A Probabilistic Deontic Argumentation Framework

Régis Riveret

Commonwealth Scientific and Industrial Research Organisation, Australia

Nir Oren

University of Aberdeen, United Kingdom

Giovanni Sartor

University of Bologna, Italy

Abstract

What does it mean that something is probably obligatory? And how does it relate to the probability that it is permitted or prohibited? In this paper, we provide a possible answer by merging deontic argumentation and probabilistic argumentation into a probabilistic deontic argumentation framework. This framework allows us to specify a semantics for the probability of deontic statuses. The deontic argumentation part builds on standard concepts from the study of computational models of argument: rule-based arguments, argumentation graphs, argument labelling semantics and statement labelling semantics. We then encapsulate this deontic composition with the approach of probabilistic labellings to probabilistic argumentation, in order to associate deontic statements with probability values. The framework is illustrated with a scenario featuring a violation and a contrary-to-duty obligation.

Keywords: Probabilistic argumentation, deontic argumentation.

1. Introduction

Research in the area of argumentation investigates how conclusions can be drawn from a set of conflicting arguments and how such arguments can be constructed from an inconsistent and incomplete knowledge base. Argumentation has become a well-established approach to defeasible reasoning by virtue of its perceived simplicity and explanatory power. Because of its logic-based underpinnings, argumentation theory has traditionally approached uncertainty from a qualitative perspective. However, recent work has begun integrating probabilistic quantitative aspects into the argumentation process, see e.g. [1, 2, 3]. Typically, arguments are associated with probability values, and quantitative relationships amongst these arguments are studied to add a probabilistic dimension to defeasible reasoning as addressed by argumentation.

It has long been recognised that normative and deontic reasoning is defeasible [4] or non-monotonic [5]. Capturing such reasoning within an argumentative context could – potentially – provide an enhanced understanding of it and foster useful applications, and consequently,

normative or deontic argumentation frameworks have attracted increasing attention, see e.g. [6, 7, 8, 9].

Integrating probabilistic uncertainty into deontic argument-based reasoning may be useful for various applications. For example, a probabilistic deontic argumentation framework may provide semantics for legal expert systems associating normative effects with probability measures. It may also be useful to norm-governed cyber-physical systems where sensors provide uncertain readings as to the state of the environment, and probability measures are attached to system's normative states. Multiple applications can be conceived.

In general, probabilistic reasoning upon deontic concepts is important for systems where norm-related statements have an associated probabilistic uncertainty. As a running example, we will consider the following simple scenario (a follow-up from [9], inspired by the Hart-Fuller debate [10, 11]).

Example 1. *Vehicles are prohibited from entering a park unless there is an emergency. At some point, sensors suggest that a vehicle has entered the park, while other sensors indicate that an emergency situation may be occurring. Given this situation, we then seek to compute the likelihood that the prohibition is violated; doing so would allow us — for example — to decide whether to dispatch a law enforcement agent to the park or not.*

In the example above, probabilities are associated with the facts matching the antecedents of conditional norms, a prohibition and an exception to it, and one is interested in the probability that the prohibition is complied with or violated. In turn, such measures can then be used to obtain descriptive or predictive insights into the system. If no proper probabilistic framework is adopted, then outcomes and uncertainty measures may be inconsistent or incoherent.

In this paper we take a first step in combining probability theory and deontic argumentation, investigating how probabilistic argumentation can be coupled with normative and deontic reasoning. Our approach builds on the notion of probabilistic labellings [3] and we show how such labellings can be applied to the types of deontic argumentation frameworks described in [9] where pieces of doctrine are expressed within defeasible theories.

The combination of deontic and probabilistic notions is certainly not new, and there exist various works where probabilistic concepts are used to capture different aspects of normative notions. In the utilitarian tradition, for example, an action is considered obligatory if it has the highest expected aggregate utility when compared to the available alternatives, and this view has influenced fields as diverse as ethics, law, and economics. Our focus here is different: we concentrate on normative conclusions, i.e. on the normative statuses (unconditional obligations, permissions, or other normative properties) that are established by a given set of norms. We seek to determine how likely it is that such normative statuses are triggered by facts matching the antecedents of conditional norms. With regard to obligations or prohibitions we also examine how likely it is that they are violated or rather complied with.

In this work we ignore dynamic and practical aspects of normative reasoning, considering only how uncertain knowledge impacts on normative concepts. In other words, we consider only epistemic issues, ignoring, for example, how to select specific actions to pursue by utilising probabilistic argumentation [12]. We believe that determining the likelihood of normative statuses is an important subject on its own, and intend to investigate the links with — for example — practical decision making as part of future work.

Our work builds on probabilistic argumentation, which has been applied to several aspects of legal and normative domains. For example, probability theory has been combined with

argument and scenario approaches to evidential reasoning [1]. In another line of work, early probabilistic argumentation frameworks (see e.g. [13]) were used together with game theory to identify optimal strategies in dialogue games [14]. More recent research has coupled probabilistic argumentation with reinforcement learning to determine what actions to perform in environments featuring basic conditional obligations [12]. However, to the best of our knowledge, probabilistic argumentation has never been integrated with any proper deontic argumentation framework, and it is precisely this gap we seek to address in the work reported here.

The remainder of the paper is structured as follows. In the next section we outline a labelling-based approach to probabilistic argumentation and describe its semantics. In Section 3 we describe how probabilistic labellings can be used to reason about deontic concepts. Section 4 provides a detailed example of our approach, after which we conclude.

2. Probabilistic Argumentation Framework

In this section we present a simple ASPIC-like argumentation system (cf. e.g. [15, 16]) and its probabilistic development [3]. We begin by defining the language for expressing rules, and then describe how arguments, attack and support relations can be constructed. Given a set of arguments, we identify which are accepted, and in turn which statements are acceptable. Acceptance is specified through labellings, which are then integrated into a probabilistic setting.

2.1. Language

Our language is propositional, with literals being constructed from propositions, negations of propositions, and modal operators (which are used later to capture deontic modalities).

Definition 2.1. *A **literal statement** is either a plain literal statement or a modal literal statement, where*

- *a **plain literal statement** is either an atomic proposition p or the negation of an atomic proposition (i.e. $\neg p$) and*
- *a **modal literal statement** is a statement of the form $\Box\gamma$ or $\neg\Box\gamma$, such that $\Box$ is a placeholder for any modal operator and γ is a plain literal statement.*

Notation 2.1. *For notational convenience we write the complement of a literal statement γ as $\overline{\gamma}$. That is, if γ is of the form p, then $\overline{\gamma}$ is $\neg p$, while if γ is $\neg p$, then $\overline{\gamma}$ is p.*

Defeasible rules constructed using literals represent conditionals of the form '*if* ... *then* ... *unless* ...'. While many argumentation formalisms distinguish between strict and defeasible rules, for simplicity we consider only defeasible rules in this work[1].

Definition 2.2. *A **defeasible rule** over a set of literal statements $\mathscr{S}$ is a construct of the form:*

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

[1] The translation of our results to strict rules can be achieved relatively simply through techniques similar to those proposed by Li and Parsons [17].

where $0 \leq n$ and $0 \leq m$, r is a unique identifier for the rule, while for any $0 \leq i \leq n$, $0 \leq j \leq m$, $\varphi_i, \varphi'_j, \varphi \in \mathscr{S}$ are all literal statements.

A rule $r : \varphi_1, \ldots, \varphi_n, \sim \varphi'_1, \ldots, \sim \varphi'_m \Rightarrow \varphi$ can be read as follows: "*if* φ_1 *and* ... *and* φ_n *are supported*, and all of φ'_1, ..., φ'_m are not supported, then φ is *defeasibly supported*. While premises prefixed with $\sim$ can be interpreted as a form of negation as failure, we view such premises as exceptions to the (default) application of the rule.

Notation 2.2. *Let r be a defeasible rule r as in Definition 2.2, and Rules a set of rules.*
- $\text{Body}(r)$ *denotes the body of r, i.e.* $\text{Body}(r) = \{\varphi_1, \ldots, \varphi_n, \sim \varphi'_1, \ldots, \sim \varphi'_m\}$.
- $\text{Head}(r)$ *denotes the head of r, i.e.* $\text{Head}(r) = \{\varphi\}$.
- $\text{Prop}(r)$ *denotes the set of propositions of r, i.e.* $\text{Prop}(r) = \{p \mid p, \neg p, \sim p, \sim \neg p, \Box p, \neg\Box p, \Box\neg p, \neg\Box\neg p, \sim \Box p, \sim \neg\Box p, \sim \Box\neg p, \sim \neg\Box\neg p \in \text{Body}(r) \cup \text{Head}(r)\}$.
- $\text{Prop}(\mathit{Rules})$ *denotes the set of propositions of Rules, i.e.* $\text{Prop}(\mathit{Rules}) = \bigcup_{r \in \mathit{Rules}} \text{Prop}(r)$.

Certain sets of literals are mutually inconsistent. In other words, they conflict, or are incompatible with each other. Perhaps the simplest form of such a conflict is between a proposition and its negation, but more complex conflicts can also exist. As done in many other systems [18, 15], we introduce an abstract conflict relation to encode incompatibilities between literal statements.

Definition 2.3. *A **conflict relation** 'Conflicts' over a set of literal statements $\mathscr{S}$ is a binary relation over $\mathscr{S}$, i.e. Conflicts $\subseteq \mathscr{S} \times \mathscr{S}$.*

Notation 2.3. *We write* $\text{Prop}(\mathit{Conflicts})$ *to denote the propositions found within a conflict relation, i.e.* $\text{Prop}(\mathit{Conflicts}) = \{p \mid (\varphi, \varphi') \in \mathit{Conflicts} : \varphi = p, \neg p, \Box p, \neg\Box p, \Box\neg p, \neg\Box\neg p,$ *or* $\varphi' = p, \neg p, \Box p, \neg\Box p, \Box\neg p, \neg\Box\neg p\}$.

The conflict relation should adhere to certain principles, in which case it is said to be *well-formed*. For example, for a purely propositional system, it should be the case that $\mathit{Conflicts}(\gamma, \overline{\gamma})$. For a deontic setting, we will specify the requirements for the conflict relation in Section 3. Note that the conflict relation can be asymmetric or symmetric to capture, for example, contrary or contradictory relationships [15], but we do not consider such aspects here.

Preferences over rules are commonly used to resolve conflicts. We capture such preferences via a *superiority* relation $\succ$ over rules. Informally, $s \succ r$ means that rule s prevails over rule r.

Definition 2.4. *A **superiority relation** $\succ$ over a set of rules Rules is an asymmetric binary relation over Rules, i.e.* $\succ \subseteq \mathit{Rules} \times \mathit{Rules}$.

As the superiority relation is asymmetric, for any rule r it is not the case that $r \succ r$, and for two distinct rules r and r' we cannot have both $r \succ r'$ and $r' \succ r$.

2.2. Defeasible theories and argumentation graphs

Having introduced rules, conflicts between literal statements and superiority relationships between rules, we now turn our attention to how rules can be combined to build arguments. Given that these arguments may conflict, we describe later how sets of accepted arguments can be identified.

A defeasible theory lists a set of rules, a conflict relation and a superiority relation. Such a defeasible theory forms the basic structure over which inferences can take place.

Definition 2.5. *A **defeasible theory** is a tuple $\langle Rules, Conflicts, \succ \rangle$ where*
- *Rules is a set of rules, and*
- *Conflicts is a conflict relation, and*
- *$\succ$ is a superiority relation over Rules.*

Notation 2.4. *Given a defeasible theory $T = \langle Rules, Conflicts, \succ \rangle$,*
- $\text{Rules}(T)$, $\text{Conflicts}(T)$, *and* $\succ(T)$ *denote the set of rules Rules, the relation Conflicts, and the relation $\succ$ of theory T respectively,*
- $\text{Prop}(T)$ *denotes the set of propositions of T, i.e.* $\text{Prop}(T) = \text{Prop}(Rules) \cup \text{Prop}(Conflicts)$.

Inferences over a defeasible theory are performed by chaining its defeasible rules to form arguments, which we define below. Our definition of arguments is inspired by, and similar to, those used in other rule-based argumentation frameworks, as described for example in [15, 16, 19].

Definition 2.6. *An **argument** A constructed from a defeasible theory $\langle Rules, Conflicts, \succ \rangle$ is a finite construct of the form:*

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

with $0 \leq n$ and $0 \leq m$, and where
- *A is the unique identifier of the argument;*
- *φ is the* conclusion *of the argument, denoted* $\text{con}(A)$*;*
- *$A_1, \ldots, A_n$ are arguments constructed from the defeasible theory $\langle Rules, Conflicts, \succ \rangle$;*
- *$r \in Rules$ is the* top rule *of the argument, and it is of the form $r : \text{con}(A_1), \ldots, \text{con}(A_n), \sim \varphi_1, \ldots, \sim \varphi_m \Rightarrow \varphi$.*

Notation 2.5. *Given an argument A as in Definition 2.6, we use the following notations.*
- $\text{Sub}(A)$ *denotes the set of subarguments of A, i.e.* $\text{Sub}(A) = \text{Sub}(A_1) \cup \ldots \cup \text{Sub}(A_n) \cup \{A\}$.
- $\text{DirectSub}(A)$ *denotes the direct subarguments of A, i.e.* $\text{DirectSub}(A) = \{A_1, \ldots, A_n\}$.
- $\text{TopRule}(A)$ *denotes the top rule of A, i.e.* $\text{TopRule}(A) = (r : \text{con}(A_1), \ldots, \text{con}(A_n), \sim \varphi_1, \ldots, \sim \varphi_m \Rightarrow \varphi)$.
- $\text{Rules}(A)$ *denotes the rules used within A, i.e.* $\text{Rules}(A) = \text{Rules}(A_1) \cup \ldots \cup \text{Rules}(A_n) \cup \{\text{TopRule}(A)\}$.

We can remark that Definition 2.6 assumes that arguments are finite, i.e. any argument has a finite set of subarguments. Accordingly, any argument has exactly one conclusion and 'bottoms out' in arguments with an empy set of subarguments (taking the form $A : \sim \varphi_1, \ldots, \sim \varphi_m \Rightarrow \varphi$ with $0 \leq m$). Nevertheless, for a given defeasible theory, we may have an infinite number of finite arguments.

Different types of inconsistencies can appear between arguments, causing them to *attack* each other. We consider two types of attacks between arguments: *rebuttals* (where two argument's conclusions are incompatible), and *undercuts*[2] (where exceptions prevent an argument's conclusion from being drawn). In the ASPIC family of argumentation frameworks, attack is differentiated from *defeat*, with the latter taking preferences between arguments into account. For simplicity, we make no such distinction in this work, instead integrating such

[2]Note that this term is overloaded in the argumentation literature, and is used with different meanings in different contexts [15].

preferences into our definition of attack. While diverse approaches to lifting preferences over rules to preferences over arguments have been described [6], in this work we utilise a simple last-link ordering [15] to compute preferences over arguments. More specifically, argument A is preferred to argument B (written $A \succ B$) iff $\text{TopRule}(A) \succ \text{TopRule}(B)$.

Definition 2.7.

