From Explicit Allowances to Defeasible Deontic Operators: A Modal View

Agata Ciabattoni[1] [0000−0001−6947−8772], Josephine Dik[1(✉)] [0009−0003−6149−5684], Emiliano Lorini[2] [0000−0002−7014−6756], Dominik Pichler[1] [0009−0003−1790−2983], and Dmitry Rozplokhas[1] [0000−0001−7882−4497]

[1] TU Wien, Austria
{agata,josephine,dominik,dmitry}@logic.at
[2] IRIT, CNRS, Toulouse University, France
emiliano.lorini@irit.fr

Abstract. Preference-based deontic logics provide a foundation for normative reasoning but fail to distinguish between explicit allowances – specified by a designer – and implicit ones derived by inference. This distinction is crucial in systems where agents may act only if (explicitly or implicitly) permitted. In this paper, we formalize this inference by grounding the preference ordering over possible worlds in a permission base, i.e., a set of explicit allowances, and derive implicit permissions, as well as defeasible prohibitions and obligations. Our framework provides solutions to key deontic paradoxes and is a conservative extension of Åqvist's dyadic deontic system **F** extended with cautious monotony. We illustrate the approach with a case study involving robotic agents operating under normative constraints and provide complexity results together with a QBF-based decision procedure to support automated reasoning.

Keywords: Permission · Preference-based semantics · Åqvist systems · Deontic logic.

1 Introduction

In designing effective and reliable AI agents, it is essential to ensure that they act only when permitted to do so. Permissions, and deontic concepts in general, are inherently conditional. Their formal analysis relies on dyadic deontic systems (see e.g. [11]), among which the family with preference-based semantics is the most well-known. This approach was originally developed by [6,13], and later adapted to a modal logic setting by [2] and [23]. Prominent preference-based deontic logics include Åqvist's systems **E**, **F** and **G** [2], and the systems in [21] and [7]. They offer an adequate treatment of one of the core challenges in normative reasoning, i.e., the handling of contrary-to-duty (CTD) norms, which are norms (obligations or prohibitions) that arise when other norms are violated. However, they do not distinguish between explicit permissions, which are allowances directly specified by a system designer, and implicit permissions, which arise through logical inference. This distinction is crucial in contexts where the system designer specifies an agent's behavior by providing a finite set of explicit permissions, with the expectation that these will fully determine the agent's actions, both directly and through any permissions that can be logically inferred from them.

In this paper, we propose a formal mechanism which answers the question: Given (i) a set of explicitly defined allowances in the form of unary permissions, and (ii) contingent information about the domain; assuming that the agent may act only if explicitly or implicitly permitted, what actions is the agent allowed to take in a specific situation? Our goal is related to that of [7], which focuses on determining ideal outcomes from a set of deontic norms. Here, in addition, we aim to answer the above question by developing a preference-based deontic logic grounded in permission bases.

Our framework builds on a computationally grounded semantics for modal logic developed in a series of works by Lorini et al., originally focused on epistemic reasoning [26,27,29,30], and later extended to model mental attitudes [25], and causal reasoning [24,28]. In this semantics, the states (or worlds) in a model are not treated as primitive entities, as in standard modal logic semantics, but are instead decomposed into two components: a knowledge base and a propositional valuation. Moreover, the accessibility relations between possible states are not given as primitives but are computed from the knowledge bases. The idea of distinguishing explicit information from implicit one is also found in prior work in linguistics [20] and knowledge representation [19].

The knowledge bases of our framework consist of explicit allowances (permission bases), are used to determine the preference ordering over possible states. We consider two variants: one in which the permission base remains fixed across all states –corresponding to the addition of the Absoluteness property within the semantics– and one in which it does not. These are conservative extensions of $\mathbb{PCLTU}$ and $\mathbb{PCLTA}$, variants of Burgess' logic $\mathbb{PCL}$ [5], respectively (see Section 2.2 for details); $\mathbb{PCLTA}$ coincides with the deontic system **F** introduced by Åqvist, augmented with cautious monotony (CM), an important property of non-monotonic systems introduced in [10]. In this paper, we focus on the case with Absoluteness, and together with the standard notions of obligation and prohibition defined in terms of permissions, we introduce a defeasible variant relative to the permission base. Our approach also provides a unified formalization of the three types of permissions from [14]: explicit, implicit, and tacit permissions, see also [4]. Explicit permissions are granted in a top-down manner, implicit permissions can be logically derived from the explicit ones, and tacit permissions correspond to the absence of (defeasible) prohibitions.

