{in}, \mathsf{no}\}$-labellings and the worst-case sensitive $\{\mathsf{in}, \mathsf{out}, \mathsf{un}, \mathsf{off}, \mathsf{unp}\}$-labellings,*

$$P(K_{\varphi_1} = \mathsf{in}) + P(K_{\varphi_2} = \mathsf{in}) \leq 1.$$

**Proof 3.2** *In a given labelling a literal $\varphi_1$ is labelled* in *or not. If the literal $\varphi_1$ is labelled* in *then there is an argument $A$ labelled* IN *such that $conc(A) = \varphi_1$. Then it must be the case that this argument attacks or is attacked by every other argument $A'$ such that $conc(A') = \varphi_2$ and given the basic conflict-freeness property satisfied by any argumentation semantics it can not be the case that any such argument $A'$ is labelled* IN *in the same labelling. Hence also $\varphi_2$ cannot be labelled* in *in the same labelling. If instead the literal $\varphi_1$ is not labelled* in*, then any conflicting literal $\varphi_2$ can labelled* in *or not. Let $\mathcal{K}_{\varphi_1,\mathsf{in}}$ denote the set of labellings of arguments such that $\varphi_1$ is labelled* in*, $\mathcal{K}_{\varphi_2,\mathsf{in}}$ the set of labellings of arguments such that $\varphi_2$ is labelled* in*, and $\mathcal{K}$ the complement set of labellings $\Omega \backslash \mathcal{K}_{\varphi_1,\mathsf{in}} \cup \mathcal{K}_{\varphi_2,\mathsf{in}}$. These three sets form a partition of $\Omega$, thus: $P(\mathcal{K}_{\varphi_1,\mathsf{in}}) + P(\mathcal{K}_{\varphi_2,\mathsf{in}}) + P(\mathcal{K}) = 1$. Since $P(\mathcal{K}) \geq 0$, we have $P(\mathcal{K}_{\varphi_1,\mathsf{in}}) + P(\mathcal{K}_{\varphi_2,\mathsf{in}}) \leq 1$, therefore $P(K_{\varphi_1} = \mathsf{in}) + P(K_{\varphi_2} = \mathsf{in}) \leq 1$.* □

The next proposition is a corollary which follows from propositions 3.3 and 3.4.

**Proposition 3.5** *Let $\langle G, \mathcal{S}, \langle \Omega, F, P\rangle\rangle$ be a PLF, and $\varphi_1$ and $\varphi_2$ literals in conflict, i.e. $conflict(\varphi_1, \varphi_2)$. Let us consider the bivalent $\{\mathsf{in}, \mathsf{no}\}$-labellings and the worst-case sensitive $\{\mathsf{in}, \mathsf{out}, \mathsf{un}, \mathsf{off}, \mathsf{unp}\}$-labellings,*

$$P(K_{\varphi_1} = \mathsf{in}) \leq P(K_{\varphi_2} \neq \mathsf{in})$$

*where*

- $P(K_{\varphi_2} \neq \mathsf{in}) = P(K_{\varphi_2} = \mathsf{no})$ *in the case of bivalent $\{\mathsf{in}, \mathsf{no}\}$-labellings;*
- $P(K_{\varphi_2} \neq \mathsf{in}) = P(K_{\varphi_2} = \mathsf{out}) + P(K_{\varphi_2} = \mathsf{un}) + P(K_{\varphi_2} = \mathsf{off}) + P(K_{\varphi_2} = \mathsf{unp})$ *in the case of worst-case sensitive $\{\mathsf{in}, \mathsf{out}, \mathsf{un}, \mathsf{off}, \mathsf{unp}\}$-labellings.*

Both propositions (3.4) and (3.5) feature inequalities which are consequences of the fact that argumentation does not necessarily fulfil (an argumentation counterpart of) the principle of excluded middle, since it is possible, for example, to have outcomes where, for any statement, neither a statement nor its complement are labelled in – they may be both labelled no for instance. These propositions can also be viewed as an expression of a probabilistic notion of consistency: it is not possible to regard two conflicting statements as highly probable at the same time, because a statement can be believed at most as much as any conflicting statement is disbelieved.

In the following, we provide a complete example of use of the various probabilistic notions we have introduced in this section.

**Example 2.1 (continuing from p. 5)** *The research scientist is asked to report on the degree of uncertainty concerning whether the critical program will be running or not at any point of time of the project. To help her, we decide to use the approach of probabilistic labellings.*

*Firstly, we deal with uncertainty at the level of the rules. The solar panels work well but they only provide power during daytime, thus, assuming that daytime and nighttime are equal, at any point of time there is 50% probability that there is enough solar power, i.e. that the rule $\mathsf{r}_{\neg \mathsf{b1}}$ applies. The battery is partially damaged, and only one expert over five advances the argument concluding that the*

| | $\mathsf{r}_{\neg b1}$ | $\mathsf{r}_{\neg b2}$ | $\mathsf{r}_{\neg b}$ | $\mathsf{r}_{c}$ | $\mathsf{r}_{\neg c}$ | $P_{\mathrm{PTF}}$ |
|---|---|---|---|---|---|---|
| $\mathsf{U}_1$ | 0 | 0 | 1 | 1 | 1 | 0.4 |
| $\mathsf{U}_2$ | 1 | 0 | 1 | 1 | 1 | 0.4 |
| $\mathsf{U}_3$ | 0 | 1 | 1 | 1 | 1 | 0.1 |
| $\mathsf{U}_4$ | 1 | 1 | 1 | 1 | 1 | 0.1 |

Figure 10: Subtheories with non-zero probability.

*battery cannot provide power, giving rise to a probability of* 0.2 *that the rule* $\mathsf{r}_{\neg b2}$ *is considered. All other rules are not affected by uncertainty. Under a reasonable assumption of independence, this gives rise to a PTF as drawn in Figure* 10 *where* $P_{\mathrm{PTF}}(\{\mathsf{U}_1\}) = 0.4$*,* $P_{\mathrm{PTF}}(\{\mathsf{U}_2\}) = 0.4$*,* $P_{\mathrm{PTF}}(\{\mathsf{U}_3\}) = 0.1$*,* $P_{\mathrm{PTF}}(\{\mathsf{U}_4\}) = 0.1$ *and the probability of all other subtheories is zero.*

*The PGF corresponding to this PTF, as illustrated in Figure 11, is such that the full argumentation graph* $\langle\{\mathsf{B1}, \mathsf{B2}, \mathsf{B}, \mathsf{C}, \mathsf{D}\}, \{(\mathsf{B}, \mathsf{C}), (\mathsf{C}, \mathsf{D}), (\mathsf{D}, \mathsf{C})\}, \{(\mathsf{B1}, \mathsf{B}), (\mathsf{B2}, \mathsf{B})\}\rangle$ *has probability* 0.1*, while the subgraphs induced by the sets of arguments* $\{\mathsf{B2}, \mathsf{C}, \mathsf{D}\}$*,* $\{\mathsf{B1}, \mathsf{C}, \mathsf{D}\}$ *and* $\{\mathsf{C}, \mathsf{D}\}$ *have probabilities* 0.1*,* 0.4*, and* 0.4 *respectively.*

| | B1 | B2 | B | C | D | $P_{\mathrm{PGF}}$ |
|---|---|---|---|---|---|---|
| $\mathsf{G}_1$ | 0 | 0 | 0 | 1 | 1 | 0.4 |
| $\mathsf{G}_2$ | 1 | 0 | 0 | 1 | 1 | 0.4 |
| $\mathsf{G}_3$ | 0 | 1 | 0 | 1 | 1 | 0.1 |
| $\mathsf{G}_4$ | 1 | 1 | 1 | 1 | 1 | 0.1 |

Figure 11: Argumentation subgraphs with non-zero probability.

*The expert has a skeptical stance, and thus she can decide to adopt grounded* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}$*-labellings. Accordingly, the argument labellings illustrated in the left part of Figure 12 are generated and we get a PLF where* $P(\{\mathrm{L}_1\}) = 0.4$*,* $P(\{\mathrm{L}_2\}) = 0.4$*,* $P(\{\mathrm{L}_3\}) = 0.1$*,* $P(\{\mathrm{L}_4\}) = 0.1$*.*

| | B1 | B2 | B | C | D | $P$ | ¬b1 | ¬b2 | ¬b | c | ¬c |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathrm{L}_1$ | OFF | OFF | OFF | UN | UN | 0.4 | off | off | off | un | un |
| $\mathrm{L}_2$ | IN | OFF | OFF | UN | UN | 0.4 | in | off | off | un | un |
| $\mathrm{L}_3$ | OFF | IN | OFF | UN | UN | 0.1 | off | in | off | un | un |
| $\mathrm{L}_4$ | IN | IN | IN | OUT | IN | 0.1 | in | in | in | out | in |

Figure 12: Argument and statement labellings with grounded semantics, with non-zero probability.

*The corresponding statement labellings according to Definition 2.40 are illustrated in the right part of Figure 12 from which the following probabilistic labelling of statements can be derived.*

$$
\begin{aligned}
&P(K_{\neg b} = \mathsf{in}) = P(\{\mathrm{L}_4\}) && (= 0.1)\\
&P(K_{\neg b} = \mathsf{off}) = P(\{\mathrm{L}_1\}) + P(\{\mathrm{L}_2\}) + P(\{\mathrm{L}_3\}) && (= 0.9)\\
&P(K_{c} = \mathsf{un}) = P(\{\mathrm{L}_1\}) + P(\{\mathrm{L}_2\}) + P(\{\mathrm{L}_3\}) && (= 0.9)\\
&P(K_{c} = \mathsf{out}) = P(\{\mathrm{L}_4\}) && (= 0.1)\\
&P(K_{\neg c} = \mathsf{in}) = P(\{\mathrm{L}_4\}) && (= 0.1)\\
&P(K_{\neg c} = \mathsf{un}) = P(\{\mathrm{L}_1\}) + P(\{\mathrm{L}_2\}) + P(\{\mathrm{L}_3\}) && (= 0.9)
\end{aligned}
$$

*Altogether, we have derived that, using the skeptical stance of grounded labellings , the overall probability of rejection of statement* c *('the program is running') is* 0.1 *while the probability of acceptance of* ¬c *('the program is not running') is estimated at* 0.1*. The status of* c *and* ¬c *is regarded as undecided with a probability* 0.9*.* □

To summarise this section, we have considered different kinds of frames for probabilistic argumentation, namely PTFs, PGFs and (the novel concept of) PLFs, each frame featuring a different probability space. While, with suitable labels, PLFs are able to encompass the uncertainty about argumentation graphs expressed by PTFs and PGFs, PLFs can also express uncertainty about the final labelling outcome, thus covering a further uncertainty dimension. Considering thus PLFs as a general approach for probabilistic argumentation, we derived some results concerning probabilistic measures of argument and statement labellings.