- *An **attack relation** $\rightsquigarrow$ over a set of arguments $\mathscr{A}$ is a binary relation over $\mathscr{A}$, i.e. $\rightsquigarrow \subseteq \mathscr{A} \times \mathscr{A}$.*
- *Let A and B be two arguments constructed from a defeasible theory $\langle Rules, Conflicts, \succ \rangle$, argument A attacks B, i.e. $(A,B) \in \rightsquigarrow$, iff A rebuts or undercuts B, where*
  - *A rebuts B (on B') iff $\exists B' \in \text{Sub}(B)$ such that $Conflicts(\text{con}(A), \text{con}(B'))$, and $B' \not\succ A$;*
  - *A undercuts B (on B') iff $\exists B' \in \text{Sub}(B)$ such that $(\sim \text{con}(A)) \in \text{Body}(\text{TopRule}(B'))$.*

In what follows, it will be useful to identify which arguments are the immediate subarguments of other arguments, leading us to the following definition of *direct subargument relations*.

Definition 2.8.

- *A **direct subargument relation** $\Mapsto$ over a set of arguments $\mathscr{A}$ is a binary relation over $\mathscr{A}$, i.e. $\Mapsto \subseteq \mathscr{A} \times \mathscr{A}$.*
- *Let A and B be two arguments constructed from a defeasible theory, argument B is a direct subargument of A, written $B \Mapsto A$, iff B belongs to the set of direct subarguments of A, i.e. $B \in \text{DirectSub}(A)$.*

Since an argument is not a direct subargument of itself and cannot be a subargument of its direct subarguments, the direct subargument relation over arguments constructed from a defeasible theory is antireflexive and acyclic. In the remainder of this work, for the sake of simplicity, we say that an argument A supports an argument B if A is a direct subargument of B. We note that other notions of support may be considered, but leave them to future work.

Given the arguments and attacks obtained from a defeasible theory, and by making use of the subargument relation, we can create an argumentation graph as defined below, cf. [20, 21, 22, 23, 24].

Definition 2.9. *An **argumentation graph** constructed from a defeasible theory T is a tuple $\langle \mathscr{A}, \rightsquigarrow, \Mapsto \rangle$ where $\mathscr{A}$ is the set of all arguments constructed from T, $\rightsquigarrow$ is an* attack *relation over $\mathscr{A}$, and $\Mapsto$ is a* direct subargument *relation over $\mathscr{A}$.*

In the rest of the paper, we assume that all argumentation graphs are constructed from a defeasible theory, which may be left unspecified.

As to the terminology, an argumentation graph may be called a *bipolar* argumentation graph/framework [21], as long as the support relation is understood as a (direct) subargument relation, every argument has a conclusion which is a statement, and such bipolar graphs enjoy all the (probabilistic) characteristics and properties discussed in the remainder of this work.

Notation 2.6. *Given an argumentation graph $G = \langle \mathscr{A}, \rightsquigarrow, \Mapsto \rangle$, we may write $\mathscr{A}_G$, $\rightsquigarrow_G$ and $\Mapsto_G$ to denote the graph's arguments $\mathscr{A}$, attacks $\rightsquigarrow$ and subargument relation $\Mapsto$ respectively.*

Example 2. *An argumentation graph is illustrated in Figure 1, with solid arrows denoting attacks and hollow arrows representing supports. From our definitions, arguments* A *and* B *are direct subarguments of* C *and therefore provide* conjunctive support *to* C.

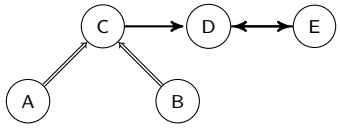

Figure 1: An argumentation graph. Arguments A and B support argument C. Argument C attacks D. Arguments D and E attack each other.

□

We note that given an argumentation graph G constructed from a defeasible theory T, since $\mathscr{A}_G$ is the set of all arguments constructed from T then the relations $\rightsquigarrow_G$ and $\mapsto_G$ of the graph are defined over $\mathscr{A}_G$. Hence $\rightsquigarrow_G$ is obtained according to Definition 2.7, and $\mapsto_G$ complies with specification of Definition 2.8 on direct subargument relations.

Given an argumentation graph, we may wish to determine what conclusions can be drawn if some arguments are omitted. We may therefore wish to deal with subgraphs of an argumentation graph.

Definition 2.10. *Let G denote an argumentation graph. The **subgraph** H of G induced by a set of arguments $\mathscr{A}_H \subseteq \mathscr{A}_G$ is an argumentation graph such that $H = \langle \mathscr{A}_H, \rightsquigarrow_G \cap (\mathscr{A}_H \times \mathscr{A}_H), \mapsto_G \cap (\mathscr{A}_H \times \mathscr{A}_H) \rangle$.*

Notation 2.7. *Given an argumentation graph G, we denote all of its subgraphs as* $\mathrm{Sub}(G)$*, i.e.* $\mathrm{Sub}(G) = \{\langle \mathscr{A}_H, \rightsquigarrow_G \cap (\mathscr{A}_H \times \mathscr{A}_H), \mapsto_G \cap (\mathscr{A}_H \times \mathscr{A}_H) \rangle \mid \mathscr{A}_H \subseteq \mathscr{A}_G\}$.

Clearly, if an argument A appears within an argumentation graph, then any subgraph should contain not only A, but also A's subarguments and any attacks related to these subarguments. Graphs obeying this property are said to be *subargument-complete*.

Definition 2.11. *A subgraph H of an argumentation graph G induced by a set of arguments $\mathscr{A}_H$ is **subargument-complete** iff for every argument $A \in \mathscr{A}_H$, if $B \mapsto_G A$ then $B \in \mathscr{A}_H$.*

Example 2 (continued)**.** *The argumentation graphs in Figures 2 (a) and (b) are subgraphs of the argumentation graph in Figure 1, and they are induced by the set of arguments* $\{\mathsf{A}, \mathsf{B}, \mathsf{C}, \mathsf{D}\}$ *and* $\{\mathsf{A}, \mathsf{C}, \mathsf{D}\}$ *respectively. The graph of Figure 2 (a) is a subargument-complete subgraph, whereas the subgraph in Figure 2 (b) is not.*

Figure 2: Two subgraphs of the argumentation graph from Figure 1.

□

Through the definitions above, given a defeasible theory we are able to build arguments and identify attacks and subarguments across arguments, from which we can construct an argumentation graph. Next, we consider how the acceptance status of arguments and statements can be determined through the use of labellings.

2.3. Argument labelling

Given an argumentation graph, the selection of arguments considered acceptable or justified is performed 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 [25], or in terms of labellings [26]. For our purposes, we adopt the labelling approach.

Argument labelling approaches usually assign one of three labels (IN, OUT and UND) to arguments, reflecting their acceptance status with respect to a specific argumentation semantics. We extend such labellings to consider whether an argument is, or is not included when evaluating acceptance by introducing an additional OFF label. The addition of this label will allow us to distinguish between the probability of using an argument as part of the *construction* of an argumentation graph, and the probability of its acceptance.

Argument labellings are commonly formalised as follows.

Definition 2.12. *Let G be an argumentation graph, and* ArgLab *a set of labels for arguments. An* ArgLab***-labelling of a set of arguments*** $\mathscr{A} \subseteq \mathscr{A}_G$ *is a total function* $\mathrm{L} : \mathscr{A} \to \mathsf{ArgLab}$.

While argument labellings map sets of arguments to a label, we will abuse notation and may speak of the 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 2.13. *Let* ArgLab *be a set of labels for arguments. An* ArgLab***-labelling of an argument graph*** *G is a total function* $\mathrm{L} : \mathscr{A}_G \to \mathsf{ArgLab}$.

Notation 2.8.
- *We write* $\mathscr{L}_{\mathsf{ArgLab}}(\mathscr{A})$ *to denote the universe of all possible* ArgLab*-labelling assignments to a set of arguments* $\mathscr{A}$*, and we write* $\mathscr{L}_{\mathsf{ArgLab}}(G)$ *for the set of all possible* ArgLab*-labellings of an argumentation graph G.*
- *If a labelling* L *assigns the label l to a set of arguments A, we may write* $l(\mathrm{L})$*. In other words,* $l(\mathrm{L}) = \{A \mid L(A) = l\}$*. Thus for example, if* IN *is a label, then* $\mathsf{IN}(\mathrm{L}) = \{A \mid \mathrm{L}(A) = \mathsf{IN}\}$.

As mentioned above, standard argument labellings make use of $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$-labellings. Here, each argument is associated with a label representing its status with regards to an argumentation semantics [26]. Intuitively, an IN labelling means that the argument is accepted, while an OUT labelling indicates its rejection. An UND labelling states that the status of an argument is undecided, i.e. it is neither accepted nor rejected. Different labelling functions may label arguments in different ways, and a single labelling function may label the same graph in multiple ways. In this work, we focus on complete labellings and from these, adopt grounded labellings.

Definition 2.14. *A* ***complete*** $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$***-labelling*** *of an argumentation graph G is a* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$*-labelling such that for every argument A in* $\mathscr{A}_G$*:*
- *A is labelled* IN *if, and only if, all attackers of A are labelled* OUT*, and*
- *A is labelled* OUT *if, and only if, A has an attacker labelled* IN.

Definition 2.15. *A* ***grounded*** $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$***-labelling*** L *of an argumentation graph G is a complete* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$*-labelling of G such that* $\mathsf{IN}(\mathrm{L})$ *is minimal (with respect to set inclusion) among all complete* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$*-labellings of G.*

Example 2 (continued)**.** *The grounded* $\{\mathsf{IN},\mathsf{OUT},\mathsf{UND}\}$*-labelling of the argumentation graph is illustrated in Figure 3.*

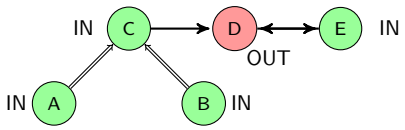

Figure 3: Argumentation graph and its grounded $\{\mathsf{IN},\mathsf{OUT},\mathsf{UND}\}$-labelling.

□

For any argumentation graph, the grounded labelling of the graph exists and is unique. Moreover, for any finite argumentation graph, it can be computed in polynomial time using for example Algorithm 1 [27]. The algorithm begins by labelling IN all arguments not being attacked or whose attackers are OUT (line 4), and then it iteratively labels OUT any argument attacked by an argument labelled IN (line 5). The iteration continues until no more arguments can be labelled IN or OUT (line 6); and it terminates by labelling UND all unlabelled arguments (line 7).

Algorithm 1 Computation of a grounded $\{\mathsf{IN},\mathsf{OUT},\mathsf{UND}\}$-labelling.

1: **input** A finite argumentation graph G,
2: $\mathrm{L}_0 = \langle \emptyset, \emptyset, \emptyset \rangle$,
3: **repeat**
4: $\quad \mathsf{IN}(\mathrm{L}_{i+1}) \leftarrow \mathsf{IN}(\mathrm{L}_i) \cup \{A \mid A \in \mathscr{A}_G$ is not labelled in L_i, and $\forall B \in \mathscr{A}_G$: if B attacks A then $B \in \mathsf{OUT}(\mathrm{L}_i)\}$
5: $\quad \mathsf{OUT}(\mathrm{L}_{i+1}) \leftarrow \mathsf{OUT}(\mathrm{L}_i) \cup \{A \mid A \in \mathscr{A}_G$ is not labelled in L_i, and $\exists B \in \mathscr{A}_G : B$ attacks A and $B \in \mathsf{IN}(\mathrm{L}_{i+1})\}$
6: **until** $\mathrm{L}_i = \mathrm{L}_{i+1}$
7: **return** $\langle \mathsf{IN}(\mathrm{L}_i), \mathsf{OUT}(\mathrm{L}_i), \mathscr{A}_G \backslash (\mathsf{IN}(\mathrm{L}_i) \cup \mathsf{OUT}(\mathrm{L}_i)) \rangle$

We can extend the grounded $\{\mathsf{IN},\mathsf{OUT},\mathsf{UND}\}$-labelling semantics through the introduction of an OFF label. Intuitively, OFF labelled arguments are not evaluated when computing argument acceptance.

Definition 2.16. *Let H be a subargument-complete subgraph of an argumentation graph G. A* ***grounded*** $\{\mathsf{IN},\mathsf{OUT},\mathsf{UND},\mathsf{OFF}\}$***-labelling*** *of G with respect to H is a* $\{\mathsf{IN},\mathsf{OUT},\mathsf{UND},\mathsf{OFF}\}$*-labelling of G such that:*

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

An argumentation graph G has a unique grounded $\{\mathsf{IN},\mathsf{OUT},\mathsf{UND}\}$-labelling, but it has as many grounded $\{\mathsf{IN},\mathsf{OUT},\mathsf{UND},\mathsf{OFF}\}$-labellings as there are subargument-complete subgraphs of G.

Example 2 (continued)**.** *A grounded* $\{\mathsf{IN},\mathsf{OUT},\mathsf{UND},\mathsf{OFF}\}$*-labelling of the argumentation graph is illustrated in Figure 4.*

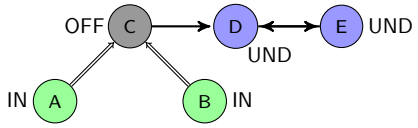

Figure 4: A grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$-labelling.

□

Argument labellings which are equivalent to an argumentation semantics are also called *argument acceptance labellings*. Given a set of labels from a set of argument acceptance labellings, one can compute justification statuses for arguments to obtain an *argument justification labelling*. From an argument justification labelling, individual statements can then be labelled, yielding a *statement justification labelling*. However, one can also directly label statements from argument acceptance labellings, and it is this latter approach the one we use in this paper.

2.4. Statement labelling

Given a set of statements, a *statement labelling* of this set is a (preferably total) function associating any statement with a label. Different specifications for statement labellings are possible [28, 29], but for our purposes, we consider what is perhaps the simplest meaningful labelling, namely the bivalent labelling semantics. Under this labelling, a statement is either accepted or not accepted. If accepted, then the statement is labelled 'in', otherwise it is labelled 'ni'.

As statements are labelled via argument acceptance labellings, the resultant statement labellings are acceptance bivalent $\{\mathsf{in}, \mathsf{ni}\}$-labellings, but we may simply call them bivalent $\{\mathsf{in}, \mathsf{ni}\}$-labellings.

Definition 2.17. *Let $\mathfrak{L}$ be a set of $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$-labellings, $\mathscr{S}$ a set of literal statements. A **bivalent** $\{\mathsf{in}, \mathsf{ni}\}$**-labelling** of $\mathscr{S}$ from $\mathfrak{L}$ is a total function $\mathrm{K} : \mathfrak{L}, \mathscr{S} \to \{\mathsf{in}, \mathsf{ni}\}$ such that for any argument labelling $\mathrm{L} \in \mathfrak{L}$ and any statement $\varphi \in \mathscr{S}$:*
- $\mathrm{K}(\mathrm{L}, \varphi) = \mathsf{in}$ *iff* $\mathsf{IN} \in \{\mathrm{L}(A) \mid \mathrm{con}(A) = \varphi\}$, *and*
- $\mathrm{K}(\mathrm{L}, \varphi) = \mathsf{ni}$ *otherwise.*

Bivalent labellings cannot differentiate between a statement being deemed unjustified and undecidable (the latter by being the conclusion of UND labelled arguments). Trivalent labellings [28] are a simple extension of bivalent labellings which include undecidability.