To analyze the behavior of our framework, we examine its response to prominent deontic paradoxes concerning permission (Free Choice Permission [18] and Ross's Paradox [37]) and demonstrate that our defeasible operators handle CTD scenarios as expected, while avoiding the problem of preference-based systems identified by [17] with the 'asparagus paradox'. We illustrate our approach with a case study involving robotic agents operating under normative constraints and provide a PSPACE complexity result together with a QBF-based decision procedure to support automated reasoning.

2 Formal Framework

We present a novel preference-based semantics for deontic reasoning. This semantics incorporates the notion of explicit allowance and uses it to compute the preference ordering over possible states within a model. We will leverage it to interpret a language that combines the notions of explicit allowance and implicit conditional permission.

2.1 Semantics and language

Assume to have a countably infinite set of atomic propositions $Atm = \{p, q, \ldots\}$. We define the language $\mathcal{L}_0$ for explicit allowances by the following grammar:

$$\mathcal{L}_0 \stackrel{\text{def}}{=} \alpha ::= p \in Atm \mid \neg\alpha \mid \alpha \wedge \alpha \mid \triangle\alpha,$$

The connectives $\top$, $\bot$, $\vee$ and $\rightarrow$ are classically defined as usual. The operator $\triangle$ is used to represent explicit allowances: the formula $\triangle\alpha$ is read as "α is explicitly permitted". $\mathcal{L}_0$ is the first layer of the language.

Unlike standard semantics of modal and deontic logic where a state is a primitive, in our semantics a state has two components: a base of explicit allowances (permission base) and a propositional valuation representing the atoms that are true at the state.

Definition 1 (State). *A state is a pair $S = (B, V)$ with $B \subseteq \mathcal{L}_0$ a finite set of explicit allowances (or permission base) and $V \subseteq Atm$ a propositional valuation. The set of states is denoted by* $\mathbf{S}$.

Formulas of the language $\mathcal{L}_0$ are interpreted relative to a state, as follows (Boolean cases are omitted, as they are defined in the usual way).

Definition 2 (Satisfaction relation). *Let $S = (B, V) \in \mathbf{S}$:*

$$S \models p \iff p \in V,$$
$$S \models \triangle\alpha \iff \alpha \in B.$$

Note in particular the interpretation of the explicit allowance modality $\triangle$: α is explicitly permitted if α is included in the permission base. Given $S, S' \in \mathbf{S}$ let

$$Sat(S', S) = \{\alpha \in B \mid S' \models \alpha\}$$

be the set of explicit allowances from state S that are satisfied at state S'. The following definition introduces the preference ordering over states. We compute it from the explicit allowances in a permission base.

Definition 3 (Preference ordering). *Let $S, S', S'' \in \mathbf{S}$, $S = (B, V), S' = (B', V')$, $S'' = (B'', V'')$. Then we define $S'' \preceq_S S'$ if and only if $Sat(S'', S) \subseteq Sat(S', S)$. Furthermore, we write $S'' \prec_S S'$ if and only if $S'' \preceq_S S'$ and $S' \not\preceq_S S''$.*

$S'' \preceq_S S'$ means that relative to the state S, state S' is at least as good (or ideal) as state S''. According to Definition 3, the latter holds if the set of explicit allowances from the permission base of S that are satisfied at S'' is included in the set of explicit allowances from the permission base of S that are satisfied at S'.

For states $S = (B, V)$ and $S' = (B', V')$, we srite $S \equiv S'$ if they share the same permission base, i.e., $S \equiv S'$ iff $B = B'$. Note that $S \equiv S'$ leads to $\preceq_S = \preceq_{S'}$.

A model is a state together with a set of states containing it, called the *context*. The context includes all states compatible with the current hard information, where hard information –facts treated as fixed and commonly known. Formally:

Definition 4 (Model). *A model is a pair* (S, U) *with* $S \in U \subseteq \mathbf{S}$. *The class of models is denoted by* $\mathbf{M}$.