# 4 On Uncertainty about Inclusion and Acceptance Status

PLFs are an expressive and flexible formalism able to capture various kinds of uncertainty because, in addition to the 'traditional' acceptance labels IN, OUT and UN, we have introduced labels ON and OFF to account for the 'inclusion' status of arguments. In this section, we further develop the analysis of this increased expressiveness by discussing, at a technical level the relationships between our proposal and the treatment given in [27] of two influential approaches to probabilistic argumentation, namely the constellations approach (Subsection 4.1) and the epistemic approach (Subsection 4.2), finally leading to a possible combination (Subsection 4.3). This is achieved by both carrying out a conceptual analysis and proving some technical properties, which provide a basis for a wider discussion of relevant literature at a general level in Section 5.

## 4.1 Constellations approach

In the constellations approach, originally investigated in [36], every argument and attack of an argumentation graph is associated with a 'likelihood'. In this section we discuss the development of this idea presented in [27], while other works related to the constellations approach (e.g. [11]) are discussed in Section 5.

We first recall the definition of probabilistic argumentation graphs (PAGs) from [27] where a probability is directly associated with each argument.

**Definition 4.1 (Probabilistic argumentation graph)** *A* probabilistic argumentation graph *(PAG) is a tuple* $\langle \mathcal{A}, \rightsquigarrow, P_{\mathrm{PAG}} \rangle$ *where* $\langle \mathcal{A}, \rightsquigarrow \rangle$ *is an abstract argumentation graph and* $P_{\mathrm{PAG}} : \mathcal{A} \to [0, 1]$.

The sample space is left implicit, but the interpretation, quoting [27], is that given an abstract argumentation graph $G$ one 'can treat the set of subgraphs of $G$ as a sample space, where one of the subgraphs is the "true" argumentation graph.'

In [27], the probability of a subgraph $H$ of $G$ induced by a set of arguments $\mathcal{A}_H \subseteq \mathcal{A}_G$ is not derived from any form of axiomatisation but is directly defined as the following product (Definition 14 of [27]):

$$P_{\mathrm{PAG}}(H) = \left( \prod_{A \in \mathcal{A}_H} P_{\mathrm{PAG}}(A) \right) \times \left( \prod_{A \notin \mathcal{A}_H} (1 - P_{\mathrm{PAG}}(A) \right). \tag{13}$$

Equation 13 relies on the assumption that for each argument $A$, the probability of $A$ appearing in the 'true' argumentation graph is independent of the probability of appearance of every other argument. This is motivated by the *justification perspective* adopted in [27]: 'knowing that one argument is a justified point does not affect the probability that another is a justified point'. This perspective leads to assume that an assignment of probability values to arguments is given as initial information. Quoting again [27], for each argument $A$, $P(A)$ 'is the probability that $A$ exists in an arbitrary full subgraph of $G$, and $1 - P(A)$ is the probability that $A$ does not exist in an arbitrary full subgraph of $G$'.

The assumption of arguments being probabilistically independent contrasts with our proposal where no assumptions of probabilistic independence is made and, actually, the subargument relation constrains the appearance of arguments in a subgraph. In particular, in the rule-based context we proposed for PTFs, a probability assignment over the subtheories of a theory $T$ is used as a starting point. Then, the probability of a subgraph $H$ is the sum of the probabilities of the subtheories generating $H$, as evidenced in Definition 3.3. Hence, our framework distinguishes and combines logical dependences and probabilistic dependences. Nevertheless, in our setting, if the appearance of arguments is assumed independent, then any PAG and its notion of 'true' argumentation graph can be captured by a PGF as follows.

**Definition 4.2 (PGF corresponding to a PAG)** *Given a PAG $\langle \mathcal{A}, \rightsquigarrow, P_{\mathrm{PAG}} \rangle$, the corresponding PGF is a tuple $\langle G, \langle \Omega_{\mathrm{PGF}}, F_{\mathrm{PGF}}, P_{\mathrm{PGF}} \rangle \rangle$ where $G = \langle \mathcal{A}, \rightsquigarrow \rangle$ and the probability distribution $P_{\mathrm{PGF}}$ is such that $\forall S \in F_{\mathrm{PGF}}$:*

$$P_{\mathrm{PGF}}(S) = \sum_{H \in S} P_{\mathrm{PAG}}(H).$$

The difference between the approach in [27] and ours can be explained by the fact that in the argumentation model based on classical logic adopted in [27] there is no explicit notion of subargument: an argument is a pair $\langle \Phi, \alpha \rangle$ where $\Phi$ is a minimal consistent set of formulae such that $\Phi \vdash \alpha$. In this perspective every argument is self-contained. We note however that it may be the case that there are two arguments $\langle \Phi, \alpha \rangle$, $\langle \Phi', \beta \rangle$ such that $\Phi \subseteq \Phi'$; then it may appear problematic to assume that the appearance of $\langle \Phi', \beta \rangle$ is independent of the appearance of $\langle \Phi, \alpha \rangle$.

A detailed discussion of the differences between logic-based and rule-based argumentation being beyond the scope of this paper, we remark that our approach provides a formal example of the legality constraints binding abstract representations like PGFs and PLFs, when considering the actual underlying argument construction process. This example can be useful as a starting point to investigate analogous legality constraints in other argumentation formalisms. Moreover the labelling representation of the constellations approach through PLFs based on $\{\mathsf{ON}, \mathsf{OFF}\}$-labellings simplifies the analysis of some basic properties, which, again, can be employed as a term of comparison in other formalisms. This is illustrated by the following propositions, whose proofs are straightforward.

First, the probability of inclusion of an argument cannot be greater than the probability of inclusion of its subarguments.

**Proposition 4.1** *Let $\langle G, \mathsf{legal}\text{-}\{\mathsf{ON}, \mathsf{OFF}\}, \langle \Omega, F, P \rangle \rangle$ be a PLF, and let $A$ and $B$ denote two arguments in $\mathcal{A}_G$ such that $A$ is a subargument of $B$.*

$$P(L_B = \mathsf{ON}) \leq P(L_A = \mathsf{ON}).$$

Further, the probability of appearance of an argument $A$ is determined by the probability of the subtheories including the rules utilised in its construction.

**Proposition 4.2** *Let $\mathfrak{F}_1 = \langle T, \langle \Omega_{\mathrm{PTF}}, F_{\mathrm{PTF}}, P_{\mathrm{PTF}} \rangle \rangle$ be a PTF, and $\mathfrak{F}_2 = \langle G_T, \mathcal{S}, \langle \Omega, F, P \rangle \rangle$ its corresponding PLF, i.e. $\mathfrak{F}_2 = PLF(PGF(\mathfrak{F}_1))$. For every argument $A$ in $\mathcal{A}_{G_T}$ it holds that*

$$P_{\mathrm{PLF}}(L_A = \mathsf{ON}) = \sum_{\{U \in Sub(T) | Rules(A) \subseteq U\}} P_{\mathrm{PTF}}(\{U\}).$$

We suggest that properties analogous to those given in propositions 4.1 and 4.2 should be regarded as basic requirements in every proposal belonging to the constellations approach. Simple as they are, such properties are out of the scope of approach focused on abstract argumentation only (see Subsection 5.2) and, as exemplified by the above discussion of [27], they are sometimes overlooked.

Turning to issues related to semantics evaluation, in Definition 15 of [27] the probability of a set of arguments $\mathcal{A}$ being an extension according to an argumentation semantics $X$, denoted $P^X_{\mathrm{PAG}}(\mathcal{A})$, is defined as the sum of the probabilities of subgraphs entailing this extension. The following equation reformulates the original definition in terms of labellings, given that the probability that a set of arguments $\mathcal{A}$ is an extension according to an argumentation semantics $X$ is the probability that all and only arguments in $\mathcal{A}$ are labelled $\mathsf{IN}$ according to an $X$-$\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$-labelling specification:

$$P^X_{\mathrm{PAG}}(\mathcal{A}) = \sum_{H \in \mathcal{H}^X(\mathcal{A})} P_{\mathrm{PAG}}(H) \tag{14}$$

where $\mathcal{H}^X(\mathcal{A}) = \{H | H \in Sub(G) : \mathcal{A} = \mathsf{IN}(\mathrm{L}) \text{ for some } \mathrm{L} \in \mathcal{L}^X_H\}$. On the basis of Definition 4.2, $P^X_{\mathrm{PAG}}(\mathcal{A})$ can be expressed equivalently in the context of PGFs given that of $P_{\mathrm{PAG}}(H) = P_{\mathrm{PGF}}(\{H\})$.

While any PAG can be captured by a PGF or the corresponding PLF (Definition 3.6) with $\{$ON, OFF$\}$-labellings, we have to emphasise that the acceptance uncertainty captured by PLFs is possibly distinct from the uncertainty captured by $P^X_{\text{PAG}}$, and provides a more fine-grained probabilistic evaluation of argument acceptance statuses, as illustrated in Example 3.1.