Definition 2.18. *Let $\mathfrak{L}$ be a set of $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$-labellings, $\mathscr{S}$ a set of literal statements. A **trivalent** $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$**-labelling** of $\mathscr{S}$ and $\mathfrak{L}$ is a total function $\mathrm{K} : \mathfrak{L}, \mathscr{S} \to \{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$ such that for any argument labelling $\mathrm{L} \in \mathfrak{L}$ and any statement $\varphi \in \mathscr{S}$:*
- $\mathrm{K}(\mathrm{L}, \varphi) = \mathsf{in}$ *iff* $\mathsf{IN} \in \{\mathrm{L}(A) \mid \mathrm{con}(A) = \varphi\}$, *and*
- $\mathrm{K}(\mathrm{L}, \varphi) = \mathsf{und}$ *iff* $\mathsf{UND} \in \{\mathrm{L}(A) \mid \mathrm{con}(A) = \varphi\}$ *and* $\mathsf{IN} \notin \{\mathrm{L}(A) \mid \mathrm{con}(A) = \varphi\}$, *and*
- $\mathrm{K}(\mathrm{L}, \varphi) = \mathsf{niund}$ *otherwise.*

Notation 2.9. *A bivalent $\{\mathsf{in}, \mathsf{ni}\}$-labelling or trivalent $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$-labelling K may be represented as a tuple $\langle \mathsf{in}(K), \mathsf{ni}(K) \rangle$ or $\langle \mathsf{in}(K), \mathsf{und}(K), \mathsf{niund}(K) \rangle$ respectively, with the obvious meaning.*

The distinction between bivalent and trivalent labellings is later exploited in our deontic setting. For now we note the following proposition.

Proposition 2.1. *Let* L *be a grounded* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$*-labelling of an argumentation graph constructed from a defeasible theory* T*,* $\mathscr{S}$ *a set of literal statements.*

- *Let* K *be a bivalent* $\{\mathsf{in}, \mathsf{ni}\}$*-labelling of* $\mathscr{S}$ *and from* $\{\mathrm{L}\}$*. For any statement* $\varphi_1, \varphi_2 \in \mathscr{S}$*, such that* $(\varphi_1, \varphi_2) \in \mathrm{Conflicts}(T)$*, if* $\mathrm{K}(\mathrm{L}, \varphi_1) = \mathsf{in}$ *then* $\mathrm{K}(\mathrm{L}, \varphi_2) = \mathsf{ni}$*.*
- *Let* K *be a trivalent* $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$*-labelling of* $\mathscr{S}$ *and from* $\{\mathrm{L}\}$*. For any statement* $\varphi_1, \varphi_2 \in \mathscr{S}$*, such that* $(\varphi_1, \varphi_2) \in \mathrm{Conflicts}(T)$*, if* $\mathrm{K}(\mathrm{L}, \varphi_1) = \mathsf{in}$ *then* $\mathrm{K}(\mathrm{L}, \varphi_2) = \mathsf{niund}$*.*

This proposition is used to show later some results in our probabilistic deontic investigation.

2.5. Probabilistic labellings

To extend our work to the probabilistic setting we take the approach of probabilistic labellings according to which, given an argumentation graph, specific sets of argument labellings are sample spaces [3]. For our purposes, we will focus on grounded labellings and thus we will work with so-called *grounded probabilistic labelling frames*.

Definition 2.19. *A* ***grounded probabilistic labelling frame*** *(or grounded PLF) based on an argumentation graph* G *is a tuple* $\langle G, \langle \Omega, \mathscr{F}, P \rangle \rangle$ *where* $\langle \Omega, \mathscr{F}, P \rangle$ *is a probability space such that:*

- *the sample space* Ω *is the set of grounded* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$*-labellings of* G*;*
- *the* σ*-algebra* $\mathscr{F}$ *is the power set of* Ω*, i.e.* $\mathscr{F} = 2^{\Omega}$*;*
- *the function* P *from* $\mathscr{F}$ *to* $[0,1]$ *is a probability measure (or probability distribution) satisfying Kolmogorov axioms.*

Example 2 (continued)**.** *Let us consider the grounded PLF based on the argumentation graph given in Figure 1 along with a uniform probability distribution. The sample space and the distribution are illustrated in Figure 5.*

□

A grounded PLF defines a probability space, and thus we can work with random variables (written using upper case letters such as X, Y or Z) from Ω to another set of elements. Therefore, we introduce a categorical random variable, which we refer to as a *random labelling*. The random labelling for an argument A, denoted L_A maps from Ω to the set of labels $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$. We may write, for example, $L_A = \mathsf{IN}$ as shorthand for the outcomes $\{\mathrm{L} \in \Omega \mid \mathrm{L}(A) = \mathsf{IN}\}$ (i.e. capturing those outcomes where A is labelled IN). We can follow a similar approach to the labelling of statements. That is, for any statement φ, we can introduce a categorical random variable K_φ which maps Ω to the set of labels $\{\mathsf{in}, \mathsf{ni}\}$ or $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$ (depending on whether we consider bivalent or trivalent labellings). Hence, $P(K_\varphi = l) = 1$, for example, is interpreted as 'the probability that statement φ is labelled l equals one', while $P(K_{\varphi'} = l' \mid K_\varphi = l) = 1$ is interpreted as 'the probability that statement φ' is labelled l', given that statement φ is labelled l, equals one'. In the remainder of this paper, we assume that all specified conditional probabilities are not undefined (a conditional probability $P(A \mid B)$ is undefined if $P(B) = 0$).

Example 2 (continued)**.** *Referring to the probability distribution in Figure 5, we can easily compute that, for example,* $P(L_{\mathsf{A}} = \mathsf{IN}) = 12/20$*,* $P(L_{\mathsf{A}} = \mathsf{OFF}) = 8/20$ *or* $P(L_{\mathsf{E}} = \mathsf{IN} \mid L_{\mathsf{D}} = \mathsf{OUT}) = 1/2$*.* □

A	B	C	D	E	$P(\cdot)$
IN	IN	IN	OUT	IN	1/20
IN	IN	IN	OUT	OFF	1/20
IN	IN	IN	OFF	IN	1/20
IN	IN	IN	OFF	OFF	1/20
IN	IN	OFF	UND	UND	1/20
IN	IN	OFF	IN	OFF	1/20
IN	IN	OFF	OFF	IN	1/20
IN	IN	OFF	OFF	OFF	1/20
IN	OFF	OFF	UND	UND	1/20
IN	OFF	OFF	IN	OFF	1/20
IN	OFF	OFF	OFF	IN	1/20
IN	OFF	OFF	OFF	OFF	1/20
OFF	IN	OFF	UND	UND	1/20
OFF	IN	OFF	IN	OFF	1/20
OFF	IN	OFF	OFF	IN	1/20
OFF	IN	OFF	OFF	OFF	1/20
OFF	OFF	OFF	UND	UND	1/20
OFF	OFF	OFF	IN	OFF	1/20
OFF	OFF	OFF	OFF	IN	1/20
OFF	OFF	OFF	OFF	OFF	1/20

Figure 5: Sample space and probability distribution.

This probabilistic argumentation framework has multiple properties, see [3] for some of them. For our purposes, let us consider the following proposition and corollary.

Proposition 2.2. *Let* $\langle G, \langle \Omega, \mathscr{F}, P \rangle \rangle$ *be a grounded PLF. Let* L *denote any grounded* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$*-labelling such that* $\mathrm{L} \in \Omega$*,* $\mathscr{S}$ *a set of literal statements, and* K *a bivalent* $\{\mathsf{in}, \mathsf{ni}\}$*-labelling or trivalent* $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$*-labelling of* $\mathscr{S}$ *and from* $\{\mathrm{L}\}$*. For any* $\varphi, \varphi' \in \mathscr{S}$ *such that if* $\mathrm{K}(\mathrm{L}, \varphi) = l$ *then* $\mathrm{K}(\mathrm{L}, \varphi') = l'$*,*

$$P(K_{\varphi'} = l' \mid K_{\varphi} = l) = 1.$$

Corollary 2.1. *Let* $\langle G, \langle \Omega, \mathscr{F}, P \rangle \rangle$ *be a grounded PLF. Let* L *denote any grounded* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$*-labelling such that* $\mathrm{L} \in \Omega$*,* $\mathscr{S}$ *a set of literal statements, and* K *a bivalent* $\{\mathsf{in}, \mathsf{ni}\}$*-labelling or trivalent* $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$*-labelling of* $\mathscr{S}$ *and from* $\{\mathrm{L}\}$*. For any* $\varphi, \varphi' \in \mathscr{S}$ *such that if* $\mathrm{K}(\mathrm{L}, \varphi) = l$ *then* $\mathrm{K}(\mathrm{L}, \varphi') = l'$*,*

$$P(K_{\varphi} = l) \leq P(K_{\varphi'} = l').$$

We will also use the proposition and corollary below [3] when later studying the probabilistic relationships amongst deontic statements.

Proposition 2.3. *Let* $\langle G, \langle \Omega, \mathscr{F}, P \rangle \rangle$ *be a grounded PLF where* G *is constructed from a defeasible theory* $T = \langle Rules, Conflicts, \succ \rangle$*,* $\mathscr{S}$ *a set of literal statements. For any* $\varphi_1, \varphi_2 \in \mathscr{S}$ *such that* $Conflicts(\varphi_1, \varphi_2)$*,*

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

Corollary 2.2. *Let* $\langle G, \langle \Omega, \mathscr{F}, P \rangle \rangle$ *be a grounded PLF where* G *is constructed from a defeasible theory* $T = \langle Rules, Conflicts, \succ \rangle$*,* $\mathscr{S}$ *a set of literal statements. For any* $\varphi_1, \varphi_2 \in \mathscr{S}$ *such that* $Conflicts(\varphi_1, \varphi_2)$*,*

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

where
- $P(K_{\varphi_2} \neq \mathsf{in}) = P(K_{\varphi_2} = \mathsf{ni})$ *in the case of bivalent* $\{\mathsf{in}, \mathsf{ni}\}$*-labellings;*
- $P(K_{\varphi_2} \neq \mathsf{in}) = P(K_{\varphi_2} = \mathsf{und}) + P(K_{\varphi_2} = \mathsf{niund})$ *in the case of trivalent* $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$*-labellings.*

These properties of the framework are used later to obtain probabilistic results in our deontic investigation.

The 'source' and interpretation of probability values can be diverse. In a classical approach for example, one may simply assume that all the possible outcomes are equally possible and thus that they should share the same probability value (as illustrated in Example 2 above). Alternatively, in a frequentist interpretation, the probability of any event is the frequency of occurrences of the event in a collection of outcomes. Clearly, there are various ways to associate an event with probability values and to interpret these probabilities. In this paper, we assume that probability distributions are simply given *a priori*.

This probabilistic argumentation framework subsumes, or can be related to, a number of existing approaches to probabilistic argumentation [3], and consequently it can be applied to various investigations in the legal domain, such as game-theoretical investigations of optimal strategies in dialogue games [14]. It can also be used in machine learning endeavours, such as structure learning [30, 31] or combinations of probabilistic argumentation with reinforcement learning to determine actions to pursue in environments featuring basic conditional obligations [12, 32]. In the remainder of the paper, we consider how deontic concepts can be represented and reasoned about by exploiting probabilistic labellings.

To recap, we described a propositional language with modal operators, which — in the next section — will be instantiated into deontic modalities. We have also described the syntax of defeasible rules, and how such rules can be combined to form arguments. We introduced a labelling approach to argument semantics, from which the acceptance status of statements can be computed. Finally, we described how uncertainty and probability values can be associated with arguments and statements within our system.

3. Probabilistic Deontic Argumentation

Having laid out a simple probabilistic rule-based argumentation system, we can now specify a deontic version of it. To do so, we adopt the deontic argumentation framework of [9] and develop it within our probabilistic setting.

3.1. Deontic language

In law and moral reasoning multiple normative concepts occur. We do not propose to consider all these concepts, but rather focus on three basic deontic notions, namely obligations, prohibitions, and permissions.

We assume that such deontic operators are interdefinable: the prohibition of something is the obligation of its opposite, and the permission of something is the negation of its prohibition. Therefore, we focus on a single deontic obligation operator O, and accordingly, we assume a language $\mathscr{L}_D$ whose literal statements are defined as follows, cf. Definition 2.1.

Definition 3.1. *A literal statement of a language $\mathscr{L}_D$ is either a plain literal statement or a deontic literal statement, where:*

- *a **plain literal statement** is either an atomic proposition p or the negation of an atomic proposition, i.e. $\neg p$, and*
- *a **deontic literal statement** is a statement of the form $\mathsf{O}\gamma$ or $\neg\mathsf{O}\gamma$ such that γ is a plain literal statement.*

As indicated above, prohibitions and permissions are captured by assuming that a prohibition on something ($\mathsf{F}\gamma$) is equivalently expressed by the obligation of its opposite ($\mathsf{O}\overline{\gamma}$), and a permission for something ($\mathsf{P}\gamma$) is equivalent to the negation of the obligation of the opposite ($\neg\mathsf{O}\overline{\gamma}$).

Notation 3.1. *As syntactic sugar, we may write $\mathsf{O}\overline{\gamma}$ as $\mathsf{F}\gamma$, and $\neg\mathsf{O}\overline{\gamma}$ as $\mathsf{P}\gamma$.*

A normative system may not be complete, i.e. it may not regulate every possible state of affairs. This means that the system does not specify, for every state of affairs, whether that state is obligatory, prohibited or permitted: there may be gaps, namely states of affairs for which no deontic position is specified. In our model, we only consider those states of affairs that are described by plain literals. Thus, by *normative completeness*, we refer to the quality of a normative system having deontic positions specified for all plain literals. We focus on normative systems that are meant to be complete, and whose gaps are filled according to the 'principle of prohibition' (see [33] p.125).

The principle of prohibition can be formulated as follows: 'everything that is not prohibited is permitted'. Although the principle is rather clear at first sight, it may not be given a unique interpretation. We can adopt a mere tautological reading, based on the above notational convention according to which being permitted for something means not being obligatory of the opposite, i.e. according to which $\mathsf{P}\gamma$ is defined as $\neg\mathsf{O}\overline{\gamma}$. Following this reading, we can say that a normative system specifies that a state of affairs is permitted when it specifies that there is no prohibition of it: the normative system states that γ is permitted, when it entails that $\neg\mathsf{O}\overline{\gamma}$. This reading does not add anything to the content of the concerned normative system. In the absence of any norms specifying that $\neg\mathsf{O}\overline{\gamma}$ or that $\mathsf{O}\gamma$ (in which case γ would be permitted, assuming that obligation entails permission) the deontic status of γ remains undetermined.

Here we adopt a different reading of the principle of prohibition, namely the reading according to which everything is permitted unless it is prohibited. More precisely, according to this reading, a normative system based on the principle of permission grants to a state of affairs the status of being permitted whenever it does not grant to the same state of affairs the status of being prohibited. Following this interpretation, the principle of prohibition is no longer tautological. Rather, it becomes a normative principle included in a normative system, which is used for its completion: every gap in the system based on the principle of permission is closed by generating permissions.

We use the terms 'strong permission' and 'weak permission' to distinguish the permission that is derived from a specific norm (whose consequent denies that a state of affair is prohibited, or equivalently, permits it) from the permission that is obtained according to the general principle of prohibition, or also, as we shall see, from the principle that whatever is obligatory is permitted.