We analyze below the properties of the preference ordering for the models of Def. 4.

Lemma 1. *Let* $(S, U) \in \mathbf{M}$. *Then, i) the ideality ordering* $\preceq_S$ *is a preorder, and ii) every nonempty* $X \subseteq U$ *contains a* $\preceq_S$*-maximal element.*

Proof. Item i) follows directly from Def. 3 and the fact that the subset relation $\subseteq$ is reflexive and transitive. For item ii), let $S = (B, V)$. The only way X could not have a maximal element would be that there exists an infinite increasing chain of states inside of X. By assumption, B is finite. Hence, such a chain cannot exist. □

We now consider the subclass of normatively absolute models, where the permission base is constant across all states, a standard assumption in preference-based deontic logics such as Åqvist [2] and Kratzer [21].

Definition 5 (Normatively absolute model). *A model* (S, U) *is normatively absolute if* $\forall S', S'' \in U, S' \equiv S''$. *The class of normatively absolute models is denoted* $\mathbf{M}^{abs}$.

Normatively absolute models satisfy the following absoluteness property:

$$\text{if } (S, U) \in \mathbf{M}^{abs} \text{ then } \forall S', S'' \in U, \preceq_{S'} = \preceq_{S''} . \tag{1}$$

We extend the language $\mathcal{L}_0$ with a dyadic modal operator for implicit conditional permission. The new language, denoted by $\mathcal{L}$, is defined by the following grammar:

$$\mathcal{L} \stackrel{\text{def}}{=} \varphi ::= \alpha \mid \neg\varphi \mid \varphi \wedge \varphi \mid \varphi \rhd \varphi,$$

where α ranges over $\mathcal{L}_0$. Again, the connectives $\top$, $\bot$, $\vee$, and $\rightarrow$ are classically defined as usual. The formula $\psi \rhd \varphi$ reads as "φ is implicitly permitted conditional to ψ".

Formulas of the language $\mathcal{L}$ are interpreted relative to a model as follows. (Boolean cases are omitted as they are defined in the usual way.)

Definition 6 (Satisfaction relation (cont.)). *Let* $(S, U) \in \mathbf{M}$. *Then:*

$$(S, U) \models \alpha \iff S \models \alpha,$$
$$(S, U) \models \psi \rhd \varphi \iff \exists S' \in \mathit{Best}(\psi,S,U) \textit{ such that } (S', U) \models \varphi,$$

$$\mathit{Best}(\psi,S,U) = \big\{S' \in U \mid (S', U) \models \psi, \nexists S'' \in U \textit{ s.t. } (S'', U) \models \psi \textit{ and } S' \prec_S S''\big\}.$$

Hence φ is implicitly permitted given ψ if there is at least one ψ-most preferred state where φ holds.

As a consequence of Lemma 1, we can conclude that for any model (S, U) and formula ψ, if there exists a state satisfying ψ there exists a $\preceq_S$-maximal ψ state.

Corollary 1. *Given a model* $(S, U) \in \mathbf{M}$. *If the set* $\{S' \in U : (S', U) \models \psi\} \neq \emptyset$ *then* $\mathit{Best}(\psi,S,U) \neq \emptyset$.

Validity and satisfiability for $\mathbf{M}$ (resp. $\mathbf{M}^{abs}$) are defined in the expected way.

Definition 7 (Validity and satisfiability). *φ is valid for the class $\mathbf{M}$ (resp. $\mathbf{M}^{abs}$), denoted by $\models_{\mathbf{M}} \varphi$ (resp. $\models_{\mathbf{M}^{abs}} \varphi$), if $(S,U) \models \varphi$ for every $(S,U) \in \mathbf{M}$ (resp. $\in \mathbf{M}^{abs}$). φ is satisfiable for the class $\mathbf{M}$ (resp. $\mathbf{M}^{abs}$) if $\not\models_{\mathbf{M}} \neg\varphi$ (resp. $\not\models_{\mathbf{M}^{abs}} \neg\varphi$).*

Furthermore, the notion of logical consequence is defined as follows.