**Example 3.1 (continuing from p. 20)** *Suppose the PAG* $\langle \mathcal{A}, \rightsquigarrow, P_{\text{PAG}} \rangle$ *and the PLF* $\langle \langle \mathcal{A}, \rightsquigarrow \rangle$, preferred-$\{$IN, OUT, UN, OFF$\}$, $\langle \Omega_{\text{PLF}}, F_{\text{PLF}}, P_{\text{PLF}} \rangle \rangle$ *have probability distributions as given in Figure 13.*

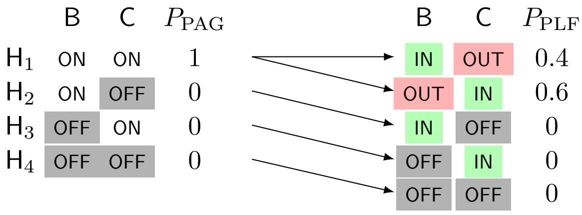


Figure 13: An example of relation and comparison between PAG and PLF.

*We have that*

- *the distribution* $P_{\text{PAG}}$ *is such that* $P_{\text{PAG}}(\text{H1}) = 1$*, reflecting that the inclusion of* B *and* C *have probability* 1*; and*
- *the distribution* $P_{\text{PLF}}$ *is such that* $P_{\text{PLF}}(\{\langle \{\text{B}\}, \{\text{C}\}, \emptyset, \emptyset \rangle\}) = 0.4$ *and* $P_{\text{PLF}}(\{\langle \{\text{C}\}, \{\text{B}\}, \emptyset, \emptyset \rangle\}) = 0.6$*, reflecting that the preferred extension* $\{\text{C}\}$ *is more likely than the preferred extension* $\{\text{B}\}$*.*

*The probability that the arguments* B *and* C *appear in a preferred extension/labelling is then different in the two approaches:*

- *the probability that* $\{\text{B}\}$ *is a preferred extension is 1, and equally for the set* $\{\text{C}\}$*, i.e* $P^{\text{pr}}_{\text{PAG}}(\{\text{B}\}) = 1$ *and* $P^{\text{pr}}_{\text{PAG}}(\{\text{C}\}) = 1$*;*
- *the probability that* B *is labelled* IN *is 0.4, while the probability that* C *is labelled* IN *is 0.6, i.e* $P^{\text{pr}}_{\text{PLF}}(L_{\text{B}} = \text{IN}) = 0.4$ *and* $P^{\text{pr}}_{\text{PLF}}(L_{\text{C}} = \text{IN}) = 0.6$.

*This example illustrates that the acceptance uncertainty addressed with PLFs is distinct from the uncertainty subject of the constellations approach, and supports a more fine-grained probabilistic evaluation of acceptance statuses of arguments.* □

To recap, we have shown that the approach of probabilistic labellings can capture the constellations approach, but the latter cannot capture the former since probabilistic labellings allow a more fine-grained probabilistic evaluation of argument acceptance statuses. Some basic properties, e.g. the fact that the probability of inclusion of an argument cannot be greater than the probability of inclusion of its subarguments, are directly derived in our framework. It turns out that the approach of probabilistic labellings can be also related to another approach to probabilistic argumentation, namely the epistemic approach, as we will see next.

## 4.2 Epistemic approach

In the epistemic approach, 'the probability distribution over arguments is used directly to identify which arguments are believed' [27]. The idea is that in this case the argumentation framework is fixed but an agent has an 'extra epistemic information' to assign an epistemic probability, denoted in the following as $P_{\text{PEF}}$, to each argument. Formally, given a set of arguments $\mathcal{A}$, $P_{\text{PEF}} : \mathcal{A} \rightarrow [0, 1]$. Given that this degree of belief is based on extra information, as exemplified in [27] it may be the case that an agent assigns a high degree of belief to an argument which would be rejected according to the acceptance labelling prescribed by the framework. Even more, in [27], it is assumed that the degrees of belief of an agent may even be inconsistent i.e. that $P_{\text{PEF}}$ may violate the probability axioms. Dealing with inconsistencies is beyond the scope of the present paper, hence we restrict to consistent epistemic probabilities in the following discussion, showing how they can be related to our approach.

First, we formalise the probability space for epistemic probabilities in our context.

**Definition 4.3 (Probabilistic epistemic frame)** *A probabilistic epistemic frame (PEF) is a tuple $\langle G, \langle \Omega_{\text{PEF}}, F_{\text{PEF}}, P_{\text{PEF}} \rangle \rangle$ where $G = \langle \mathcal{A}, \rightsquigarrow, \Rightarrow \rangle$ denotes an argumentation graph, and $\langle \Omega_{\text{PEF}}, F_{\text{PEF}}, P_{\text{PEF}} \rangle$ is a probability space such that:*

- *the sample space $\Omega_{\text{PEF}}$ is the set of subsets of $\mathcal{A}$, i.e. $\Omega = 2^{\mathcal{A}}$;*
- *the $\sigma$-algebra $F_{\text{PEF}}$ is the power set of $\Omega_{\text{PEF}}$, i.e. $F_{\text{PEF}} = 2^{\Omega_{\text{PEF}}}$;*
- *the function $P_{\text{PEF}}$ from $F_{\text{PEF}}$ to $[0, 1]$ is a probability distribution satisfying Kolmogorov axioms.*

Each element of the sample space of a PEF is a set of arguments, an option for the unconstrained choice of which arguments are believed. The epistemic probability, i.e. the degree of belief, of an argument is the sum of the probabilities of the sets of arguments including it. So given a PEF tuple $\langle G, \langle \Omega_{\text{PEF}}, F_{\text{PEF}}, P_{\text{PEF}} \rangle \rangle$, we have:

$$P_{\text{PEF}}(A) = \sum_{\{\mathcal{E} \in \Omega_{\text{PEF}} | A \in \mathcal{E}\}} P_{\text{PEF}}(\{\mathcal{E}\}). \tag{15}$$

Every PEF can be put in correspondence with a PLF based on the $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$ labels, where the label $\mathsf{UN}$ is not used. Basically the epistemic probability of each element of the sample space of a PEF, i.e. of each set of arguments, is put in correspondence with the probability of a labelling where the members of the set are labelled $\mathsf{IN}$ and the other arguments are labelled $\mathsf{OUT}$.

**Definition 4.4 (PLF corresponding to a PEF)** *Given a PEF $\langle G, \langle \Omega_{\text{PEF}}, F_{\text{PEF}}, P_{\text{PEF}} \rangle \rangle$, its corresponding PLF is a PLF tuple $\langle G, \{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}, \langle \Omega_{\text{PLF}}, F_{\text{PLF}}, P_{\text{PLF}} \rangle \rangle$ whose probability distribution $P_{\text{PLF}}$ is such that $\forall S \in F_{\text{PLF}}$:*

$$P_{\text{PLF}}(S) = \sum_{\mathcal{E} \in \mathfrak{E}} P_{\text{PEF}}(\{\mathcal{E}\})$$

*where $\mathfrak{E} = \{\mathcal{E} \in \Omega_{\text{PEF}} | \exists \mathrm{L} \in S : \mathsf{IN}(\mathrm{L}) = \mathcal{E}, \forall A \notin \mathcal{E} : \mathrm{L}(A) = \mathsf{OUT}\}$.*

**Definition 4.5 (PEF corresponding to a PLF)**

*Given a PLF $\langle G, \text{all-}\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}, \langle \Omega_{\text{PLF}}, F_{\text{PLF}}, P_{\text{PLF}} \rangle \rangle$, its corresponding PEF is a PEF tuple $\langle G, \langle \Omega_{\text{PEF}}, F_{\text{PEF}}, P_{\text{PEF}} \rangle \rangle$ whose probability distribution $P_{\text{PEF}}$ is such that $\forall S \in F$:*

$$P_{\text{PEF}}(S) = \sum_{L \in \mathcal{L}} P_{\text{PLF}}(\{\mathrm{L}\})$$

*where $\mathcal{L} = \{\mathrm{L} \in \mathcal{L}_{\{\mathsf{IN},\mathsf{OUT},\mathsf{UN}\}}(G) \mid \exists \mathcal{E} \in S : \mathcal{E} = \mathsf{IN}(\mathrm{L})\}$.*

So, each labelling L in the sample space of the PLF is put in correspondence with the set of arguments $\mathsf{IN}(\mathrm{L})$ belonging to the sample space of a corresponding PEF. In particular, from Equation 15 and taking into account Proposition (4), we directly obtain for every argument $A$ in $\mathcal{A}$:

$$P_{\text{PEF}}(A) = P_{\text{PLF}}(L_A = \mathsf{IN}). \tag{16}$$

**Example 3.1 (continuing from p. 20)** *Suppose the PEF $\langle \mathsf{G}, \langle \Omega_{\text{PEF}}, F_{\text{PEF}}, P_{\text{PEF}} \rangle \rangle$ and the PLF $\langle \mathsf{G}, \text{preferred-}\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}, \langle \Omega_{\text{PLF}}, F_{\text{PLF}}, P_{\text{PLF}} \rangle \rangle$, where the argumentation graph $\mathsf{G}$ is pictured in Figure 8. Suppose that the probability acceptance of $\mathsf{B}$ and $\mathsf{C}$ are $0.4$ and $0.6$, respectively, then we have $P_{\text{PEF}}(\mathsf{B}) = P(L_{\mathsf{B}} = \mathsf{IN}) = 0.4$ and $P_{\text{PEF}}(\mathsf{C}) = P(L_{\mathsf{C}} = \mathsf{IN}) = 0.6$. In this example, values for the probabilities of acceptance are chosen to be consistent with respect to the preferred extension/labellings. However, a distribution $P_{\text{PEF}}$ does not necessarily satisfy this constraint. For example one may set $P_{\text{PEF}}(\{\mathsf{B}, \mathsf{C}\}) > 0$.* □

As mentioned in Example 3.1, the epistemic approach includes the consideration of possibly anomalous probability distributions, and for this reason one may/should consider some desirable properties for epistemic probabilities. In that regard, the proposed correspondences between PEFs and PLFs preserve some desirable properties stated in [27] for such epistemic probabilities. In particular, an epistemic probability $P_{\text{PEF}}$ is called 'coherent' if for every pair of arguments $A$ and $B$ such that $A$ attacks $B$ it holds that $P_{\text{PEF}}(A) + P_{\text{PEF}}(B) \leq 1$. Coherence implies a weaker property, called 'rationality' in [27], namely that if $P_{\text{PEF}}(A) > 0.5$ then $P_{\text{PEF}}(B) \leq 0.5$. In our framework, coherence of $P(L_A = \mathsf{IN})$ is ensured, provided that the labelling satisfies the minimal property of conflict freeness.