Note that our concept of a weak permission differs from the concept of a weak permission, as characterised by G. H. von Wright [34], or C.E. Alchourrón and E. Bulygin [33]. For these authors a weak permission does not derive from a normative specification (for example from a permissive norm): it is an assertion about a normative system, namely, the assertion that the

normative system does not entail a prohibition. On the contrary, we view a weak permission as an additional normative specification, which is added to a normative system, according to the principle of prohibition, whenever the normative system does not entail a corresponding prohibition. Finally note that the distinction between strong and weak permissions is not directly represented in our deontic language; this distinction depends in the way in which the permission at issue has been derived, from an explicit permissive norm, or rather according to the general principle of prohibition.

A defeasible rule can specify varied relationships amongst (deontic) literal statements of a given language $\mathscr{L}_D$. Such rules are called normative defeasible rules.

Definition 3.2. *Given a language $\mathscr{L}_D$, a **normative defeasible rule** is a defeasible rule over a set of literal statements in $\mathscr{L}_D$.*

For the sake of simplicity, norms potentially captured by 'modalised rules', e.g. rules of the form $\mathsf{O}(r : \varphi_1, \ldots, \varphi_n, \sim \varphi'_1, \ldots, \sim \varphi'_m \Rightarrow \varphi)$, are not accounted for in this paper. Such constructs and their meanings are left for future work.

Normative rules are partitioned into *foreground rules* and *background rules*. Foreground rules provide substantive normative regulations for particular normative domains, while background rules express general deontic principles underlying the normative system being considered.

Foreground rules are domain-dependent, and represent primary norms such as *constitutive rules* or *regulative rules*. The effect of a constitutive rule is to define a term as understood in a given situation or to 'create' an institutional entity from a set of brute or institutional facts. A regulative rule, on the other hand, determines the conditions (premises) when a 'deontic' effect (obligation, prohibition, permission) is in force. While constitutive and regulative norms have been formally approached in various (and sometimes sophisticated) ways in the literature [35], the distinction is simply addressed in the present system: the consequent of the rule is a plain literal for constitutive rules, and a deontic literal for regulative rules. A regulative rule whose head is a (strong) permission would typically be used to specify an exception to a prohibition (as discussed by A. Ross [36]), but such a rule can also be used to stress a permission and clarify its conditions without being an exception to any existing prohibitions.

Background rules are domain-independent and apply to a whole legal system, or to broad sections of it (e.g. it has often been argued that the principle of prohibition applies to criminal law). They express general deontic assumptions. These background rules can be viewed as defeasible rule schemata which are isomorphic to some pieces of very basic legal doctrines or some axioms of a deontic system. Various background defeasible rules may be proposed, we provide some examples of such rules below.

d_γ:	$\mathsf{O}\gamma \Rightarrow \mathsf{P}\gamma$	An obligation $\mathsf{O}\gamma$ implies a permission $\mathsf{P}\gamma$ (cf. Axiom 'D' in deontic logics),
p_γ:	$\Rightarrow \mathsf{P}\gamma$	Anything is permitted prima facie.
k_γ:	$\sim \mathsf{O}\overline{\gamma} \Rightarrow \mathsf{P}\gamma$	Anything that is not prohibited is permitted.

These background rules can be employed to build arguments supporting permissions. These arguments are successful as long as no corresponding prohibitions can be accepted. We

recognise that such rules do not capture all possible deontic schemata, but note they are relevant in many contexts.

Different sets of background rules result in systems with different behaviour. In this work, we focus on two sets of background rules: $\{\mathsf{d}\_\gamma, \mathsf{p}\_\gamma\}$ and $\{\mathsf{d}\_\gamma, \mathsf{k}\_\gamma\}$.

Definition 3.3. *A set of background defeasible rule schemata B is*
- *a* ***prima facie permissive set of background defeasible rule schemata*** *iff* $B = \{\mathsf{d}\_\gamma, \mathsf{p}\_\gamma\}$*;*
- *a* ***Kelsenian permissive set of background defeasible rule schemata*** *iff* $B = \{\mathsf{d}\_\gamma, \mathsf{k}\_\gamma\}$*;*
- *a* ***permissive set of background defeasible rule schemata*** *iff* $B = \{\mathsf{d}\_\gamma, \mathsf{p}\_\gamma\}$ *or* $B = \{\mathsf{d}\_\gamma, \mathsf{k}\_\gamma\}$.

'Prima facie permissive' and 'Kelsenian permissive' sets of rules both indicate that anything is defeasibly permitted. However, the Kelsenian permissive set may better reflect the principle of prohibition as exposed by the legal theorist H. Kelsen (thus its name). For both sets, we will see that such background rules are not enough to obtain normative completeness when using bivalent labellings.

Whatever the set of background defeasible rules, we will ground the rules over a set of propositions. For our purposes, we do so over the propositions of an input (domain-dependent) defeasible theory to yield a set of background rules. Then background rules are combined with the rules of the theory to obtain backgrounded rules.

Definition 3.4. *A set of rules is a* ***set of background rules*** *with respect to a defeasible theory T and a set of background defeasible rule schemata B, denoted* $\mathrm{BackRules}(T,B)$*, iff*
- $\mathrm{BackRules}(T,B) = \{\mathsf{d}\_\gamma, \mathsf{d}\_\overline{\gamma}, \mathsf{p}\_\gamma, \mathsf{p}\_\overline{\gamma} \mid \gamma \in \mathrm{Prop}(T)\}$ *if B is a prima facie permissive set of background defeasible rule schemata;*
- $\mathrm{BackRules}(T,B) = \{\mathsf{d}\_\gamma, \mathsf{d}\_\overline{\gamma}, \mathsf{k}\_\gamma, \mathsf{k}\_\overline{\gamma} \mid \gamma \in \mathrm{Prop}(T)\}$ *if B is a Kelsenian permissive set of background defeasible rule schemata.*

Definition 3.5. *A set of rules Rules is a* ***backgrounded set of rules*** *with respect to a defeasible theory T and a set of background defeasible rule schemata B iff* $\mathit{Rules} = \mathrm{Rules}(T) \cup \mathrm{BackRules}(T,B)$.

Example 3. *Returning to our park example, let us assume the presence of a policy stating that vehicles are prohibited from entering the park, unless there is an emergency. Let us consider the following atoms and informal meanings.*

> vehi*: a vehicle stays at the entrance of the park.*
> emer*: there is an emergency.*
> enter*: the vehicle enters into the park.*

Accordingly, Fenter *means that the vehicle is forbidden to enter into the park, and* Penter *means that the vehicle is permitted to enter.* Fvehi *means that it is forbidden that a vehicle stays at the entrance of the park,* Pvehi *means that it is permitted that the vehicle stays at the park entrance, and so on.*

The policy may be easily formalised by the foreground defeasible theory $\langle\{\mathsf{r}\}, \emptyset, \emptyset\rangle$ *where*

$\mathsf{r}:\quad \mathsf{vehi}, \sim \mathsf{emer} \Rightarrow \mathsf{Fenter}$

The Kelsenian permissive set of background rules with respect to theory $\langle\{\mathsf{r}\}, \emptyset, \emptyset\}\rangle$ *includes all the following rules.*

d_vehi :	Ovehi	$\Rightarrow$ Pvehi	d_$\neg$vehi :	O$\neg$vehi	$\Rightarrow$ P$\neg$vehi
k_vehi :	$\sim$ Fvehi	$\Rightarrow$ Pvehi	k_$\neg$vehi :	$\sim$ F$\neg$vehi	$\Rightarrow$ P$\neg$vehi
d_emer :	Oemer	$\Rightarrow$ Pemer	d_$\neg$emer :	O$\neg$emer	$\Rightarrow$ P$\neg$emer
k_emer :	$\sim$ Femer	$\Rightarrow$ Pemer	k_$\neg$emer :	$\sim$ F$\neg$emer	$\Rightarrow$ P$\neg$emer
d_enter :	Oenter	$\Rightarrow$ Penter	d_$\neg$enter :	O$\neg$enter	$\Rightarrow$ P$\neg$enter
k_enter :	$\sim$ Fenter	$\Rightarrow$ Penter	k_$\neg$enter :	$\sim$ F$\neg$enter	$\Rightarrow$ P$\neg$enter

□

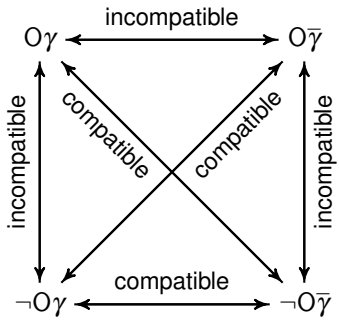

Figure 6: Deontic square of compatibility relation.

Concerning conflicts, we have normative conflicts, and any normative conflict can be a foreground conflict or a background conflict, cf. [9]. A normative conflict has the form $(\gamma, \overline{\gamma})$ or $(\mathsf{O}\gamma, \mathsf{O}\overline{\gamma})$ or $(\neg\mathsf{O}\gamma, \mathsf{O}\gamma)$ or $(\mathsf{O}\gamma, \neg\mathsf{O}\gamma)$. The latter three types of conflict are deontic conflicts, and they can be visualised in the deontic square drawn in Figure 6.

Definition 3.6. *A **normative conflict** is a conflict of the form* $(\gamma, \overline{\gamma})$ *or* $(\mathsf{O}\gamma, \mathsf{O}\overline{\gamma})$ *or* $(\neg\mathsf{O}\gamma, \mathsf{O}\gamma)$ *or* $(\mathsf{O}\gamma, \neg\mathsf{O}\gamma)$.

For example, for any atomic proposition p, a conflict $(\mathsf{O}p, \neg\mathsf{O}p)$ is captured by the conflict of the form $(\mathsf{O}\gamma, \neg\mathsf{O}\gamma)$ where $\gamma = p$. Similarly, a conflict $(\mathsf{O}\neg p, \neg\mathsf{O}\neg p)$ is captured by $(\mathsf{O}\gamma, \neg\mathsf{O}\gamma)$ where $\gamma = \neg p$.

Foreground conflicts are normative conflicts. They are meant to be conflicts which are specified in foreground theories (see Definition 3.11). However, such foreground conflicts may be incomplete — some (intuitively desirable) conflicts may not be included in the foreground relation. To ensure completeness of conflicts, we utilise background conflicts. Background conflicts are domain-independent in normative reasoning, and a conflict relation is backgrounded by such conflicts if, and only if, they are included in the conflict relation.

Definition 3.7. *A set of conflicts is a **set of background conflicts** with respect to a defeasible theory T, denoted* BackConflicts(T)*, iff* BackConflicts$(T) = \{(\gamma, \overline{\gamma}), (\overline{\gamma}, \gamma), (\mathsf{O}\gamma, \mathsf{O}\overline{\gamma}), (\mathsf{O}\overline{\gamma}, \mathsf{O}\gamma), (\mathsf{O}\gamma, \neg\mathsf{O}\gamma), (\neg\mathsf{O}\gamma, \mathsf{O}\gamma), (\mathsf{O}\overline{\gamma}, \neg\mathsf{O}\overline{\gamma}), (\neg\mathsf{O}\overline{\gamma}, \mathsf{O}\overline{\gamma}) \mid \gamma \in$ Prop$(T)\}$.

Definition 3.8. *A conflict relation $Conflicts$ is a **backgrounded conflict relation** with respect to a defeasible theory T iff $Conflicts =$* Conflicts$(T) \cup$ BackConflicts(T).

Example 3 (continued)**.** *The background conflict pairs are as follows.*

$(\mathsf{vehi}, \neg\mathsf{vehi})$	$(\mathsf{emer}, \neg\mathsf{emer})$	$(\mathsf{enter}, \neg\mathsf{enter})$
$(\neg\mathsf{vehi}, \mathsf{vehi})$	$(\neg\mathsf{emer}, \mathsf{emer})$	$(\neg\mathsf{enter}, \mathsf{enter})$
$(\mathsf{O}\mathsf{vehi}, \mathsf{O}\neg\mathsf{vehi})$	$(\mathsf{O}\mathsf{emer}, \mathsf{O}\neg\mathsf{emer})$	$(\mathsf{O}\mathsf{enter}, \mathsf{O}\neg\mathsf{enter})$
$(\mathsf{O}\neg\mathsf{vehi}, \mathsf{O}\mathsf{vehi})$	$(\mathsf{O}\neg\mathsf{emer}, \mathsf{O}\mathsf{emer})$	$(\mathsf{O}\neg\mathsf{enter}, \mathsf{O}\mathsf{enter})$
$(\mathsf{O}\mathsf{vehi}, \neg\mathsf{O}\mathsf{vehi})$	$(\mathsf{O}\mathsf{emer}, \neg\mathsf{O}\mathsf{emer})$	$(\mathsf{O}\mathsf{enter}, \neg\mathsf{O}\mathsf{enter})$
$(\neg\mathsf{O}\mathsf{vehi}, \mathsf{O}\mathsf{vehi})$	$(\neg\mathsf{O}\mathsf{emer}, \mathsf{O}\mathsf{emer})$	$(\neg\mathsf{O}\mathsf{enter}, \mathsf{O}\mathsf{enter})$
$(\mathsf{O}\neg\mathsf{vehi}, \neg\mathsf{O}\neg\mathsf{vehi})$	$(\mathsf{O}\neg\mathsf{emer}, \neg\mathsf{O}\neg\mathsf{emer})$	$(\mathsf{O}\neg\mathsf{enter}, \neg\mathsf{O}\neg\mathsf{enter})$
$(\neg\mathsf{O}\neg\mathsf{vehi}, \mathsf{O}\neg\mathsf{vehi})$	$(\neg\mathsf{O}\neg\mathsf{emer}, \mathsf{O}\neg\mathsf{emer})$	$(\neg\mathsf{O}\neg\mathsf{enter}, \mathsf{O}\neg\mathsf{enter})$

□

Given a defeasible theory T where any conflict in the relation $\text{Conflicts}(T)$ is a normative conflict, we have that $\text{Conflicts}(T) \subseteq \text{BackConflicts}(T)$. Consequently, if $\text{Conflicts}(T)$ is specified as a set of foreground conflicts which is backgrounded, then every conflict in $\text{Conflicts}(T)$ is a foreground conflict which is also a background conflict in $\text{BackConflicts}(T)$, while in some cases a background conflict may not be a foreground conflict. Thus, if one works with backgrounded conflicts relations of a foreground defeasible theory, as we will do, foreground conflicts are not necessary within the domain specification. Foreground conflicts are nevertheless necessary when specifying conflicts of foreground theories (as per Definition 3.11 below), especially in any setting where such conflicts would not be backgrounded

It is worth recapping our setting on conflicts. We have defined normative conflicts which are conflicts capturing propositional conflicts, as well as deontic conflicts. A normative conflict can be a foreground conflict or a background conflict. Foreground conflicts will be those conflicts which are made explicit when specifying a foreground theory. A backgrounded conflict relation is one for which all possible normative conflicts are included.

Foreground and background deontic rules may have conflicting heads, and to ensure correct reasoning patterns, background superiorities can be proposed. We assume that background rules are inferior to any foreground rules.