Definition 8 (Logical consequence). *Given a finite set of formulas Σ and a formula φ we say that φ is a logical consequence of Σ in $\mathbf{M}$ (resp. $\mathbf{M}^{abs}$), denoted $\Sigma \models_{\mathbf{M}} \varphi$ (resp. $\Sigma \models_{\mathbf{M}^{abs}} \varphi$), if for all $(S,U) \in \mathbf{M}$ (resp. for all $(S,U) \in \mathbf{M}^{abs}$):*

$$\text{if } \forall S' \in U, (S',U) \models \bigwedge_{\psi \in \Sigma} \psi \text{ then } \forall S' \in U, (S',U) \models \varphi.$$

As shown below, the universal modality can be defined in terms of $\rhd$. This will permit us to express the previous notion of logical consequence in the language $\mathcal{L}$.

2.2 Properties

In this section, we first analyze the key properties of the operator $\rhd$ in isolation and then examine its interaction with the operator $\triangle$ to highlight the relationship between explicit and implicit permission. We then show that our models generalize the preference-based models underlying the well-known conditional logics $\mathbb{PCLTU}$ and $\mathbb{PCLTA}$ (i.e. Åqvist's logic **F** + cautious monotony) for $\mathbf{M}$ and $\mathbf{M}^{abs}$, respectively.

The universal modality is definable. We begin by showing that the universal modality, along with its dual, the existential modality, can be defined through the following abbreviations using the dyadic modality $\rhd$: $\Diamond\varphi \stackrel{\text{def}}{=} \varphi \rhd \varphi$, and $\Box\varphi \stackrel{\text{def}}{=} \neg\Diamond\neg\varphi$.

Lemma 2. *Let $(S,U) \in \mathbf{M}$. Then, the following are equivalent: (i) there exists $S' \in U$ such that $(S',U) \models \varphi$, and (ii) $(S,U) \models \Diamond\varphi$.*

Proof. Assume (i). By Lemma 1, there is a state $S'' \in Best(\varphi, S, U)$, and by Definition 6, this state satisfies φ. Hence, by the semantics of the conditional modality, we have (ii). For the converse, assume (ii). By the satisfaction condition for $\rhd$, it follows that there exists a state in $Best(\varphi, S, U)$ that satisfies φ, which implies (i). □

Notice that, in the light of Lemma 2, it is easy to show that the modality $\Box$ is an S5 modality. Moreover, we have the following validities for the class $\mathbf{M}^{abs}$:

$$\models_{\mathbf{M}^{abs}} \triangle\alpha \to \Box\triangle\alpha, \tag{2}$$

$$\models_{\mathbf{M}^{abs}} \neg\triangle\alpha \to \Box\neg\triangle\alpha, \tag{3}$$

$$\models_{\mathbf{M}^{abs}} (\varphi \rhd \psi) \to \Box(\varphi \rhd \psi), \tag{4}$$

$$\models_{\mathbf{M}^{abs}} \neg(\varphi \rhd \psi) \to \Box\neg(\varphi \rhd \psi). \tag{5}$$

The following deduction theorem is a direct corollary of Lemma 2. Since the universal modality can be represented in the language $\mathcal{L}$, the notion of logical consequence of Definition 8 is also syntactically expressible.

Theorem 1. *Let Σ be a finite set of formulas and φ a formula. Then,*

$$\Sigma \models_{\mathbf{M}} \varphi \text{ iff } \models_{\mathbf{M}} \Box(\bigwedge_{\psi\in\Sigma} \psi) \to \varphi \quad \text{and} \quad \Sigma \models_{\mathbf{M}^{abs}} \varphi \text{ iff } \models_{\mathbf{M}^{abs}} \Box(\bigwedge_{\psi\in\Sigma} \psi) \to \varphi.$$

Interaction between explicit allowance and implicit conditional permission. As a next step, we analyze the $\triangle$ operator and its interaction with the conditional modality $\rhd$. We begin by noting that $\triangle$ is a syntactic operator, meaning it is fully intensional. That is, $\triangle\alpha$ and $\triangle\beta$ are not necessarily semantically equivalent, even if $\alpha \leftrightarrow \beta$ is a propositional tautology. This is due to the fact that the evaluation of an explicit allowance $\triangle\alpha$ depends solely on whether $\alpha \in B$, where B is the permission base associated with a state. Since B is predefined and syntactic in nature, the truth value of the formula α does not influence the truth value of $\triangle\alpha$. For instance, consider a model with state $S = (\{\alpha\}, \emptyset)$. Then it holds that $S \models \triangle\alpha \wedge \neg\triangle(\alpha \vee \alpha)$.