**Proposition 4.3** *Let $\langle G, \mathsf{cf}\text{-}\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}, \langle \Omega, F, P\rangle\rangle$ be a PLF. Let $A$ and $B$ denote two arguments in $\mathcal{A}_G$ such that $B$ attacks $A$.*

$$P(L_A = \mathsf{IN}) + P(L_B = \mathsf{IN}) \leq 1.$$

**Proof 4.1** *Let us denote $\mathcal{L}_A$ the set of labellings where $\mathrm{L}(A) = \mathsf{IN}$, $\mathcal{L}_B$ the set of labellings where $\mathrm{L}(B) = \mathsf{IN}$, and $\mathcal{L}$ the complement set of labellings $\Omega \backslash (\mathcal{L}_A \cup \mathcal{L}_B)$. By conflict freeness, $\mathcal{L}_A$ and $\mathcal{L}_B$ are disjoint, we have thus: $P(\mathcal{L}_A) + P(\mathcal{L}_B) + P(\mathcal{L}) = 1$. By definition, $P(\mathcal{L}_A) = P(L_A = \mathsf{IN})$ and $P(\mathcal{L}_B) = P(L_B = \mathsf{IN})$. We also have that $P(\mathcal{L}) \geq 0$. Therefore $P(L_A = \mathsf{IN}) + P(L_B = \mathsf{IN}) \leq 1$.* $\square$

Further it is easy to see that $P(L_A = \mathsf{IN})$ satisfies other properties under additional mild assumptions. First of all an epistemic probability is said to be 'founded' [28] if every argument not receiving any attack has probability 1. This property is directly achieved if one assumes complete labellings.

**Proposition 4.4** *Let $\langle G, \mathsf{complete}\text{-}\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}, \langle \Omega, F, P\rangle\rangle$ be a PLF. Let $A$ be any argument in $\mathcal{A}_G$ such that no arguments attack $A$.*

$$P(L_A = \mathsf{IN}) = 1.$$

**Proof 4.2** *In the case of the* complete-$\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$*-labelling, for any labelling* $\mathrm{L}$ *in the sample space* $\Omega$*, we have* $\mathrm{L}(A) = \mathsf{IN}$*, therefore* $P(L_A = \mathsf{IN}) = 1$*. In the case of the* complete-$\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}$*-labelling, for any labelling* $\mathrm{L}$ *in the sample space* $\Omega$*, we have* $\mathrm{L}(A) \neq \mathsf{IN}$ *only if* $\mathrm{L}(A) = \mathsf{OFF}$*, therefore* $P(L_A = \mathsf{IN}) + P(L_A = \mathsf{OFF}) = 1$. $\square$

While some properties considered in the literature have a direct counterpart within the proposed framework, as discussed above, for some others the situation is not so clearcut. Consider for instance 'optimistic' distributions proposed in [28]. Letting $\mathcal{B}$ denote the set of all the attackers of an argument $A$, a distribution is optimistic if for every argument $A$

$$P(L_A = \mathsf{IN}) \geq 1 - \sum_{B \in \mathcal{B}} P(L_B = \mathsf{IN}). \tag{17}$$

Considering $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$-labellings, the constraint may become trivial in the case the probability of some of the attackers of $A$ is high (i.e. when $\sum_{B \in \mathcal{B}} P(L_B = \mathsf{IN}) \geq 1$). However, the constraint does not always hold. For instance, if for every attacker $B$ of $A$ the distribution is such that $P(L_B = \mathsf{IN}) = 0$, then the constraint becomes an equality, i.e. $P(L_A = \mathsf{IN}) = 1$ and is not necessarily satisfied.

So, epistemic probabilities can be put in direct correspondence with PLFs based on $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$-labellings, where the focus is on arguments labelled $\mathsf{IN}$ only. The basic properties of coherence and foundedness introduced for epistemic probabilities have a counterpart in PLFs in terms of properties of labellings, while other properties of epistemic probabilities like optimism have no counterparts and their conceptual status is open to discussion (see [4] for a more extensive analysis on this point).

Moreover, our proposal allows one to directly extend the notion of epistemic probability from arguments to argument conclusions: it is rather natural to assume that the final goal of an agent is to express his/her degrees of belief on the statements about which arguments are built, rather than just about arguments.

As a matter of fact, in [27] the probability of a claim is derived from the probability on the underlying logical models, which is used as input information, independently of the probability of arguments. Indeed, in this context, the probability of each claim can be computed directly from the input information without argument construction. The probability of each argument is also computed from the input information and has no effect on the probability of the relevant claim (cf. Proposition 5 of [27] where it is shown that the probability of a claim is not less than the probability of any argument supporting it). In our rule-based approach, using the probability on subtheories as input, the probability of claims is evaluated through the argument construction process, which, more coherently with the notion of argument-based reasoning, plays a central role. Further, our approach allows a fine-grained uncertain evaluation of claims, with probabilities associated with all the relevant labels, while in [27] a binary evaluation of claims is implicitly adopted.

To recap, the epistemic approach can be put in correspondence with the approach of probabilistic labellings through PLFs based on $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$-labellings. In such PLFs, some desirable epistemic properties like 'coherence' and 'foundedness' are guaranteed under some basic assumptions, like conflict-freeness, whereas other properties do not hold in general. Then, while in [27] the probability of a claim can be computed without argument construction, in our proposal the probability of claims is evaluated through the construction of arguments, an account which appears more coherent with the notion of argument-based reasoning. Eventually, probabilistic labellings can capture a combination of the constellations and epistemic approaches, as we will see next.

## 4.3 Combining the constellations and the epistemic approach

The constellations and epistemic approaches have been treated separately in the previous literature as referring to different types of uncertainty. It is worth remarking, however, that they are, so to say, orthogonal. Consequently, a situation where both types of uncertainty are present cannot be excluded: an agent may be uncertain about which argumentation graph to actually consider and, at the same time, may have different degrees of belief about the arguments accepted in the actual (unknown) argumentation graph.

PLFs based on $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}$-labellings provide a way to encompass jointly the two kinds of uncertainty. In fact, extending to this kind of labelling the reasoning illustrated in the previous sections, the probability of inclusion of an argument $A$ corresponds to $P(L_A \neq \mathsf{OFF})$ (also denoted as $P(L_A = \mathsf{ON})$), while the epistemic probability of $A$ corresponds to $P(L_A = \mathsf{IN})$.

The correspondences between the different probabilistic approaches preserve essential properties in the generalised framework. For example, propositions 4.3 and 4.4 can be turned into propositions 4.5 and 4.6 to take into account the inclusion status of arguments.

**Proposition 4.5** *Let $\langle G, \mathsf{cf}\text{-}\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}, \langle \Omega, F, P\rangle\rangle$ be a PLF. Let $A$ and $B$ denote two arguments in $\mathcal{A}_G$ such that $B$ attacks $A$.*

$$P(L_A = \mathsf{IN}) + P(L_B = \mathsf{IN}) \leq 1.$$

**Proposition 4.6** *Let $\langle G, \mathsf{complete}\text{-}\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}, \langle \Omega, F, P\rangle\rangle$ be a PLF. Let $A$ be any argument in $\mathcal{A}_G$ such that no arguments attack $A$.*

$$P(L_A = \mathsf{IN}) + P(L_A = \mathsf{OFF}) = 1.$$

Other propositions can be derived, for example, to relate the probability of acceptance and inclusion of arguments, cf. Proposition 4.1.

**Proposition 4.7** *Let $\langle G, \mathsf{cf}\text{-}\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}, \langle \Omega, F, P\rangle\rangle$ be a PLF, and let $A$ denote any argument in $\mathcal{A}_G$.*

$$P(L_A = \mathsf{IN}) \leq P(L_A = \mathsf{ON}).$$

**Corollary 4.1** *Let $\langle G, \mathsf{cf}\text{-}\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}, \langle \Omega, F, P\rangle\rangle$ be a PLF, and let $A$ and $B$ denote two arguments in $\mathcal{A}_G$ such that $A$ is a subargument of $B$.*

$$P(L_B = \mathsf{IN}) \leq P(L_A = \mathsf{ON}).$$

Eventually, our proposal is able to capture in a formal way the effect of the combination of these different kinds of uncertainty on argument conclusions, in particular in the case where a multi-labelling semantics is adopted, as illustrated in the following example.

**Example 2.1 (continuing from p. 5)** *Suppose that the research scientist, in light of some previous experiences, estimates that the acceptance of* C *is twice more credible than the acceptance of* D*, whenever the choice is open, namely in the cases where they are both labelled* UN *by grounded semantics in the example at the end of Section 3. To convey this, she decides to adopt preferred* {IN, OUT, UN, OFF}*-labellings, and thus the labellings for arguments and statements illustrated in Figure 14 are now deemed possible.*

| | B1 | B2 | B | C | D | ¬b1 | ¬b2 | ¬b | c | ¬c |
|---|---|---|---|---|---|---|---|---|---|---|
| $L_1$ | OFF | OFF | OFF | OUT | IN | off | off | off | out | in |
| $L_2$ | OFF | OFF | OFF | IN | OUT | off | off | off | in | out |
| $L_3$ | IN | OFF | OFF | OUT | IN | in | off | off | out | in |
| $L_4$ | IN | OFF | OFF | IN | OUT | in | off | off | in | out |
| $L_5$ | OFF | IN | OFF | OUT | IN | off | in | off | out | in |
| $L_6$ | OFF | IN | OFF | IN | OUT | off | in | off | in | out |
| $L_7$ | IN | IN | IN | OUT | IN | in | in | in | out | in |

Figure 14: Argument and statement labellings with preferred semantics.