Definition 3.9. *A superiority relation is a* ***background superiority relation*** *with respect to a defeasible theory T and a set of background defeasible rule schemata B, denoted* $\text{BackSup}(T,B)$*, iff* $\text{BackSup}(T,B) = \{(s,r) \mid s \in \text{Rules}(T), r \in \text{BackRules}(T,B)\}$.

Definition 3.10. *A superiority relation $\succ$ is a* ***backgrounded superiority relation*** *with respect to a defeasible theory T and a set of background defeasible rule schemata B iff* $\succ = \succ(T) \cup \text{BackSup}(T,B)$.

Example 3 (continued)**.** *The background pairs in the superiority relation are as follows.*

$(\mathsf{r}, \mathsf{d\_vehi})$	$(\mathsf{r}, \mathsf{d\_emer})$	$(\mathsf{r}, \mathsf{d\_enter})$
$(\mathsf{r}, \mathsf{k\_vehi})$	$(\mathsf{r}, \mathsf{k\_emer})$	$(\mathsf{r}, \mathsf{k\_enter})$
$(\mathsf{r}, \mathsf{d\_}\neg\mathsf{vehi})$	$(\mathsf{r}, \mathsf{d\_}\neg\mathsf{emer})$	$(\mathsf{r}, \mathsf{d\_}\neg\mathsf{enter})$
$(\mathsf{r}, \mathsf{k\_}\neg\mathsf{vehi})$	$(\mathsf{r}, \mathsf{k\_}\neg\mathsf{emer})$	$(\mathsf{r}, \mathsf{k\_}\neg\mathsf{enter})$

□

3.2. Deontic defeasible theory and argumentation graphs

We now propose to 'background' defeasible theories where rules, conflicts and superiority relationships are backgrounded with respect to any foreground defeasible theory. A foreground

defeasible theory is a defeasible theory where rules are not background rules, i.e. rules whose identifiers are not identifiers of any background rules.

Definition 3.11. *A defeasible theory $\langle Rules, Conflicts, \succ \rangle$ is a **foreground defeasible theory** iff*
- *every defeasible rule in Rules is a (foreground) normative defeasible rule which is not a background defeasible rule, and*
- *every conflict in Conflicts is a (foreground) normative conflict.*

Definition 3.12. *A defeasible theory $\langle Rules, Conflicts, \succ \rangle$ is a **backgrounded defeasible theory** of a foreground defeasible theory T with a set of background defeasible rule schemata B iff*
- *Rules is a backgrounded set of rules with respect to T and B, and*
- *Conflicts is a backgrounded conflict relation with respect to T, and*
- *$\succ$ is a backgrounded superiority relation with respect to T and B.*

In practice, we will first construct a foreground defeasible theory which can be backgrounded to obtain a backgrounded defeasible theory. In the remainder of the paper, we assume that any defeasible theory is backgrounded with a permissive set of background defeasible rule schemata to obtain a permissive defeasible theory.

Definition 3.13. *A defeasible theory is a **permissive defeasible theory** iff it is a backgrounded defeasible theory with a permissive set of background defeasible rule schemata.*

We can build arguments from a backgrounded defeasible theory. When building arguments, chaining rules implicitly use the detachment of the consequent of rules. In that regard, deontic studies usually distinguish between factual detachment and deontic detachment. In this work, we restrict ourselves to factual detachment, leaving (defeasible) deontic detachments (which are considered somewhat controversial in the literature [37]) to future work. Once arguments are built we can form an argumentation graph, and then label arguments and (deontic) statements to determine their statuses, as discussed next.

3.3. Probabilistic labellings

Given a backgrounded defeasible theory and an argumentation graph built from it, we can now label (deontic) literal statements and associate their acceptance statuses with probability values.

To label arguments and (deontic) statements, we first consider labelling semantics as described in Section 2. Hence, given the argumentation graph built from a backgrounded defeasible theory, arguments are labelled according to the grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$-labelling semantics. Then, literal statements are labelled with acceptance statement labelling semantics. Such labellings are thus a straightforward application of standard labelling semantics corresponding to some common modes of reasoning.

However, bare $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$-labellings are cumbersome in that they allow any 'doctrinal' arguments, i.e. arguments based on background principles, to be excluded, i.e. labelled OFF, whereas such doctrinal arguments should always be included as long as the corresponding background principles are endorsed in the legal system being considered. For this reason, we may introduce 'legitimate' (legit) grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$-labellings and legit grounded probabilistic labelling frames.

Definition 3.14. *Let G be an argumentation graph constructed from a permissive defeasible theory. A grounded* $\{\mathsf{IN},\mathsf{OUT},\mathsf{UND},\mathsf{OFF}\}$*-labelling* L *of G is a* ***legit grounded*** $\{\mathsf{IN},\mathsf{OUT},\mathsf{UND},\mathsf{OFF}\}$***-labelling*** *of G iff for any argument* $A \in \mathscr{A}_G$*:*

- *if* $\mathrm{TopRule}(A) = \mathsf{d}\_\gamma$ *and* $\mathrm{DirectSub}(A) = \{B\}$ *and* $\mathrm{L}(B) \neq \mathsf{OFF}$ *then* $\mathrm{L}(A) \neq \mathsf{OFF}$*, and*
- *if* $\mathrm{TopRule}(A) = \mathsf{p}\_\gamma$ *or* $\mathrm{TopRule}(A) = \mathsf{k}\_\gamma$ *then* $\mathrm{L}(A) \neq \mathsf{OFF}$.

Example 4. *Let us consider the following arguments, along with the associated argumentation graph* G *drawn in Figure 7. All the legit grounded* $\{\mathsf{IN},\mathsf{OUT},\mathsf{UND},\mathsf{OFF}\}$*-labellings of* G *are given in Table 1.*

O1 :		$\Rightarrow_{\mathsf{r}}$ Oa	O2 :		$\Rightarrow_{\mathsf{r}'}$ O¬a
P1 :	O1	$\Rightarrow_{\mathsf{d\_a}}$ Pa	P2 :	O2	$\Rightarrow_{\mathsf{d\_\neg a}}$ P¬a
W1 :	$\sim$ Fa	$\Rightarrow_{\mathsf{k\_a}}$ Pa	W2 :	$\sim$ F¬a	$\Rightarrow_{\mathsf{k\_\neg a}}$ P¬a

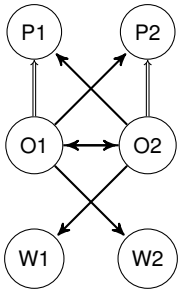

Figure 7: Argumentation graph G.

O1	O2	P1	P2	W1	W2
OFF	OFF	OFF	OFF	IN	IN
IN	OFF	IN	OFF	IN	OUT
OFF	IN	OFF	IN	OUT	IN
UND	UND	UND	UND	UND	UND

Table 1: Legit grounded $\{\mathsf{IN},\mathsf{OUT},\mathsf{UND},\mathsf{OFF}\}$-labellings of argumentation graph G.

□

Definition 3.15. *Let G be an argumentation graph constructed from a permissive defeasible theory. A grounded PLF* $\langle G, \langle \Omega, \mathscr{F}, P\rangle\rangle$ *is a* ***legit grounded probabilistic labelling frame*** *iff for any* $\{\mathsf{IN},\mathsf{OUT},\mathsf{UND},\mathsf{OFF}\}$*-labelling* $\mathrm{L} \in \Omega$ *such that* L *is not a legit grounded* $\{\mathsf{IN},\mathsf{OUT},\mathsf{UND},\mathsf{OFF}\}$*-labelling,* $P(\mathrm{L}) = 0$.

Moving to statement labellings, (legit) grounded probabilistic labelling frames have multiple properties. Let us mention a few of them. First of all, bivalent $\{\mathsf{in},\mathsf{ni}\}$-labelling and trivalent $\{\mathsf{in},\mathsf{und},\mathsf{niund}\}$-labelling semantics imply that two conflicting deontic statements cannot be simultaneously labelled in: if a deontic statement is labelled in then any conflicting statement is labelled ni (in the case of $\{\mathsf{in},\mathsf{ni}\}$-labellings) or niund (in the case of $\{\mathsf{in},\mathsf{und},\mathsf{niund}\}$-labellings), see Proposition 2.1.

Proposition 3.1. *Let* L *be a grounded* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$*-labelling of an argumentation graph constructed from a permissive defeasible theory* T*,* $\mathscr{S}$ *a set of literal statements such that* $\mathscr{S} = \{p, \neg p \mid p \in \mathrm{Prop}(T)\}$*.*

- *Let* K *be a bivalent* $\{\mathsf{in}, \mathsf{ni}\}$*-labelling of* $\mathscr{S}$ *and from* $\{\mathrm{L}\}$*. For any* $\gamma \in \mathscr{S}$*:*
  - *if* $\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{in}$ *then* $\mathrm{K}(\mathrm{L}, \mathsf{O}\overline{\gamma}) = \mathsf{ni}$*;*
  - *if* $\mathrm{K}(\mathrm{L}, \neg\mathsf{O}\gamma) = \mathsf{in}$ *then* $\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{ni}$*;*
  - *if* $\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{in}$ *then* $\mathrm{K}(\mathrm{L}, \neg\mathsf{O}\gamma) = \mathsf{ni}$*.*
- *Let* K *be a trivalent* $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$*-labelling of* $\mathscr{S}$ *and from* $\{\mathrm{L}\}$*. For any* $\gamma \in \mathscr{S}$*:*
  - *if* $\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{in}$ *then* $\mathrm{K}(\mathrm{L}, \mathsf{O}\overline{\gamma}) = \mathsf{niund}$*;*
  - *if* $\mathrm{K}(\mathrm{L}, \neg\mathsf{O}\gamma) = \mathsf{in}$ *then* $\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{niund}$*;*
  - *if* $\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{in}$ *then* $\mathrm{K}(\mathrm{L}, \neg\mathsf{O}\gamma) = \mathsf{niund}$*.*

By Proposition 2.2 and Corollary 2.1, results in Proposition 3.1 can be understood in our probabilistic development as follows.

Proposition 3.2. *Let* $\langle G, \langle \Omega, \mathscr{F}, P\rangle\rangle$ *be a grounded PLF where* G *is constructed from a permissive defeasible theory* T*, and* $\mathscr{S}$ *is a set of literal statements such that* $\mathscr{S} = \{p, \neg p \mid p \in \mathrm{Prop}(T)\}$*.*

- *In case of bivalent* $\{\mathsf{in}, \mathsf{ni}\}$*-labellings, for any* $\gamma \in \mathscr{S}$*:*
  - $P(K_{\mathsf{O}\overline{\gamma}} = \mathsf{ni} \mid K_{\mathsf{O}\gamma} = \mathsf{in}) = 1$*;*
  - $P(K_{\mathsf{O}\gamma} = \mathsf{ni} \mid K_{\neg\mathsf{O}\gamma} = \mathsf{in}) = 1$*;*
  - $P(K_{\neg\mathsf{O}\gamma} = \mathsf{ni} \mid K_{\mathsf{O}\gamma} = \mathsf{in}) = 1$*.*
- *In case of trivalent* $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$*-labellings, for any* $\gamma \in \mathscr{S}$*:*
  - $P(K_{\mathsf{O}\overline{\gamma}} = \mathsf{niund} \mid K_{\mathsf{O}\gamma} = \mathsf{in}) = 1$*;*
  - $P(K_{\mathsf{O}\gamma} = \mathsf{niund} \mid K_{\neg\mathsf{O}\gamma} = \mathsf{in}) = 1$*;*
  - $P(K_{\neg\mathsf{O}\gamma} = \mathsf{niund} \mid K_{\mathsf{O}\gamma} = \mathsf{in}) = 1$*.*

Proof. Let us provide the proof for the first item only in the case of bivalent $\{\mathsf{in}, \mathsf{ni}\}$-labellings (proofs for other items follow the same structure). From Proposition 3.1, if $\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{in}$ then $\mathrm{K}(\mathrm{L}, \mathsf{O}\overline{\gamma}) = \mathsf{ni}$. Therefore, by Proposition 2.2, $P(K_{\mathsf{O}\overline{\gamma}} = \mathsf{ni} \mid K_{\mathsf{O}\gamma} = \mathsf{in}) = 1$. □

Corollary 3.1. *Let* $\langle G, \langle \Omega, \mathscr{F}, P\rangle\rangle$ *be a grounded PLF where* G *is constructed from a permissive defeasible theory* T*, and* $\mathscr{S}$ *is a set of literal statements such that* $\mathscr{S} = \{p, \neg p \mid p \in \mathrm{Prop}(T)\}$*.*

- *In case of bivalent* $\{\mathsf{in}, \mathsf{ni}\}$*-labellings, for any* $\gamma \in \mathscr{S}$*:*
  - $P(K_{\mathsf{O}\gamma} = \mathsf{in}) \leq P(K_{\mathsf{O}\overline{\gamma}} = \mathsf{ni})$*;*
  - $P(K_{\neg\mathsf{O}\gamma} = \mathsf{in}) \leq P(K_{\mathsf{O}\gamma} = \mathsf{ni})$*;*
  - $P(K_{\mathsf{O}\gamma} = \mathsf{in}) \leq P(K_{\neg\mathsf{O}\gamma} = \mathsf{ni})$*.*
- *In case of trivalent* $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$*-labellings, for any* $\gamma \in \mathscr{S}$*:*
  - $P(K_{\mathsf{O}\gamma} = \mathsf{in}) \leq P(K_{\mathsf{O}\overline{\gamma}} = \mathsf{niund})$*;*
  - $P(K_{\neg\mathsf{O}\gamma} = \mathsf{in}) \leq P(K_{\mathsf{O}\gamma} = \mathsf{niund})$*;*
  - $P(K_{\mathsf{O}\gamma} = \mathsf{in}) \leq P(K_{\neg\mathsf{O}\gamma} = \mathsf{niund})$*.*

Proof. Let us provide the proof for the first item only in the case of bivalent $\{\mathsf{in}, \mathsf{ni}\}$-labellings (proofs for other items follow the same structure). From Proposition 3.1, if $\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{in}$ then $\mathrm{K}(\mathrm{L}, \mathsf{O}\overline{\gamma}) = \mathsf{ni}$. Therefore, by Corollary 2.1, $P(K_{\mathsf{O}\gamma} = \mathsf{in}) \leq P(K_{\mathsf{O}\overline{\gamma}} = \mathsf{ni})$. □

On the same note, Proposition 2.3 can be instantiated with deontic statements which are

inherently conflicting as specified in Definition 3.6. For instance

$$P(K_{\mathsf{F}\gamma} = \mathsf{in}) + P(K_{\mathsf{O}\gamma} = \mathsf{in}) \leq 1 \tag{1}$$
$$P(K_{\mathsf{F}\gamma} = \mathsf{in}) + P(K_{\mathsf{P}\gamma} = \mathsf{in}) \leq 1 \tag{2}$$

and thus by Corollary 2.2 we obtain

$$P(K_{\mathsf{O}\gamma} = \mathsf{in}) \leq P(K_{\mathsf{F}\gamma} = \mathsf{ni}) \tag{3}$$
$$P(K_{\mathsf{O}\gamma} = \mathsf{in}) \leq P(K_{\mathsf{F}\gamma} = \mathsf{niund}) + P(K_{\mathsf{F}\gamma} = \mathsf{und}). \tag{4}$$

$$P(K_{\mathsf{P}\gamma} = \mathsf{in}) \leq P(K_{\mathsf{F}\gamma} = \mathsf{ni}) \tag{5}$$
$$P(K_{\mathsf{P}\gamma} = \mathsf{in}) \leq P(K_{\mathsf{F}\gamma} = \mathsf{niund}) + P(K_{\mathsf{F}\gamma} = \mathsf{und}). \tag{6}$$

These results are more specific than results given in Corollary 3.1 in the case of trivalent $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$-labellings.