*Then, the different argument labellings are assigned the following probabilities:*

$$
\begin{array}{ll}
P(\{L_1\}) = (1 - 0.5) \times (1 - 0.2) \times (1/3) & (\approx 0.13) \\
P(\{L_2\}) = (1 - 0.5) \times (1 - 0.2) \times (2/3) & (\approx 0.27) \\
P(\{L_3\}) = 0.5 \times (1 - 0.2) \times (1/3) & (\approx 0.13) \\
P(\{L_4\}) = 0.5 \times (1 - 0.2) \times (2/3) & (\approx 0.27) \\
P(\{L_5\}) = (1 - 0.5) \times 0.2 \times (1/3) & (\approx 0.03) \\
P(\{L_6\}) = (1 - 0.5) \times 0.2 \times (2/3) & (\approx 0.07) \\
P(\{L_7\}) = 0.5 \times 0.2 & (= 0.1)
\end{array}
$$

*The following probabilistic labelling of statements can then be derived.*

$$
\begin{array}{ll}
P(K_{\neg b} = \mathsf{in}) = P(\{L_7\}) & (= 0.1) \\
P(K_{\neg b} = \mathsf{off}) = 1 - P(\{L_7\}) & (= 0.9) \\
P(K_{c} = \mathsf{in}) = P(\{L_2\}) + P(\{L_4\}) + P(\{L_6\}) & (\approx 0.61) \\
P(K_{c} = \mathsf{out}) = 1 - P(K_{c} = \mathsf{in}) & (\approx 0.39) \\
P(K_{\neg c} = \mathsf{in}) = P(\{L_1\}) + P(\{L_3\}) + P(\{L_5\}) + P(\{L_7\}) & (\approx 0.39) \\
P(K_{\neg c} = \mathsf{out}) = 1 - P(K_{\neg c} = \mathsf{in}) & (\approx 0.61)
\end{array}
$$

*On the basis of arguments and their approximate probabilities of being included or accepted in an argumentative reasoning, we have derived that the overall probability of acceptance of statement* c *('the program is running') is approximately* 0.61 *while the probability of acceptance of* ¬c *('the program is not running') is estimated at* 0.39*.* □

To sum up this section, probabilistic labellings can capture the constellations approach but not vice versa because probabilistic labellings yield a more fine-grained probabilistic evaluation of argument acceptance statuses. Probabilistic labellings can also be put in correspondence with the epistemic approach, and provide a formal tool to back or question its assertions. Finally, the constellations and epistemic approaches can be seamlessly combined using probabilistic labellings.

# 5 Discussion of Related Literature

We relate in this section our proposal with other relevant literature on probabilistic argumentation. We classify these works into three groups: probabilistic structured argumentation (Subsection 5.1), probabilistic abstract argumentation (Subsection 5.2), and other approaches connecting argumentation and probability at different levels (Subsection 5.3).

## 5.1 Probabilistic structured argumentation

A significant amount of work on probabilistic argumentation concerned structured argumentation where the origin and structure of arguments are explicitly dealt with. These are discussed next, distinguishing between those which, like in our proposal, use or can be easily related to abstract argumentation frameworks for argument or statement acceptance evaluation and those which have little or no relation with this formalism.

### 5.1.1 Probabilistic structured argumentation related to Dung's framework

Several structured probabilistic argumentation constructs in the literature use or can be easily related to Dung's framework, as a component for the evaluation of argument or statement acceptance in the context of a staged process analogous to the one presented in this paper.

For instance, the probability of the defeasible status of a conclusion is employed in [54] to determine strategies to maximise the chances of winning legal disputes, in the context of a multi-layer argumentation formalism based on Defeasible Logic, which can be interpreted in terms of Dung's framework [22, 34]. In this work, the probability of the defeasible status of a conclusion is the sum of cases where this status is entailed. A case is a set of premises which are assumed independent; the treatment of uncertainty in this context can be seen as a restricted account of PTFs.

The probability to win a legal dispute, and thus the probability that arguments and statements get accepted by an adjudicator is explored in [53], on the basis of a probabilistic variant of a rule-based argumentation framework akin to the framework in [47]. The proposal has a focus on the computation of probabilities of acceptance reflecting moves in a dialogue, with a particular emphasis on the distinction between the probability of the event of construction of an argument and the probability of acceptance of an argument. Premises of arguments are assumed independent, so that the probability of an argument is the product of the probability of its premises. [52] further developed the idea of computing the probability of the event of an argument to a rule-based argumentation framework where premises are rules, a treatment which is similar to PTFs. The setting in [52] relaxes the assumption of independent rules by associating a potential with theories. The potentials are then learned through reinforcement learning to match the softmax policies of reinforced learning agents.

A probabilistic development of a rule-based argumentation system inspired from the ASPIC$^+$ formalism is proposed in [48]. The framework shares with [52] multiple intuitions and similar results, in particular a treatment of probabilistic rules which is close to our PTFs, and a notion of theory state which corresponds to our notion of subtheory.

A probabilistic argumentation formalism built on Assumption-Based Argumentation [7] is introduced in [13] for jury-based dispute resolution. Arguments are constructed from a set of probabilistic rules and they are evaluated in the context of probabilistic frameworks with Dung's grounded semantics. Every juror has a different probability space to model the fact that they may reach different probabilistic conclusions for the same case.

The present work can be regarded as a systematic generalisation of the ideas presented in these papers at a more initial level, with dedicated formalisms and/or within specific application contexts. It provides a framework where these proposals can be placed, their structure analysed in a principled way and their properties investigated in a domain independent manner.

### 5.1.2 Other probabilistic structured argumentation

A few other works in the literature deal with probabilistic argumentation within formalisms which are not relying on Dung's framework for the evaluation of arguments.

A probabilistic setting for non-monotonic reasoning is investigated in [24]. In this proposal, uncertainty concerns the truth values assumed by some propositional variables called probabilistic variables which are a strict subset of the whole set of variables of interest for an agent. A total assignment of truth values to the probabilistic variables is called a scenario, and basically the agent is uncertain about which scenario to choose. The agent is also equipped with a knowledge base $\Phi$, which is assumed to be certain, and the chosen scenario must be consistent with $\Phi$, i.e. the scenario and $\Phi$ together must not entail contradictory conclusions. A consistent scenario is called an argument for a conclusion $\varphi$ if, together with $\Phi$, it entails the truth of the conclusion, while it is called counterargument for $\varphi$ if, together with the given knowledge base, it entails the falsity of the conclusion. In this sense, an argument can be understood as a consistent set of assumptions, somehow similarly to assumption-based argumentation, but with the difference that this set of assumptions, being a scenario, is exhaustive. Note that inconsistent scenarios are ruled out and that a consistent scenario may neither be an argument nor a counterargument for $\varphi$, if it does not entail neither the truth nor the falsity of $\varphi$. Moreover, a scenario may be, at the same time, an argument for several conclusions and a counterargument for several others. According to this peculiar notion of argument, there is no notion of argument acceptance in this context, since every scenario stands alone and the arguments and counterarguments corresponding to each scenario cannot be in conflict. For this reason the approach focuses on argument conclusions: on the basis of the probability of each scenario, it is possible to define the probability of a conclusion being supported, being rejected and of being neither supported nor rejected. From these values the notions of degree of support, degree of possibility, and degree of ignorance of a conclusion are defined. While, as evidenced above, this proposal adopts a quite specific notion of argument with respect to the literature, it may be remarked that some of its elements can be put in correspondence with our approach: uncertainty about the probabilistic variables can be regarded as a quite restricted case of PTF, and the degrees of support, possibility, and ignorance can be connected to, respectively, the probability for a statement of being labelled in, of not being labelled out and of being labelled unp.

The probability on statement statuses is catered for in [57], tackling similar issues as introduced in [53], but without concerns for reflecting the structure of dialogues and without using Dung's framework. The authors defined a probabilistic argumentation logic and implemented it with the language CHRiSM [56], which is a rule-based probabilistic logic programming language based on Constraint Handling Rules (CHR) [19] associated with a high-level probabilistic programming language called PRISM [55][5]. They discussed how it can be seen as a probabilistic generalization of the defeasible logic proposed by [41], and showed how it provides a method to determine the initial probabilities from a given body of precedents. Rules in CHRiSM have an attached probability and each rule is applied with respect to its probability. Thus, also in this case, probabilistic uncertainty can be related to a sort of PTF.

## 5.2 Probabilistic abstract argumentation

Several works focus on extending Dung's formalism of abstract argumentation frameworks with probabilistic information, independently of any underlying mechanism of argument construction and of any intended goal of statement evaluation.

### 5.2.1 Constellations approach

The proposal in [36], which can be regarded as the starting point of the constellations approach, makes a strong assumption of independence on the inclusion of arguments, as already commented in Section 4. It has also to be mentioned that, differently from our proposal, attacks are also uncertain in [36]: given a pair of arguments $A$ and $B$ the fact that $A$ attacks $B$ is subject to a probabilistic evaluation and this attack may be present or not in different frameworks where both $A$ and $B$ are included. In the rule-based construction of arguments adopted in the present paper, we do not consider uncertainty about attacks, since in our context the fact that an argument attacks another is deterministic. Encompassing uncertainty about attacks through a generalised labelling approach, where also attacks are labelled, does not appear to pose specific technical difficulties and can be exploited in future work.

[5] This language must not be confused with the system PRISM for probabilistic model checking developed by [26].

In the context of the formal framework of [36], further work [15] investigates the computational complexity of computing the probability of a family of binary predicates, whose truth value depends on the actual structure of the argumentation framework. These predicates concern the fact that a given set of arguments is an extension according to a given semantics, so they are parametric with respect to the chosen semantics. Admissible and stable semantics turn out to be tractable in this respect, while other semantics are not. This analysis has been extended in [16] where it is shown that lifting the assumption of independence leads to intractability in all cases. The analysis is complemented in [17] by Monte-Carlo simulations for the efficient estimation of the probability values in the intractable cases. In a parallel line of research, the semantics of probabilistic argumentation is formulated in [38, 39] by characterising subgraphs in relation to an extension, possibly leading to more efficient algorithms to deal with the complexity of the constellations approach.