The above results are based on the fact that if a deontic statement is labelled in then any conflicting statement is labelled ni or niund. Let us now inspect probabilistic relationships between obligations and permissions. We can first note that if an obligation $\mathsf{O}\gamma$ is labelled in then the implied permission $\neg\mathsf{O}\overline{\gamma}$ (i.e. $\mathsf{P}\gamma$) is also labelled in.

Proposition 3.3. *Let* L *be a legit grounded* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$*-labelling of an argumentation graph constructed from a permissive defeasible theory* T*,* $\mathscr{S}$ *a set of literal statements such that* $\mathscr{S} = \{p, \neg p \mid p \in \mathrm{Prop}(T)\}$*, and* K *a bivalent* $\{\mathsf{in}, \mathsf{ni}\}$*-labelling or trivalent* $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$*-labelling of* $\mathscr{S}$ *and from* $\{\mathrm{L}\}$*. For any* $\gamma \in \mathscr{S}$*, if* $\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{in}$ *then* $\mathrm{K}(\mathrm{L}, \neg\mathsf{O}\overline{\gamma}) = \mathsf{in}$*.*

Proof. Given an argumentation graph G constructed from a permissive defeasible theory, if $\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{in}$, then there exist an argument $A \in \mathscr{A}_G$ and $B \in \mathscr{A}_G$ such that $\mathrm{con}(A) = \mathsf{O}\gamma$ and $B : A \Rightarrow_{\mathsf{d\_}\gamma} \mathsf{P}\gamma$. Let $\mathscr{A}^{\rightsquigarrow} \subseteq \mathscr{A}_G$ be the set of attackers of A, $\mathscr{A}^{\leftrightsquigarrow} \subseteq \mathscr{A}_G$ the set of arguments attacked by A, and $\mathscr{B}^{\rightsquigarrow} \subseteq \mathscr{A}_G$ the set of attackers of B. We have that $\mathscr{B}^{\rightsquigarrow} \subseteq \mathscr{A}^{\rightsquigarrow} \cup \mathscr{A}^{\leftrightsquigarrow}$. By Definition 2.16, if $\mathrm{L}(A) = \mathsf{IN}$ then for any argument $C \in \mathscr{A}^{\rightsquigarrow} \cup \mathscr{A}^{\leftrightsquigarrow}$ $\mathrm{L}(C) = \mathsf{OUT}$ or $\mathrm{L}(C) = \mathsf{OFF}$. Since $\mathscr{B}^{\rightsquigarrow} \subseteq \mathscr{A}^{\rightsquigarrow} \cup \mathscr{A}^{\leftrightsquigarrow}$, for any argument $C \in \mathscr{B}^{\rightsquigarrow}$ $\mathrm{L}(C) = \mathsf{OUT}$ or $\mathrm{L}(C) = \mathsf{OFF}$, and thus $\mathrm{L}(B) = \mathsf{IN}$. Therefore, if $\mathrm{L}(A) = \mathsf{IN}$ then $\mathrm{L}(B) = \mathsf{IN}$, and thus if $\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{in}$ then $\mathrm{K}(\mathrm{L}, \mathsf{P}\gamma) = \mathsf{in}$, i.e. if $\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{in}$ then $\mathrm{K}(\mathrm{L}, \neg\mathsf{O}\overline{\gamma}) = \mathsf{in}$. □

In our probabilistic framework, Proposition 3.3 tells us that for any outcome where $\mathsf{O}\gamma$ is labelled in, the implied permission $\mathsf{P}\gamma$ is also labelled in. Consequently, the probability that something is permitted given that it is obligatory is one, and the probability that something is permitted is necessarily greater or equal than the probability that it is obligatory.

Proposition 3.4. *Let* $\langle G, \langle \Omega, \mathscr{F}, P \rangle \rangle$ *be a legit grounded PLF where* G *is constructed from a permissive defeasible theory* T*, and* $\mathscr{S}$ *is a set of literal statements such that* $\mathscr{S} = \{p, \neg p \mid p \in \mathrm{Prop}(T)\}$*. For any* $\gamma \in \mathscr{S}$*:*

$$P(K_{\mathsf{P}\gamma} = \mathsf{in} \mid K_{\mathsf{O}\gamma} = \mathsf{in}) = 1.$$

Proof. From Proposition 3.3, if $\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{in}$ then $\mathrm{K}(\mathrm{L}, \neg\mathsf{O}\overline{\gamma}) = \mathsf{in}$. By Proposition 2.2 $P(K_{\mathsf{P}\gamma} = \mathsf{in} \mid K_{\mathsf{O}\gamma} = \mathsf{in}) = 1$. □

Corollary 3.2. *Let $\langle G, \langle \Omega, \mathscr{F}, P\rangle\rangle$ be a legit grounded PLF where G is constructed from a permissive defeasible theory T, and $\mathscr{S}$ is a set of literal statements such that $\mathscr{S} = \{p, \neg p \mid p \in \mathrm{Prop}(T)\}$. For any $\gamma \in \mathscr{S}$:*

$$P(K_{\mathsf{O}\gamma} = \mathsf{in}) \leq P(K_{\mathsf{P}\gamma} = \mathsf{in}).$$

Proof. From Proposition 3.3, if $\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{in}$ then $\mathrm{K}(\mathrm{L}, \neg\mathsf{O}\overline{\gamma}) = \mathsf{in}$. By Corollary 2.1, $P(K_{\mathsf{O}\gamma} = \mathsf{in}) \leq P(K_{\mathsf{P}\gamma} = \mathsf{in})$. □

We can also observe that any backgrounded defeasible theory along with a trivalent labelling semantics leads to a third interpretation of the principle of prohibition in terms of labelling: if something is not prohibited (the prohibition is labelled niund) then it is permitted (the permission is labelled in).

Proposition 3.5. *Let* L *be a legit grounded* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$*-labelling of an argumentation graph constructed from a permissive defeasible theory T, $\mathscr{S}$ a set of literal statements such that $\mathscr{S} = \{p, \neg p \mid p \in \mathrm{Prop}(T)\}$. Let* K *be a trivalent* $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$*-labelling of $\mathscr{S}$ and from* $\{\mathrm{L}\}$*. For any $\gamma \in \mathscr{S}$: if* $\mathrm{K}(\mathrm{L}, \mathsf{F}\gamma) = \mathsf{niund}$ *then* $\mathrm{K}(\mathrm{L}, \mathsf{P}\gamma) = \mathsf{in}$.

Proof. Given an argumentation graph constructed from a permissive defeasible theory T, for any $\gamma \in \mathscr{S} = \{p, \neg p \mid p \in \mathrm{Prop}(T)\}$, there exists a unique argument $W : \sim \mathsf{F}\gamma \Rightarrow_{\mathsf{k\_}\gamma} \mathsf{P}\gamma$. All attackers of W are arguments whose conclusion is $\mathsf{F}\gamma$ (i.e. $\mathsf{O}\overline{\gamma}$). If $K(\mathrm{L}, \mathsf{O}\overline{\gamma}) = \mathsf{niund}$ then all attackers of W are OUT or OFF, and thus W is labelled IN, and $\mathsf{P}\gamma$ is labelled in. i.e. $K(\mathrm{L}, \neg\mathsf{O}\overline{\gamma}) = \mathsf{in}$. Therefore, if $K(\mathrm{L}, \mathsf{F}\gamma) = \mathsf{niund}$ then $K(\mathrm{L}, \mathsf{P}\gamma) = \mathsf{in}$. □

Corollary 3.3. *Let* L *be a legit grounded* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$*-labelling of an argumentation graph constructed from a permissive defeasible theory T, and $\mathscr{S}$ a set of literal statements such that $\mathscr{S} = \{p, \neg p \mid p \in \mathrm{Prop}(T)\}$. Let* K *be a trivalent* $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$*-labelling of $\mathscr{S}$ and from* $\{\mathrm{L}\}$*. For any $\gamma \in \mathscr{S}$,* $\mathrm{K}(\mathrm{L}, \mathsf{F}\gamma) = \mathsf{niund}$ *iff* $\mathrm{K}(\mathrm{L}, \mathsf{P}\gamma) = \mathsf{in}$.

Proof. From Proposition 3.1, if $\mathrm{K}(\mathrm{L}, \neg\mathsf{O}\gamma) = \mathsf{in}$ then $\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{niund}$, that is, if $\mathrm{K}(\mathrm{L}, \mathsf{P}\gamma) = \mathsf{in}$ then $\mathrm{K}(\mathrm{L}, \mathsf{F}\gamma) = \mathsf{niund}$. In addition, from Proposition 3.5, if $\mathrm{K}(\mathrm{L}, \mathsf{F}\gamma) = \mathsf{niund}$ then $\mathrm{K}(\mathrm{L}, \mathsf{P}\gamma) = \mathsf{in}$. Hence, $\mathrm{K}(\mathrm{L}, \mathsf{F}\gamma) = \mathsf{niund}$ iff $\mathrm{K}(\mathrm{L}, \mathsf{P}\gamma) = \mathsf{in}$. □

In our probabilistic framework, Corollary 3.3 implies that for any outcome, something is permitted if, and only if, it is not prohibited. Consequently, the probability that something is permitted and the probability that it is not prohibited are equal.

Proposition 3.6. *Let $\langle G, \langle \Omega, \mathscr{F}, P\rangle\rangle$ be a legit grounded PLF where G is constructed from a permissive defeasible theory T, $\mathscr{S}$ a set of literal statements such that $\mathscr{S} = \{p, \neg p \mid p \in \mathrm{Prop}(T)\}$. For any $\gamma \in \mathscr{S}$:*

$$P(K_{\mathsf{F}\gamma} = \mathsf{niund}) = P(K_{\mathsf{P}\gamma} = \mathsf{in}).$$

Now, on the basis of legit grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$-labellings, jurists may argue that bivalent $\{\mathsf{in}, \mathsf{ni}\}$-labellings are not satisfactory. Given an argumentation graph from any backgrounded defeasible theory (which can contain arguments supporting implicit permissions due to background rules), it may be the case that all arguments are labelled UND. Consequently, using a bivalent $\{\mathsf{in}, \mathsf{ni}\}$-labelling, deontic statements $\mathsf{O}\gamma$, $\mathsf{F}\gamma$ and $\mathsf{P}\gamma$ may be labelled ni. This is illustrated in Example 4 below, and suggests that bivalent labellings are insufficient to capture normative completeness. For this reason, we discard bivalent $\{\mathsf{in}, \mathsf{ni}\}$-labelling semantics, and instead propose trivalent $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$-labelling semantics to cater for deontic reasoning to address normative completeness.

Example 4 (continued)**.** *Let us consider the legit grounded* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$*-labelling drawn in Figure 8.*

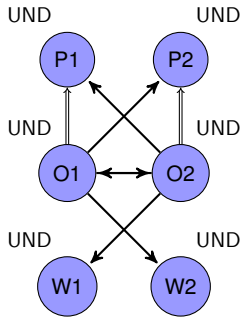

Figure 8: A legit grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$-labelling.

First, naive common sense (bivalent) reasoning can be modelled via the following acceptance bivalent $\{\mathsf{in}, \mathsf{ni}\}$*-labelling:*

$$\langle \emptyset, \{\mathsf{Oa}, \mathsf{Pa}, \mathsf{O}\neg\mathsf{a}, \neg\mathsf{Pa}\} \rangle .$$

However, this bivalent labelling is problematic from a legal stance because statement a *is here not obligated, permitted or prohibited. To address this gap, we can employ a trivalent* $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$*-labelling:*

$$\langle \emptyset, \{\mathsf{Oa}, \mathsf{Pa}, \mathsf{O}\neg\mathsf{a}, \neg\mathsf{Pa}\}, \emptyset \rangle$$

according to which the deontic status of a *is undecided.* □

More formally, the definition of normative gaps — which represent statements that are not permitted, obliged or prohibited — as we may conceive them in terms of statement labellings, depends on whether bivalent $\{\mathsf{in}, \mathsf{ni}\}$-labellings or trivalent $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$-labellings are employed.

Definition 3.16. *Let* L *be a grounded* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$*-labelling of an argumentation graph constructed from a defeasible theory* T*, and* $\mathscr{S}$ *the set of literal statements such that* $\mathscr{S} = \{p, \neg p \mid p \in \mathrm{Prop}(T)\}$.

- *Let* K *denote a bivalent* $\{\mathsf{in}, \mathsf{ni}\}$*-labelling of* $\mathscr{S}$ *and from* $\{\mathrm{L}\}$*. There is a* $\{\mathsf{in}, \mathsf{ni}\}$***-labelling normative gap*** *iff there exists a literal statement* $\gamma \in \mathscr{S}$ *such that*

$$\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{ni} \textit{ and } \mathrm{K}(\mathrm{L}, \mathsf{O}\overline{\gamma}) = \mathsf{ni} \textit{ and } \mathrm{K}(\mathrm{L}, \neg\mathsf{O}\overline{\gamma}) = \mathsf{ni}.$$

- *Let* K' *denote a trivalent* $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$*-labelling of* $\mathscr{S}$ *and from* $\{\mathrm{L}\}$*. There is a* $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$***-labelling normative gap*** *iff there exists a literal statement* $\gamma \in \mathscr{S}$ *such that*

$$\mathrm{K}'(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{niund} \textit{ and } \mathrm{K}'(\mathrm{L}, \mathsf{O}\overline{\gamma}) = \mathsf{niund} \textit{ and } \mathrm{K}'(\mathrm{L}, \neg\mathsf{O}\overline{\gamma}) = \mathsf{niund}.$$

In the case of a legit grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$-labelling of an argumentation graph constructed from a permissive defeasible theory, the definition of gaps may be simplified. We can remark, for example, that the conjunction in the first item can be simplified into '$\mathrm{K}(\mathrm{L}, \mathsf{O}\overline{\gamma}) = \mathsf{ni}$ and $\mathrm{K}(\mathrm{L}, \neg\mathsf{O}\overline{\gamma}) = \mathsf{ni}$', because $\mathrm{K}(\mathrm{L}, \neg\mathsf{O}\overline{\gamma}) = \mathsf{ni}$ implies $\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{ni}$ (by Proposition 3.3). In some other deontic frameworks, however, these two conjuncts may hold while $\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) = \mathsf{in}$, suggesting a gap where in fact there is none. Hence, a conjunction specifying the labelling of the three deontic statuses may appear as a stronger definition of gaps, and eventually it may be easier to grasp.

In the case of a legit grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$-labelling of an argumentation graph constructed from a permissive defeasible theory, and as illustrated in Example 4, bivalent $\{\mathsf{in}, \mathsf{ni}\}$-labellings may lead to $\{\mathsf{in}, \mathsf{ni}\}$-labelling normative gaps, whereas, trivalent $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$-labellings have no normative gaps, i.e. they provide normative completeness.

Theorem 3.1. *Let* L *be a legit grounded* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$*-labelling of an argumentation graph constructed from a permissive defeasible theory* T*,* $\mathscr{S}$ *a set of literal statements such that* $\mathscr{S} = \{p, \neg p \mid p \in \mathrm{Prop}(T)\}$*, and* K *a trivalent* $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$*-labelling of* $\mathscr{S}$ *and from* $\{\mathrm{L}\}$*. For any* $\gamma \in \mathscr{S}$*:*

$$\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) \neq \mathsf{niund} \textit{ or } \mathrm{K}(\mathrm{L}, \mathsf{O}\overline{\gamma}) \neq \mathsf{niund} \textit{ or } \mathrm{K}(\mathrm{L}, \neg\mathsf{O}\overline{\gamma}) \neq \mathsf{niund}.$$

Proof. Any trivalent $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$-labelling K is a total function. Thus, for any $\gamma \in \mathscr{S}$, there are three cases which must be considered: 1. $\mathrm{K}(\mathrm{L}, \mathsf{O}\overline{\gamma}) = \mathsf{in}$, 2. $\mathrm{K}(\mathrm{L}, \mathsf{O}\overline{\gamma}) = \mathsf{und}$, and 3. $\mathrm{K}(\mathrm{L}, \mathsf{O}\overline{\gamma}) = \mathsf{niund}$. If $\mathrm{K}(\mathrm{L}, \mathsf{O}\overline{\gamma}) = \mathsf{niund}$ then $\mathrm{K}(\mathrm{L}, \mathsf{P}\gamma) = \mathsf{in}$ (Proposition 3.5), and thus in the last case, $\mathrm{K}(\mathrm{L}, \mathsf{P}\gamma) = \mathsf{in}$, i.e. $\mathrm{K}(\mathrm{L}, \neg\mathsf{O}\overline{\gamma}) = \mathsf{in}$. Therefore in any case, $\mathrm{K}(\mathrm{L}, \mathsf{O}\gamma) \neq \mathsf{niund}$ or $\mathrm{K}(\mathrm{L}, \mathsf{O}\overline{\gamma}) \neq \mathsf{niund}$ or $\mathrm{K}(\mathrm{L}, \neg\mathsf{O}\overline{\gamma}) \neq \mathsf{niund}$. □

Hence, trivalent $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$-labellings address normative completeness by means of the status 'undecided' for deontic statements. Eventually, such undecidedness can be disentangled in various ways, typically by a competent authority such as a judge. The trivalent labellings make such undecidedness explicit.

In our probabilistic framework, Theorem 3.1 implies that for any outcome there is no normative gap, and therefore the probability of a normative gap is zero.

Corollary 3.4. *Let* $\langle G, \langle \Omega, \mathscr{F}, P\rangle\rangle$ *be a legit grounded PLF where G is constructed from a permissive defeasible theory* T*,* $\mathscr{S}$ *a set of literal statements such that* $\mathscr{S} = \{p, \neg p \mid p \in \mathrm{Prop}(T)\}$*. For any* $\gamma \in \mathscr{S}$*:*

$$P(K_{\mathsf{O}\gamma} = \mathsf{niund},\ K_{\mathsf{F}\gamma} = \mathsf{niund},\ K_{\mathsf{P}\gamma} = \mathsf{niund}) = 0.$$

The results presented above hold for any permissive theory. Consequently, they hold for any backgrounded defeasible theory of a (foreground) defeasible theory with a prima facie or Kelsenian permissive set of background defeasible rule schemata. In general, it turns out that both sets yield the same trivalent $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$-labelling. To show this result, let us first introduce some definitions for ease of presentation.

Definition 3.17. *Two arguments A and B are* ***k-equivalent arguments*** *iff*
- *if A is of the form* $A: \ \Rightarrow_{\mathsf{p\_}\gamma} \mathsf{P}\gamma$ *then B is of the form* $B: \ \sim \mathsf{F}\gamma \Rightarrow_{\mathsf{k\_}\gamma} \mathsf{P}\gamma$*, and*
- *if A is of the form* $A: A_1, \ldots A_n, \sim \varphi_1, \ldots, \sim \varphi_m \Rightarrow_r \varphi$ $(r \neq \mathsf{p\_}\gamma)$ *then B is of the form* $B: B_1, \ldots, B_n, \sim \varphi_1, \ldots, \sim \varphi_m \Rightarrow_r \varphi$*, where* A_i *and* B_i $(1 \leq i \leq n)$ *are k-equivalent.*

Definition 3.18. *Argumentation graphs* $G1$ *and* $G2$ *are* ***k-equivalent argumentation graphs*** *iff there exists a bijection* $\mathsf{f}: \mathscr{A}_{G1} \rightarrow \mathscr{A}_{G2}$ *such that for any argument* $A \in \mathscr{A}_{G1}$*:*
- $\mathsf{f}(A) = B$ *iff A and B are k-equivalent, and*
- $(A, A') \in \rightsquigarrow_{G1}$ *iff* $(\mathsf{f}(A), \mathsf{f}(A')) \in \rightsquigarrow_{G2}$.

Definition 3.19. *Let*
- $G1$ *and* $G2$ *be two argumentation graphs;*
- $H1$ *(*$H2$ *resp.) be a subargument-complete subgraph of* $G1$ *(*$G2$ *resp.);*
- L_1 *(*L_2 *resp.) be a legit grounded* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$*-labelling of the argumentation graph* $G1$ *(*$G2$ *resp.) with respect to subgraph* $H1$ *(*$H2$ *resp.);*

L_1 *and* L_2 *are* ***k-equivalent labellings*** *iff* $G1$ *and* $G2$ *are k-equivalent, and* $H1$ *and* $H2$ *are k-equivalent.*

The work in [9] showed that both types of permissive theories yield the same trivalent $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$-labelling for $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$-labellings. Theorem 3.2 is a generalisation of this result to grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$-labellings.

Theorem 3.2. *Let*
- T *be a (foreground) defeasible theory;*
- U *be the backgrounded defeasible theory of* T *with a prima facie permissive set of background defeasible rule schemata;*
- V *be the backgrounded defeasible theory of* T *with a Kelsenian permissive set of background defeasible rule schemata;*
- L_U *(*L_V *resp.) be a legit grounded* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$*-labelling of the argumentation graph constructed from U (V resp.);*
- $\mathscr{S}$ *be a set of literal statements such that* $\mathscr{S} = \{p, \neg p \mid p \in \mathrm{Prop}(T)\}$*, and* K *a trivalent* $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$*-labelling of* $\mathscr{S}$ *and from* $\{\mathrm{L}\}$.

For any $\gamma \in \mathscr{S}$*, if* L_U *and* L_V *are k-equivalent then*

$$\mathrm{K}(\mathrm{L}_U, \gamma) = \mathrm{K}(\mathrm{L}_V, \gamma) \text{ and } \mathrm{K}(\mathrm{L}_U, \mathsf{O}\gamma) = \mathrm{K}(\mathrm{L}_V, \mathsf{O}\gamma) \text{ and } \mathrm{K}(\mathrm{L}_U, \neg\mathsf{O}\gamma) = \mathrm{K}(\mathrm{L}_V, \neg\mathsf{O}\gamma).$$

Proof. Let G_U (G_V resp.) denote the argumentation graph constructed from U (V resp.), and H_U (H_V resp.) a subgraph of G_U (G_V resp.) such that L_U (L_V resp.) the legit grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$-labelling with respect to H_U (H_V resp.). If L_U and L_V are k-equivalent then G_U and G_V are k-equivalent, and H_U and H_V are k-equivalent. By Theorem 3.2 in [9], for any $\gamma \in \mathscr{S}$: $\mathrm{K}(\mathrm{L}_U, \gamma) = \mathrm{K}(\mathrm{L}_V, \gamma)$ and $\mathrm{K}(\mathrm{L}_U, \mathsf{O}\gamma) = \mathrm{K}(\mathrm{L}_V, \mathsf{O}\gamma)$ and $\mathrm{K}(\mathrm{L}_U, \neg\mathsf{O}\gamma) = \mathrm{K}(\mathrm{L}_V, \neg\mathsf{O}\gamma)$. □

Theorem 3.2 extends representation results to our probabilistic setting. The principle of prohibition 'anything that is not prohibited is permitted' as a schema $\mathsf{k\_}\gamma: \sim \mathsf{O}\overline{\gamma} \Rightarrow \mathsf{P}\gamma$ or a principle such as 'anything is permitted prima facie' as a schema $\mathsf{p\_}\gamma: \ \Rightarrow \mathsf{P}\gamma$ are two alternatives to cater for normative completeness, and both alternatives lead to the same results in terms of statement labellings in the case of grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$-labellings [9]. Theorem 3.2 extends the result to legit grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$-labellings and thus to our probabilistic setting.

4. Worked Example

To illustrate our system, let us reappraise the policy stating that vehicles are forbidden to enter a park unless there is an emergency. This policy and its associated assumptions may be formalised in various ways. We illustrate our system with one option which is developed in the remainder of this section.

4.1. Backgrounded defeasible theory

We assume that there is a vehicle at the entrance of the park, and that there may *possibly* be an emergency. Let us capture this with the foreground defeasible theory $\langle\{\mathsf{rv}, \mathsf{re}, \mathsf{r\bar{e}}, \mathsf{r}\}, \emptyset, \emptyset\rangle$ where

$\mathsf{rv} \quad \Rightarrow \mathsf{vehi}$
$\mathsf{re} \quad \Rightarrow \mathsf{emer}$
$\mathsf{r\bar{e}} \quad \Rightarrow \neg\mathsf{emer}$
$\mathsf{r}: \quad \mathsf{vehi}, \sim \mathsf{emer} \Rightarrow \mathsf{Fenter}$

The foreground theory can be then backgrounded to yield a backgrounded theory featuring, amongst others, background rules as exposed in Example 3.

4.2. Argument and argumentation graph construction

Let us consider a Kelsenian permissive set of background defeasible rule schemata. Accordingly, we can build the following arguments from background rules:

$\mathsf{W1}: \sim \mathsf{Fvehi} \Rightarrow_{\mathsf{k\_vehi}} \mathsf{Pvehi}$ $\quad$ $\mathsf{W4}: \sim \mathsf{F}\neg\mathsf{vehi} \Rightarrow_{\mathsf{k\_}\neg\mathsf{vehi}} \mathsf{P}\neg\mathsf{vehi}$
$\mathsf{W2}: \sim \mathsf{Femer} \Rightarrow_{\mathsf{k\_emer}} \mathsf{Pemer}$ $\quad$ $\mathsf{W5}: \sim \mathsf{F}\neg\mathsf{emer} \Rightarrow_{\mathsf{k\_}\neg\mathsf{emer}} \mathsf{P}\neg\mathsf{emer}$
$\mathsf{W3}: \sim \mathsf{Fenter} \Rightarrow_{\mathsf{k\_enter}} \mathsf{Penter}$ $\quad$ $\mathsf{W6}: \sim \mathsf{F}\neg\mathsf{enter} \Rightarrow_{\mathsf{k\_}\neg\mathsf{enter}} \mathsf{P}\neg\mathsf{enter}$

In addition, we can build the following arguments:

$\mathsf{A1}: \Rightarrow_{\mathsf{rv}} \mathsf{vehi}$ $\quad$ $\mathsf{B1}: \Rightarrow_{\mathsf{re}} \mathsf{emer}$
$\mathsf{A2}: \mathsf{A1}, \sim \mathsf{emer} \Rightarrow_{\mathsf{r}} \mathsf{Fenter}$ $\quad$ $\mathsf{C1}: \Rightarrow_{\mathsf{r\bar{e}}} \neg\mathsf{emer}$
$\mathsf{A3}: \mathsf{A2} \Rightarrow_{\mathsf{d\_}\neg\mathsf{enter}} \mathsf{P}\neg\mathsf{enter}$

Consequently, we can build the argumentation graph G such that: $\mathscr{A}_{\mathsf{G}} = \{\mathsf{A1}, \mathsf{A2}, \mathsf{A3}, \mathsf{B1}, \mathsf{C1}, \mathsf{W1}, \mathsf{W2}, \mathsf{W3}, \mathsf{W4}, \mathsf{W5}, \mathsf{W6}\}$, and $\rightsquigarrow_{\mathsf{G}} = \{(\mathsf{B1}, \mathsf{C1}), (\mathsf{C1}, \mathsf{B1}), (\mathsf{B1}, \mathsf{A2}), (\mathsf{B1}, \mathsf{A3}), (\mathsf{A2}, \mathsf{W3})\}$, and $\mapsto_{\mathsf{G}} = \{(\mathsf{A1}, \mathsf{A2}), (\mathsf{A2}, \mathsf{A3})\}$ (see Figure 9).

We note that we have built arguments to support weak/doctrinal permissions, thus we can argue and present full-fledged arguments about such permissions (here arguments $\mathsf{W1} \ldots \mathsf{W6}$).

4.3. Argument labelling

The grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$-labelling of argumentation graph G is illustrated in Figure 9, and all legit grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$-labelling outcomes which are deemed possible are exposed in Table 2.

We assume that the cyber-physical system is such that arguments B1 and C1 cannot both be labelled OFF, so that it is always indicated whether it is the case that there is an emergency or no emergency, or whether the case is undecided. We also assume that rule r is always applied when its antecedents hold. Finally we adopt the principle of indifference on

all possible legit grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$-labelling outcomes, so that they are equally distributed as in Table 2.

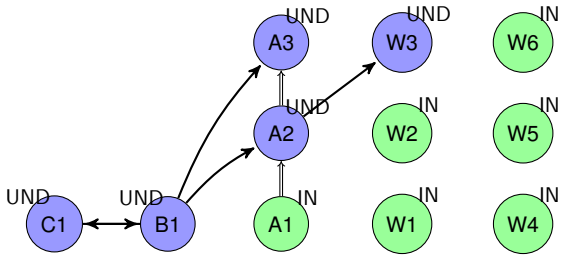

Figure 9: Grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$-labelling of argumentation graph G.

A1	A2	A3	B1	C1	W1	W2	W3	W4	W5	W6	$P(\cdot)$
OFF	OFF	OFF	IN	OFF	IN	IN	IN	IN	IN	IN	1/6
OFF	OFF	OFF	OFF	IN	IN	IN	IN	IN	IN	IN	1/6
OFF	OFF	OFF	UND	UND	IN	IN	IN	IN	IN	IN	1/6
IN	OUT	OUT	IN	OFF	IN	IN	IN	IN	IN	IN	1/6
IN	IN	IN	OFF	IN	IN	IN	OUT	IN	IN	IN	1/6
IN	UND	UND	UND	UND	IN	IN	UND	IN	IN	IN	1/6

Table 2: Legit grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$-labellings of argumentation graph G with non-zero probabilities.

Given probabilities in Table 2, we can easily see that for example

$$P(L_{\mathsf{A3}} = \mathsf{IN}) = 1/6 \tag{7}$$

$$P(L_{\mathsf{A3}} = \mathsf{IN} \mid L_{\mathsf{B1}} = \mathsf{IN}) = 0. \tag{8}$$

In words, the marginal probability that argument A3 is labelled IN is 1/6; and the probability that argument A3 is labelled IN, given that B1 is labelled IN, is zero.