The limits of the independence assumption about the inclusion of arguments is also acknowledged in [37], which utilises evidential argumentation frameworks/systems [42, 43] to lift the assumption. The formalism uses a special argument, denoted as $\eta$, to represent evidence and introduces a relation of support between arguments, which is regarded as uncertain, leading to the notion of Probabilistic Evidential Argumentation Frameworks (PrEAFs), where a probability value is associated with each support. Our proposal does resort to neither specific notions like incontrovertible premises nor to a specific evidence-based interpretation of the notion of support. We rather point out that the generic notion of subargument creates by itself a dependence among arguments and induces some legality constraints, as explained in Section 2.

The work presented in [11] shows some closer similarities to ours. First, it introduces a notion of probabilistic argumentation framework where a joint probability distribution $P$ over the set of arguments is given. For any argument $A$, $P(A)$ denotes the probability that $A$ holds 'in isolation, before the dialectical process starts.' This notion is coherent with the constellations approach and with the kind of uncertainty we capture with $\{\mathsf{ON}, \mathsf{OFF}\}$-labellings. It is worth remarking that, by assuming a joint probability distribution over arguments, this work does not rely on any independence assumption. Similarly to our approach, [11] advances a distinct probability assessment with reference to $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$-labellings for arguments. Several differences are however worth pointing out. First, at a representation level, the label OUT also covers the cases where an argument is not included in a framework, namely the cases which we separately label as OFF. Second, only grounded and preferred semantics are explicitly considered in [11]. Third, and more importantly, in [11], the probabilities of acceptance, rejection and undecidedness of an argument $A$ (i.e. the probabilities that $A$ is labelled IN, OUT, or UN respectively) are assumed to be computable from the initial probabilities of the arguments in isolation (and indeed an algorithm is provided to carry out this computation), while this is not the case in our approach. This point deserves a specific comment also because an analogous assumption is adopted in other works belonging to the constellations approach [36, 17] in the context of extension-based semantics. In our approach, a semantics can produce multiple labellings for a given (sub) framework. So, an agent can assign different probability values to each labelling. This captures uncertainty about the selection of acceptance labellings even when there is no uncertainty about the inclusion of arguments in the frameworks; this type of acceptance uncertainty is not addressed in [11]. To exemplify, if for every argument $A$, $P(A) = 1$ or $P(A) = 0$, i.e. there is only one possible (and actually certain) subframework of the original framework, in [11] the probability of acceptance, rejection or undecidedness of every arguments is either 1 or 0 in turn (this holds for all the evaluation alternatives in the paper, namely grounded, sceptical preferred, and credulous preferred). As a consequence of the different assumptions on the nature of uncertainty to be represented, this constraint does not hold in our approach as far as multiple status semantics, like preferred semantics, are concerned. Clearly, these different assumptions may be appropriate in different domains: comparing the suitability of the various existing approaches to the needs of distinct application contexts is left to future work.

The interest in combinations of argumentation and probability is also witnessed by the MARF (Markov Argumentation Random Field) software system based on the combination of abstract argumentation and Markov random fields presented in [58]. In [58], argument acceptability status can take four values: accepted (A), rejected (R), undecided (U), and ignored (I). and the system computes a probabilistic acceptability distribution on these values. Exploring in detail the relationships between these values and our labels and investigating the integration of our approach with Markov Random

Field theory appear to be interesting lines of future development.

### 5.2.2 Epistemic approach

Besides the constellations approach, other works interpret probabilities associated with abstract arguments according to the epistemic approach which is originally introduced in [59] and further developed in [29, 30]. An equational approach to probabilistic abstract argumentation which bears some similarity with the epistemic approach of [59] is studied in [20], where a syntactical and a semantical method to the definition of the probabilities of arguments are proposed. [4] explored an alternative setting for epistemic probabilities in abstract argumentation by using De Finetti's theory of subjective probabilities [10] rather than Kolmogorov's axiomatization. Moreover they preliminarily investigated the extension of the approach to the case of imprecise probabilities [71].

The variety of flavours in probabilistic argumentation presented in the above cited papers suggests that the interpretation of the probability of an argument is, in an epistemic sense, a largely open issue and that a finer classification could be devised, where the generic notion of epistemic probability considered up to now in the literature is replaced by a more detailed taxonomy. While this interesting direction of investigation is beyond the scope of this paper, we remark that the above mentioned papers typically resort to an intuitive explanation of the intended meaning of epistemic probability and assume as a starting point that the epistemic probability is given, leaving implicit the underlying probability space. As discussed in Section 4, our formalism is able to capture, at a formal level, an assignment of epistemic probabilities to arguments as a special case of probabilistic $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$-labellings based on a well-identified probability space. In the same formal setting, an epistemic probability assignment corresponding to a generic probabilistic $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}\}$-labelling is identified, showing that we recover, in our context, some of the desirable technical properties introduced in other papers. In this sense, we do not commit to any specific intuitive interpretation of epistemic probabilities but rather provide a reference formal framework which can be employed for further analyses of this notion not only at the abstract level, on which the above cited papers are focused, but also concerning its origin at the structured level and its impact on the evaluation of argument conclusions.

## 5.3 Other connections between argumentation and probability

While the approach proposed in this paper and those previously overviewed focus on modelling probabilistic uncertainty in the context of the various steps of formal argumentative reasoning, other kinds of connections and interactions between argumentation and probabilistic notions have been conceived; we discuss them in the following.

At a foundational level, some works use basic probabilistic notions to give an interpretation of defeasible rules in argumentative reasoning. In particular Pollock [45] uses the notion of statistical syllogism to define *prima facie reasons*. Briefly, in the simplest form, if $F$ holds and the conditional probability that $G$ holds given that $F$ holds is above a given threshold (typically $0.5$) then there is a prima facie reason to (defeasibly) infer $G$. In this view, basic probability concepts (in particular the notion of conditioning) are, in a sense, regarded as more fundamental than argumentative notions which are a sort of derived concept, at least from an epistemological point of view. As discussed by Verheij in [67], Pollock however did criticise other aspects of probabilistic reasoning leading to a sort of 'anti-probabilistic stance'. To reconcile Pollock's foundational perspective with standard probability laws, Verheij proposes a reformulation of the notion of prima facie reason, and shows how it addresses some of Pollock's criticisms. In the same foundational strain, Verheij [66] proposes to combine basic notions from nonmonotonic reasoning and probability theories to formalise ampliative arguments, namely defeasible arguments going 'beyond their premises'. Again, a standpoint is assumed that argumentative notions can be interpreted in terms of more fundamental (or more primitive) concepts, probability being one of them. Discussing foundational issues is beyond the scope of the present paper, which deals with combining probability and argumentation at a different level: independently of the possible interpretation of defeasible rules in probabilistic terms, there can be uncertainty about which rules are adopted, which arguments are built, which of the built arguments are accepted and so on. In this sense, our study of probabilistic uncertainty inside the process of argumentative reasoning can be understood as complementary to the study of how the defeasible rules in argumentative reasoning can be interpreted in probabilistic terms.

Connections between argumentation and one of the most popular probability-based formalisms, namely Bayesian networks, have also attracted a lot of attention in the recent years, also due to some basic similarities in the graph-based representations adopted in both areas. For instance, in [25] Wigmore charts (a graphical formalism useful for 'describing and organizing the available evidence in a case and in following reasoning processes through sequential steps') and Bayesian networks are compared by using them to model the infamous Sacco and Vanzetti case. Several pros and cons of both formalisms are discussed and then attention is focused on making the construction of Bayesian networks for complex legal cases more manageable by introducing a hierarchical object-oriented approach utilising 'small modular networks (or network fragments) as building blocks'. This idea is further pursued in [18] where in order to avoid *ad hoc* construction of Bayesian networks several basic structures, called *idioms*, for the representation of legal cases are introduced. Differently from [25], this work does not assume a specific argument representation. In the same line, but adopting a different formal background, [23] proposes a translation from the Carneades argument model into a Bayesian network providing it a probabilistic semantics.

Other approaches follow, so to say, the converse direction: they assume that a Bayesian network representing a case is given and investigate how to produce an argument-based representation for it [69, 33], under the assumption that this provides a simpler and more intuitive presentation. For example, given a Bayesian network, in [60, 61, 62] a support graph is derived whose nodes can be associated with a numerical support, based on the original probabilistic information [62]. Arguments can then be identified on the basis of the support graph [61].

Further works [68, 64] incorporate a third kind of representation commonly adopted in legal reasoning, namely *scenarios* and explore various kind of (pairwise and threewise) connections and combinations between them, probability and argumentation.

All the above cited works assume as a starting point that argumentation and probability theory provide complementary/competing formalisms for representing the same situation and, investigate, from a general knowledge representation perspective, their relative merits and disadvantages, and various options for connecting, deriving or combining these formalisms. In this context, argumentation formalisms are essentially seen as descriptive tools (as in the case of Wigmore charts) while the (numeric) assessment of conclusions is typically assigned by Bayesian networks.

Our approach is different in that it does not follow the perspective of two competing formalisms and involves a simple but full-fledged argumentation-based reasoning process where argument and conclusion assessments are carried out on the basis of argumentation semantics and the related notions of justifications. As already remarked, probability theory is used with the specific purpose of capturing various kinds of uncertainty that can be present in this process. Addressing knowledge representation issues and inter-formalism comparisons is beyond the scope of the present paper and represents an interesting direction of future work.

# 6 Conclusion

The main contribution of the paper is a labelling-oriented framework for probabilistic argumentation. The approach builds on an articulated account of an argumentation process, involving the four main stages of rule-based argument construction, argument graph definition, argument evaluation, and statement evaluation. We have considered various forms of uncertainty which can be present in these stages, casted them into a formal probabilistic setting and analysed their relationships.

Differently from other proposals in the literature employing the well-established Dung's abstract argumentation frameworks, we have adopted semi-abstract argumentation graphs where the key role of the subargument relation is taken into account to express dependences between the events of inclusion/exclusion of arguments. Moreover, uncertainty about the inclusion of arguments can be combined with uncertainty about their acceptance through the notion of $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UN}, \mathsf{OFF}\}$-labellings under suitable legality constraints. In particular, we have shown that the proposed approach can capture key concepts underlying the so-called constellations and epistemic approaches or a combination of both, and that it can ease the analysis of their properties. Some of these properties are easily retrieved by derivation in our context, while it has been evidenced that the status of other properties is more open to discussion.