4.4. Statement labelling

Let L denote the grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$-labelling of argumentation graph G, i.e. the labelling in the last row in Table 2. From this, we can obtain the bivalent $\{\mathsf{in}, \mathsf{ni}\}$-labelling and trivalent $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$-labelling as exposed in Table 3.

	vehi	¬vehi	emer	¬emer	enter	¬enter
K(L,·)	in	ni	ni	ni	ni	ni
K(L,·)	in	niund	und	und	niund	niund

	Ovehi	O¬vehi	Oemer	O¬emer	Oenter	O¬enter
K(L,·)	ni	ni	ni	ni	ni	ni
K(L,·)	niund	niund	niund	niund	niund	und

	Pvehi	P¬vehi	Pemer	P¬emer	Penter	P¬enter
K(L,·)	in	in	in	in	ni	in
K(L,·)	in	in	in	in	und	in

Table 3: Bivalent $\{\mathsf{in}, \mathsf{ni}\}$-labelling and trivalent $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$-labelling.

We can remark that the $\{\mathsf{in}, \mathsf{ni}\}$-bivalent labelling results in a normative gap (enter is neither obligatory nor prohibited nor permitted), whereas the trivalent $\{\mathsf{in}, \mathsf{und}, \mathsf{niund}\}$-labelling fills this gap by labelling the permission to enter as undecided.

More generally, for every legit grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$-labelling, we can draw a statement labelling, and from all the statement labellings we can determine probability values on the acceptance statuses of statements. For instance, from the argument labelling probabilities as given in Table 2, we can compute that:

$$P(K_{\mathsf{vehi}} = \mathsf{in}) = 1/2 \tag{9}$$

$$P(K_{\mathsf{Fenter}} = \mathsf{in}) = 1/6 \tag{10}$$

$$P(K_{\mathsf{Fenter}} = \mathsf{in} \mid K_{\mathsf{vehi}} = \mathsf{in}) = 1/3. \tag{11}$$

In words, the marginal probability that there is a vehicle is 1/2, and the marginal probability that there is a detached prohibition from entering the park is 1/6. If it is accepted that there is a vehicle (a vehicle is perceived), then the probability that the vehicle should not enter is 1/3 (there may be an emergency, or the case is undecided). Such likelihoods may be useful, for example, in determining whether to monitor specific portions of the system.

4.5. Violation and contrary-to-duty obligation

Let us extend the example with the formalisation of a violation and a contrary-to-duty obligation. Such obligations can be a pitfall for deontic formalisms which have a more sophisticated conception of deontic modalities [38, 39], and we would like to illustrate how such obligations can be handled in our probabilistic deontic argumentation framework.

Let us suppose that the park policy also states that a violation of the prohibition would be sanctioned by a fine (the amount does not matter for our purposes). To capture such a policy, we can add the following rules.

v :	Fenter, enter $\Rightarrow$ violation	$\bar{\mathsf{v}}$:	$\Rightarrow \neg$violation
f :	violation $\Rightarrow$ fine	$\bar{\mathsf{f}}$:	$\Rightarrow \neg$fine

such that $\mathsf{v} \succ \bar{\mathsf{v}}$ and $\mathsf{f} \succ \bar{\mathsf{f}}$. Rules $\bar{\mathsf{v}}$ and $\bar{\mathsf{f}}$ specify that, by default, we can derive that there is neither violation nor fine, unless the contrary is shown.

Furthermore, assume that the park's management adds the following contrary-to-duty obligation: if the prohibition is violated then one should stop driving (in the park). We can thus add the following rule.

$\mathsf{s}:\quad \mathsf{violation} \Rightarrow \mathsf{Ostop}$

A sequence of compensatory obligations can be added along similar lines. We also assume that a vehicle enters in the park:

$\mathsf{e}:\quad \Rightarrow \mathsf{enter}$

In addition to the previous arguments, we can thus build the following arguments (amongst others).

$\mathsf{E1}: \quad \Rightarrow_{\mathsf{e}} \mathsf{enter}$ $\qquad$ $\mathsf{V2}: \quad \Rightarrow_{\overline{\mathsf{v}}} \neg\mathsf{violation}$

$\mathsf{V1}: \quad \mathsf{E1}, \mathsf{A2} \Rightarrow_{\mathsf{r}} \mathsf{violation}$ $\qquad$ $\mathsf{F2}: \quad \Rightarrow_{\overline{\mathsf{f}}} \neg\mathsf{fine}$

$\mathsf{F1}: \quad \mathsf{V1} \Rightarrow_{\mathsf{f}} \mathsf{fine}$

A1	A2	A3	B1	C1	W1	W2	W3	W4	W5	W6	E1	V1	$P(\cdot)$
OFF	OFF	OFF	IN	OFF	IN	IN	IN	IN	IN	IN	OFF	OFF	1/9
OFF	OFF	OFF	OFF	IN	IN	IN	IN	IN	IN	IN	OFF	OFF	1/9
OFF	OFF	OFF	UND	UND	IN	IN	IN	IN	IN	IN	OFF	OFF	1/9
IN	OUT	OUT	IN	OFF	IN	IN	IN	IN	IN	IN	OFF	OFF	1/9
IN	IN	IN	OFF	IN	IN	IN	OUT	IN	IN	IN	OFF	OFF	1/9
IN	UND	UND	UND	UND	IN	IN	UND	IN	IN	IN	OFF	OFF	1/9
IN	OUT	OUT	IN	OFF	IN	IN	IN	IN	IN	IN	IN	OUT	1/9
IN	IN	IN	OFF	IN	IN	IN	OUT	IN	IN	IN	IN	IN	1/9
IN	UND	UND	UND	UND	IN	IN	UND	IN	IN	IN	IN	UND	1/9

Table 4: Legit grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$-labellings with non-zero probabilities (not all possible arguments are labelled due to the lack of space).

Again, for every legit grounded $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}, \mathsf{OFF}\}$-labelling of the resulting argumentation graph, we can draw a statement labelling, and from all the statement labellings we can determine probability values on the labellings of statements. For instance, from the argument labelling probabilities given in Table 4, and assuming that rules v, f, $\overline{\mathsf{v}}$, $\overline{\mathsf{f}}$ and s apply in all outcomes where their antecedents hold, and assuming that there is no doubt that a vehicle has entered in the park, we can compute the following probabilities.

$P(K_{\mathsf{violation}} = \mathsf{in} \mid K_{\mathsf{vehi}} = \mathsf{in}) = 1/6$ $\qquad$ $P(K_{\mathsf{violation}} = \mathsf{in} \mid K_{\mathsf{vehi}} = \mathsf{in}, K_{\mathsf{enter}} = \mathsf{in}) = 1/3$

$P(K_{\mathsf{fine}} = \mathsf{in} \mid K_{\mathsf{vehi}} = \mathsf{in}) = 1/6$ $\qquad$ $P(K_{\mathsf{fine}} = \mathsf{in} \mid K_{\mathsf{vehi}} = \mathsf{in}, K_{\mathsf{enter}} = \mathsf{in}) = 1/3$

$P(K_{\mathsf{Ostop}} = \mathsf{in} \mid K_{\mathsf{vehi}} = \mathsf{in}) = 1/6$ $\qquad$ $P(K_{\mathsf{Ostop}} = \mathsf{in} \mid K_{\mathsf{vehi}} = \mathsf{in}, K_{\mathsf{enter}} = \mathsf{in}) = 1/3$

In words, if a vehicle is perceived, then a violation, its sanction, and the associated contrary-to-duty obligation will occur with a probability value 1/6. If, in addition the vehicle has entered the park, then the probability values increase to 1/3.

5. Conclusion

We have introduced a probabilistic deontic argumentation framework to reason about probability values attached to deontic statements and arguments supporting these statements. Given a set of norms and a state of affairs modelled as a defeasible theory, the framework allows one to associate acceptance statuses of deontic statements and related violations with probability values, in a principled way.

To do so, we have combined a deontic argumentation approach where deontic principles are reified [9] and a probabilistic argumentation approach where argument labellings are associated with probabilities values [3]. The deontic argumentation composition relies on common concepts taken from computational models of argument: rule-based arguments, argumentation graphs, argument labelling semantics and statement labelling semantics. In this framework, normative completeness is addressed by reifying the principle of prohibition. This deontic composition is then combined with probabilistic labellings used in probabilistic argumentation, enabling us to associate statement acceptance statuses with probability values.

We have learnt that the construction of a probabilistic deontic argumentation framework can be achieved by using standard constructs from computational models of argument and probabilistic labellings. Then, we could have appreciated that argumentation properties of the deontic framework can be coupled with probabilistic counterparts, i.e. relationships amongst acceptance statuses of deontic statements can be understood in meaningful relationships of probabilities of acceptance statuses of these statements.

Throughout the paper we have identified several potential directions for future research. In addition, we note that no algorithms to learn or compute probability statuses of arguments and statements have been considered, and exploring efficient algorithms for this purpose would be interesting. Finally, throughout the paper, we have avoided relating the probabilistic framework with possible probability interpretations (such as classical, frequentist, or Bayesian views on these matters), and thus it may be interesting to investigate the impact of such interpretations on probabilistic deontic issues.

References

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

[2] A. Hunter, M. Thimm, Probabilistic reasoning with abstract argumentation frameworks, Journal of Artificial Intelligence Research 59 (2017) 565–611.

[3] R. Riveret, P. Baroni, Y. Gao, G. Governatori, A. Rotolo, G. Sartor, A labelling framework for probabilistic argumentation, Annals of Mathematics and Artificial Intelligence 83 (1) (2018) 21–71.

[4] H. Hart, The Concept of Law, Oxford University Press, 1961.

[5] J. Horty, Nonmonotonic foundations for deontic logic, in: Defeasible Deontic Logic, Kluwer Academic Publishers, 1997, pp. 17–44.

[6] B. Liao, N. Oren, L. Van Der Torre, S. Villata, Prioritized norms in formal argumentation, Journal of Logic and Computation 29 (2) (2018) 215–240.

[7] M. Beirlaen, J. Heyninck, C. Straßer, Structured argumentation with prioritized conditional obligations and permissions, Journal of Logic and Computation 29 (2) (2019) 187–214.

[8] G. Pigozzi, L. van der Torre, Arguing about constitutive and regulative norms, Journal of Applied Non-Classical Logics 28 (2-3) (2018) 189–217.

[9] R. Riveret, A. Rotolo, G. Sartor, A deontic argumentation framework towards doctrine reification, IfCoLog Journal of Logics and their Applications 6 (5) (2019) 903–939.

[10] L. L. Fuller, Positivism and fidelity to law: A reply to Professor Hart, Harvard Law Review 71 (4) (1958) 630–672.

[11] H. L. A. Hart, Positivism and the separation of law and morals, Harvard Law Review 71 (4) (1958) 593–629.

[12] R. Riveret, Y. Gao, G. Governatori, A. Rotolo, J. Pitt, G. Sartor, A probabilistic argumentation framework for reinforcement learning agents - towards a mentalistic approach to agent profiles, Autonomous Agents and Multi-Agent Systems 33 (1-2) (2019) 216–274.

[13] R. Riveret, A. Rotolo, G. Sartor, H. Prakken, B. Roth, Success chances in argument games: A probabilistic approach to legal disputes, in: Proc. of the 20th Conference on Legal Knowledge and Information Systems, IOS Press, 2007, pp. 99–108.

[14] R. Riveret, H. Prakken, A. Rotolo, G. Sartor, Heuristics in argumentation: A game theory investigation, in: Proc. of the 2nd Conference on Computational Models of Argument, IOS Press, 2008, pp. 324–335.

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

[16] M. Caminada, L. Amgoud, On the evaluation of argumentation formalisms, Artificial Intelligence 171 (5-6) (2007) 286–310.

[17] Z. Li, S. Parsons, On argumentation with purely defeasible rules, in: Proc. of the 9th International Conference Scalable Uncertainty Management, Vol. 9310, Springer, 2015, pp. 330–343.

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

[19] G. A. Vreeswijk, Abstract argumentation systems, Artificial Intelligence 90 (1) (1997) 225 – 279.

[20] H. Prakken, On support relations in abstract argumentation as abstractions of inferential relations, in: Proc. of the 21st European Conference on Artificial Intelligence, IOS Press, 2014, pp. 735–740.

[21] C. Cayrol, M.-C. Lagasquie-Schiex, Bipolarity in argumentation graphs: Towards a better understanding, International Journal of Approximate Reasoning 54 (7) (2013) 876 – 899.

[22] A. Cohen, S. Gottifredi, A. J. Garca, G. R. Simari, A survey of different approaches to support in argumentation systems, The Knowledge Engineering Review 29 (5) (2014) 513550.

[23] S. Polberg, N. Oren, Revisting support in abstract argumentation systems, in: Proc. of the 5th International Conference on Computational Models of Argument, 2014, pp. 369–376.

[24] A. Cohen, S. Parsons, E. I. Sklar, P. McBurney, A characterization of types of support between structured arguments and their relationship with support in abstract argumentation, International Journal of Approximate Reasoning 94 (2018) 76 – 104.

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

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

[27] S. Modgil, M. Caminada, Proof theories and algorithms for abstract argumentation frameworks, in: Argumentation in Artificial Intelligence, Springer, 2009, pp. 105–129.

[28] P. Baroni, G. Governatori, R. Riveret, On labelling statements in multi-labelling argumentation, in: Proc. of the 22nd European Conference on Artificial Intelligence, IOS Press, 2016, pp. 489–497.

[29] P. Baroni, R. Riveret, Enhancing statement evaluation in argumentation via multi-labelling systems, Journal of Artificial Intelligence Research 66 (2019) 793–860.

[30] R. Riveret, G. Governatori, On learning attacks in probabilistic abstract argumentation, in: Proc. of the 15th International Conference on Autonomous Agents & Multiagent Systems, ACM, 2016, pp. 653–661.

[31] R. Riveret, On searching explanatory argumentation graphs, Journal of Applied Non-Classical Logics 30 (2) (2020) 123–192.

[32] R. RIveret, A. Artikis, J. Pitt, E. G. Nepomuceno, Self-governance by transfiguration: From learning to prescription changes, in: Proc. of the 8th International Conference on Self-Adaptive and Self-Organizing Systems, 2014, pp. 70–79.

[33] C. E. Alchourrón, E. Bulygin, Normative systems, Springer-Verlag, 1971.

[34] G. H. von Wright, Norm and Action: A Logical Enquiry, Routledge and Kegan Paul, 1963.

[35] D. Gabbay, J. Horty, X. Parent, R. van der Meyden, L. van der Torre (Eds.), Handbook of Deontic Logic and Normative Systems, College Publications, 2013.

[36] A. Ross, Directives and Norms, Humanities Press, 1967.

[37] R. Mullins, Legal positivism and deontic detachment, Ratio Juris 31 (1) (2018) 4–8.

[38] J. Carmo, A. J. I. Jones, Deontic Logic and Contrary-to-Duties, Springer, 2002, pp. 265–343.

[39] P. McNamara, Deontic logic, in: D. Gabbay, J. Woods (Eds.), The Handbook of the History of Logic, vol. 7: Logic and the Modalities in the Twentieth Century, Elsevier Press, 2006, pp. 197–288.