Along the road, we also evidenced key correspondences between probabilistic settings at different levels. Since different models of probabilistic argumentation found in the literature are akin to these settings, we believe that these correspondences can be helpful to position novel proposals too within the whole picture of probabilistic argumentation.

At a general level, the main lessons learnt along this journey concern diversity and unification in this research field. Diversity arises from the fact that 'probabilistic argumentation' is far from being a univocal term, leading in the literature to statements like 'What is meant by the probability of an argument holding is an open question. Indeed, there seem to be various ways that we could answer this.' [27].

To account for this diversity while avoiding the risk of confusion and ambiguity, two aspects turned out to be crucial: analysing the argumentation process in its entirety, from argument construction to statement assessment, and specifying in detail the probability space (in particular the sample space) whenever some form of uncertainty arises. Analysing all phases of the argumentation process allows to situate distinct forms of uncertainty 'in the right place', and detailed specification of the probability space avoids ambiguity about their nature, since the events one is uncertain about are formally identified. Simple as they are, these specifications are most often left implicit or ambiguous in the literature. We believe therefore that both aspects provide a contribution to the need of conceptual clarity in the field. In particular, they are essential to carry out an analysis of the relationships and possible dependences between the different forms of uncertainty, which is only partially doable when they are held in isolation or are not clearly specified. An example of this kind of analysis is the distinction between the probability of a set of arguments being an extension as usually conceived in the constellations approach and the probability of a labelling in our probabilistic labelling frames. While, at a superficial level, they could be seen as similar, we have shown that the former can be regarded as a derived notion in the context of probabilistic graph frames, while the latter refers to a completely distinct form of uncertainty in the context of probabilistic labelling frames. Suggestions for a critical reappraisal of some concepts adopted in non-probabilistic argumentation are another byproduct of this analysis, as exemplified by the discussion about probabilistic argument justification labellings.

While accounting for a diversity of uncertainties is essential, it has also to be acknowledged that they may be present together, leading to the unification problem: how to combine them within a formal framework, able to generalise some existing approaches and to provide a uniform treatment of diverse uncertainties. Probabilistic labellings turned out to be a suitable tool to this purpose.

Future directions are multiple. We may lay down compact representations through probabilistic rules or factors, paving the way to possible integration of probabilistic argumentation with well-established probabilistic graphical models and machine learning techniques to address the parametrisation and valuation of such models. In addition, we mentioned possible use of non-classical approaches to probability (e.g. De Finetti's subjective probabilities) in this context as an alternative to the classical definition of probability function based on Kolmogorov axioms we adopted here. Moreover, the argumentation process in this paper is rule-based: extending our analysis to other forms of argumentation (e.g. dialogical) is another direction of future work. Finally, our investigation is entirely formal and independent from any particular domain. However, some domains may require to capture some particular features. For example, as mentioned in Section 5, a variety of research proposed a combination of different argumentation and probabilistic models to address diverse aspects of uncertainty in legal reasoning. More generally, it would be interesting to see how probabilistic labellings could better help to characterise diverse aspects of uncertainty in specific application domains.

# Acknowledgement

Yang Gao was supported by National Natural Science Foundation of China (NSFC) grant 61602453. Régis Riveret was supported by the Marie Curie Intra-European Fellowship PIEF-GA-2012-331472. Antonino Rotolo was supported by the European Unions Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No 690974 for the project

MIREL: MIning and REasoning with Legal texts. The final publication is available at Springer via http://dx.doi.org/10.1007/s10472-018-9574-1

# Notation

Some key notations used in the paper.

| | |
|---|---|
| $T$ | A defeasible theory |
| $G$ | An argumentation graph |
| $\mathsf{ArgLab}$ | A set of labels for arguments |
| $l$ | A label for arguments |
| L | A labelling function of arguments |
| $L_A$ | The random labelling of an argument $A$ |
| **L** | A set of random labellings of arguments |
| $\mathsf{LitLabels}$ | A set of labels for literals |
| $k$ | A label for literals |
| K | A labelling function of literals |
| $K_\varphi$ | The random labelling of literal $\varphi$ |
| **K** | A set of random labellings of literals |
| $\mathcal{L}_{\mathsf{ArgLab}}(G)$ | The set of ArgLab-labellings of the argumentation graph $G$ |
| $\mathcal{S}$ | An $X$-ArgLab-labelling specification |
| $\mathcal{L}^X_{\mathsf{ArgLab}}(G)$ | The set of $X$-ArgLab-labellings of the argumentation graph $G$, identified by the $X$-ArgLab-labelling specification |
| PTF | A probabilistic theory frame |
| PGF | A probabilistic graph frame |
| PLF | A probabilistic labelling frame |
| PAG | A probabilistic argumentation graph (from [27]) |
| PEF | A probabilistic epistemic frame |

# References


[1] Atkinson, K., Baroni, P., Giacomin, M., Hunter, A., Prakken, H., Reed, C., Simari, G.R., Thimm, M., Villata, S.: Towards artificial argumentation. AI Magazine **38**(3), 25–36 (2017)

[2] Baroni, P., Caminada, M., Giacomin, M.: An introduction to argumentation semantics. Knowledge Engineering Review **26**(4), 365–410 (2011)

[3] Baroni, P., Cerutti, F., Giacomin, M., Guida, G.: AFRA: argumentation framework with recursive attacks. Int. J. Approx. Reasoning **52**(1), 19–37 (2011)

[4] Baroni, P., Giacomin, M., Vicig, P.: On rationality conditions for epistemic probabilities in abstract argumentation. In: Proc. of the 5th Inf. Conf. on Computational Models of Argument (COMMA 2014), *Frontiers in Artificial Intelligence and Applications*, vol. 266, pp. 121–132. IOS Press (2014)

[5] Baroni, P., Governatori, G., Riveret, R.: On labelling statements in multi-labelling argumentation. In: Proc. of the 22nd European Conf. on Artificial Intelligence (ECAI 2016), *Frontiers in Artificial Intelligence and Applications*, vol. 285, pp. 489–497. IOS Press (2016)

[6] Bench-Capon, T., Dunne, P.E.: Argumentation in artificial intelligence. Artificial Intelligence **171**(10-15), 619 – 641 (2007)

[7] Bondarenko, A., Dung, P.M., Kowalski, R.A., Toni, F.: An abstract, argumentation-theoretic approach to default reasoning. Artificial Intellgence **93**, 63–101 (1997)

[8] Cayrol, C., Lagasquie-Schiex, M.: Bipolarity in argumentation graphs: Towards a better understanding. Int. J. Approx. Reasoning **54**(7), 876–899 (2013)

[9] Cohen, A., Gottifredi, S., García, A.J., Simari, G.R.: An approach to abstract argumentation with recursive attack and support. J. Applied Logic **13**(4), 509–533 (2015)

[10] De Finetti, B.: Theory of probability: a critical introductory treatment. Wiley (1974)

[11] Dondio, P.: Toward a computational analysis of probabilistic argumentation frameworks. Cybernetics and Systems **45**(3), 254–278 (2014)

[12] Dung, P.M.: 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)

[13] Dung, P.M., Thang, P.M.: Towards (probabilistic) argumentation for jury-based dispute resolution. In: Proc. of the 3rd Int. Conf. on Computational Models of Argument (COMMA 2010), *Frontiers in Artificial Intelligence and Applications*, vol. 216, pp. 171–182. IOS Press (2010)

[14] Dung, P.M., Thang, P.M.: Closure and consistency in logic-associated argumentation. J. Artificial Inteligence Research **49**(1), 79–109 (2014)

[15] Fazzinga, B., Flesca, S., Parisi, F.: On the complexity of probabilistic abstract argumentation. In: Proc. of the 23rd Int. Joint Conf. on Artificial Intelligence, IJCAI 2013, pp. 898–904. AAAI Press (2013)

[16] Fazzinga, B., Flesca, S., Parisi, F.: On the complexity of probabilistic abstract argumentation frameworks. ACM Trans. Comput. Log. **16**(3), 22 (2015)

[17] Fazzinga, B., Flesca, S., Parisi, F.: On efficiently estimating the probability of extensions in abstract argumentation frameworks. Int. J. Approx. Reasoning **69**, 106–132 (2016)

[18] Fenton, N.E., Neil, M., Lagnado, D.A.: A general structure for legal arguments about evidence using bayesian networks. Cognitive Science **37**(1), 61–102 (2013)

[19] Fruhwirth, T.: Constraint Handling Rules, 1st edn. Cambridge University Press (2009)

[20] Gabbay, D.M., Rodrigues, O.: Probabilistic argumentation: An equational approach. Logica Universalis **9**(3), 345–382 (2015)

[21] Garcia, A.J., Simari, G.R.: Defeasible logic programming: DeLP servers, contextual queries, and explanations for answers. Argument & Computation **5**(1), 63 – 88 (2014)

[22] Governatori, G., Maher, M.J., Antoniou, G., Billington, D.: Argumentation semantics for defeasible logic. J. Log. Comput. **14**(5), 675–702 (2004)

[23] Grabmair, M., Gordon, T.F., Walton, D.: Probabilistic semantics for the Carneades argument model using bayesian networks. In: Proc. of the 3rd Int. Conf. on Computational Models of Argument (COMMA 2010), *Frontiers in Artificial Intelligence and Applications*, vol. 216, pp. 255–266. IOS Press (2010)

[24] Haenni, R.: Probabilistic argumentation. J. Applied Logic **7**(2), 155–176 (2009)

[25] Hepler, A.B., Dawid, P., Leucari, V.: Object-oriented graphical representations of complex patterns of evidence. Law, Probability & Risk **6**(1-4), 275–293 (2007)

[26] Hinton, A., Kwiatkowska, M., Norman, G., Parker, D.: PRISM: A tool for automatic verification of probabilistic systems. In: Proc. of the 12th Int. Conf. on Tools and Algorithms for the Construction and Analysis of Systems (TACAS 2006), *LNCS*, vol. 3920, pp. 441–444. Springer-Verlag (2006)

[27] Hunter, A.: A probabilistic approach to modelling uncertain logical arguments. Int. J. Approx. Reasoning **54**(1), 47–81 (2013)

[28] Hunter, A., Thimm, M.: Probabilistic argumentation with epistemic extensions. In: Proc. of the Int. Workshop on Defeasible and Ampliative Reasoning, vol. 1212. CEUR-WS.org (2014)

[29] Hunter, A., Thimm, M.: Probabilistic argumentation with epistemic extensions and incomplete information. CoRR **abs/1405.3376** (2014)

[30] Hunter, A., Thimm., M.: On partial information and contradictions in probabilistic abstract argumentation. In: Proc. of the 15th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR 2016), pp. 53–62 (2016)

[31] Jakobovits, H., Vermeir, D.: Robust semantics for argumentation frameworks. J. Log. Comput. **9**(2), 215–261 (1999)

[32] Jech, T.: Set Theory, 3rd millennium edition, revised and expanded edn. Springer (2003)

[33] Keppens, J.: Argument diagram extraction from bayesian networks. Artificial Intelligence and Law **20**(2), 109–143 (2012)

[34] Lam, H.P., Governatori, G., Riveret, R.: On ASPIC+ and Defeasible Logic. In: Proc. of the 6th Int. Conf. on Computational Models of Argument (COMMA 2016), *Frontiers in Artificial Intelligence and Applications*, vol. 287, pp. 359–370. IOS Press (2016)

[35] Li, H.: Probabilistic argumentation. Ph.D. thesis, Department of Computing Science, University of Aberdeen (2015)

[36] Li, H., Oren, N., Norman, T.J.: Probabilistic argumentation frameworks. In: Revised Selected Papers of the 1st Int. Workshop on Theory and Applications of Formal Argumentation (TAFA 2011), *LNAI*, vol. 7132, pp. 1–16. Springer (2011)

[37] Li, H., Oren, N., Norman, T.J.: Relaxing independence assumptions in probabilistic argumentation. In: Proc. of the 10th Int. Workshop on Argumentation in Multi-Agent Systems (2013)

[38] Liao, B., Huang, H.: Formulating semantics of probabilistic argumentation by characterizing subgraphs. In: Proc. of the 5th Int. Workshop on Logic, Rationality, and Interaction (LORI 2015), *LNCS*, vol. 9394, pp. 243–254. Springer (2015)

[39] Liao, B., Xu, K., Huang, H.: Formulating semantics of probabilistic argumentation by characterizing subgraphs: Theory and empirical results. CoRR **abs/1608.00302** (2016)

[40] Modgil, S., Prakken, H.: The *ASPIC*$^+$ framework for structured argumentation: a tutorial. Argument & Computation **5**(1), 31–62 (2014)

[41] Nute, D.: Defeasible logic. In: Handbook of Logic in Artificial Intelligence and Logic Programming, pp. 353–395. Oxford University Press (2001)

[42] Oren, N., Norman, T.J.: Semantics for evidence-based argumentation. In: Proc. of the 2nd Int. Conf. on Computational Models of Argument (COMMA 2008), *Frontiers in Artificial Intelligence and Applications*, vol. 172, pp. 276–284. IOS Press (2008)

[43] Polberg, S., Oren, N.: Revisiting support in abstract argumentation systems. In: Proc. of the 5th Int. Conf. on Computational Models of Argument (COMMA 2014), *Frontiers in Artificial Intelligence and Applications*, vol. 266, pp. 369–376. IOS Press (2014)

[44] Pollock, J.L.: Justification and defeat. Artificial Intelligence **67**(2), 377–407 (1994)

[45] Pollock, J.L.: Cognitive Carpentry: A Blueprint for How to Build a Person. MIT Press (1995)

[46] Prakken, H.: On support relations in abstract argumentation as abstractions of inferential relations. In: Proc. of the 21st European Conf. on Artificial Intelligence (ECAI 2014), *Frontiers in Artificial Intelligence and Applications*, vol. 263, pp. 735–740. IOS Press (2014)

[47] Prakken, H., Sartor, G.: Argument-based extended logic programming with defeasible priorities. Journal of Applied Non-Classical Logics **7**(1-2), 25–75 (1997)

[48] Rienstra, T.: Towards a probabilistic Dung-style argumentation system. In: Proc. of the 1st Int. Conf. on Agreement Technologies, vol. 918, pp. 138–152. CEUR (2012)

[49] Riveret, R., Governatori, G.: On learning attacks in probabilistic abstract argumentation. In: Proc. of the 15th Int. Conf. on Autonomous Agents & Multiagent Systems (AAMAS 2016), pp. 653–661. International Foundation for Autonomous Agents and Multiagent Systems (2016)

[50] Riveret, R., Korkinof, D., Draief, M., Pitt, J.V.: Probabilistic abstract argumentation: an investigation with Boltzmann machines. Argument & Computation **6**(2), 178–218 (2015)

[51] Riveret, R., Pitt, J.V., Korkinof, D., Draief, M.: Neuro-symbolic agents: Boltzmann machines and probabilistic abstract argumentation with sub-arguments. In: Proc. of the14th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2015), pp. 1481–1489. International Foundation for Autonomous Agents and Multiagent Systems (2015)

[52] Riveret, R., Rotolo, A., Sartor, G.: Probabilistic rule-based argumentation for norm-governed learning agents. Artificial Intelligence and Law **20**(4), 383–420 (2012)

[53] Riveret, R., Rotolo, A., Sartor, G., Prakken, H., Roth, B.: Success chances in argument games: a probabilistic approach to legal disputes. In: Proc. of the 20th Conf. on Legal Knowledge and Information Systems (JURIX 2007), *Frontiers in Artificial Intelligence and Applications*, vol. 165, pp. 99–108. IOS Press (2007)

[54] Roth, B., Riveret, R., Rotolo, A., Governatori, G.: Strategic argumentation: a game theoretical investigation. In: Proc. of the 11th Int. Conf. on Artificial Intelligence and Law (ICAIL 07), pp. 81–90. ACM (2007)

[55] Sato, T.: A glimpse of symbolic-statistical modeling by PRISM. Journal of Intelligent Information Systems **31**(2), 161–176 (2008)

[56] Sneyers, J., Meert, W., Vennekens, J., Kameya, Y., Sato, T.: Chr(prism)-based probabilistic logic learning. CoRR **abs/1007.3858** (2010)

[57] Sneyers, J., Schreye, D.D., Fruhwirth, T.: Probabilistic legal reasoning in CHRiSM. Theory and Practice of Logic Programming **13**, 769–781 (2013)

[58] Tang, Y., Oren, N., Sycara, K.P.: Markov argumentation random fields. In: Proc. of the 30th AAAI Conference on Artificial Intelligence (AAAI 2016), pp. 4397–4398. AAAI Press (2016)

[59] Thimm, M.: A probabilistic semantics for abstract argumentation. In: Proc. of the 20th European Conf. on Artificial Intelligence (ECAI 2012), *Frontiers in Artificial Intelligence and Applications*, vol. 242, pp. 750–755. IOS Press (2012)

[60] Timmer, S.T., Meyer, J.J.C., Prakken, H., Renooij, S., Verheij, B.: Extracting legal arguments from forensic Bayesian networks. In: Proc. of the 27th Conf. on Legal Knowledge and Information Systems (JURIX 2014), *Frontiers in Artificial Intelligence and Applications*, vol. 271, pp. 71–80. IOS Press (2014)

[61] Timmer, S.T., Meyer, J.J.C., Prakken, H., Renooij, S., Verheij, B.: Explaining Bayesian networks using argumentation. In: Proc. of the 13th European Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2015), *LNAI*, vol. 9161, pp. 83–92. Springer (2015)

[62] Timmer, S.T., Meyer, J.J.C., Prakken, H., Renooij, S., Verheij, B.: A structure-guided approach to capturing Bayesian reasoning about legal evidence in argumentation. In: Proc. of the 15th Int. Conf. on Artificial Intelligence and Law (ICAIL 2015), pp. 109–118 (2015)

[63] Toni, F.: A tutorial on assumption-based argumentation. Argument & Computation **5**(1), 89 – 117 (2014)

[64] Verheij, B.,  , F., Timmer, S.T., Vlek, C.S., Meyer, J.J.C., Renooij, S., Prakken, H.: Arguments, scenarios and probabilities: connections between three normative frameworks for evidential reasoning. Law, Probability & Risk **15**(1), 35–70 (2015)

[65] Verheij, B.: Two approaches to dialectical argumentation: Admissible sets and argumentation stages. In: Proc. of the Int. Conf. on Formal and Applied Practical Reasoning (FAPR 1996), *LNAI*, vol. 1085, pp. 357–368. Springer (1996)

[66] Verheij, B.: Jumping to conclusions - A logico-probabilistic foundation for defeasible rule-based arguments. In: Proc. of the 13th European Conf. on Logics in Artificial Intelligence (JELIA 2012), *LNCS*, vol. 7519, pp. 411–423. Springer (2012)

[67] Verheij, B.: Arguments and their strength: Revisiting Pollock's anti-probabilistic starting points. In: Proc. of the 5th Int. Conf. on Computational Models of Argument (COMMA 2014), *Frontiers in Artificial Intelligence and Applications*, vol. 266, pp. 433–444. IOS Press (2014)

[68] Verheij, B.: To catch a thief with and without numbers: arguments, scenarios and probabilities in evidential reasoning. Law, Probability & Risk **13**(3-4), 307–325 (2014)

[69] Vreeswijk, G.A.W.: Argumentation in Bayesian belief networks. In: Revised Selected and Invited Papers of the 1st Int.Workshop on Argumentation in Multi-Agent Systems (ARGMAS 2004), *LNCS*, vol. 3366, pp. 111–129. Springer (2005)

[70] Vreeswijk, G.A.W.: An algorithm to compute minimally grounded and admissible defence sets in argument systems. In: Proc. of the 1st Int. Conf. on Computational Models of Argument (COMMA 2006), *Frontiers in Artificial Intelligence and Applications*, vol. 144, pp. 109–120. IOS Press (2006)

[71] Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall (1991)