This intensionality is desirable for explicit allowances, as such permissions have no logical consequences beyond their syntax. Thus, stating α or β is not equivalent—even if α is semantically equivalent to β—since only α may appear in the permission base. In particular, by not closing the permission base under logical equivalence, we maintain its finiteness, aligning with the notion of issuing a finite set of explicit instructions.

Given a model (S, U) and states $S' \in U$, Definition 3 implies the following: if $\alpha \in B$ and $(S', U) \models \alpha$, then for all $S'' \in U$ such that $S' \preceq_S S''$, we have $(S'', U) \models \alpha$. This captures the idea that explicit allowance cannot make a state worse by being true—anything explicitly permitted preserves or improves the deontic status.

We now examine the interaction between explicit allowances and conditional implicit permissions. A key observation is that explicit permissions generate implicit ones. Specifically, if α is explicitly allowed in a state, then α is implicitly permitted under any condition φ such that $\varphi \wedge \alpha$ is possible. Formally: $(\triangle\alpha \wedge \Diamond(\varphi \wedge \alpha)) \to (\varphi \rhd \alpha)$. In particular, α is implicitly permitted unconditionally (i.e., under condition $\top$): $(\triangle\alpha \wedge \Diamond\alpha) \to (\top \rhd \alpha)$. Moreover, if an explicit conditional permission $\triangle(\beta \to \alpha)$ is given and $\beta \wedge \alpha$ is possible, then the corresponding implicit conditional permission also holds: $(\triangle(\beta \to \alpha) \wedge \Diamond(\beta \wedge \alpha)) \to (\beta \rhd \alpha)$. We prove these in the following Theorem 2.

Theorem 2. *We have the following validities:*

$$\models_{\mathbf{M}} (\triangle\alpha \wedge \Diamond(\varphi \wedge \alpha)) \to \varphi \rhd \alpha \tag{6}$$

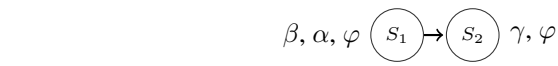
$$\models_{\mathbf{M}} (\triangle\alpha \wedge \Diamond\alpha) \to \top \rhd \alpha \tag{7}$$

$$\models_{\mathbf{M}} (\triangle(\beta \to \alpha) \wedge \Diamond(\beta \wedge \alpha)) \to \beta \rhd \alpha \tag{8}$$

Proof. For the first validity, assume $S = (B, V) \in U$ and $(S, U) \models \triangle\alpha \wedge \Diamond(\varphi \wedge \alpha)$. Then there exists a state $S' \in U$ such that $(S', U) \models \varphi \wedge \alpha$. Consider the following set $X = \{S'' \in U \colon (S'', U) \models \varphi \textit{ and } S' \preceq_S S''\}$. $S' \in X$, so by Lemma 1, there exists a $\preceq_S$-maximal state S^* in X. $S^* \in \mathit{Best}(\varphi, S, U)$ (otherwise S^* would not be maximal in X) and $(S^*, U) \models \alpha$ by Definition 3 (since $\alpha \in B$ and $(S', U) \models \alpha$ and $S' \preceq_S S^*$). So $(S, U) \models \varphi \rhd \alpha$ by Definition 6. The second validity is the instance of the first one with $\varphi = \top$. The proof of the third validity is analogous: for state $S' \in U$ such that $(S', U) \models \beta \wedge \alpha$ we consider set $X = \{S'' \in U : (S'', U) \models \beta \textit{ and } S' \preceq_S S''\}$ and $\preceq_S$-maximal state S^* in it. Since $(S', U) \models \beta \to \alpha$, also $(S^*, U) \models \beta \to \alpha$ (since $(\beta \to \alpha) \in B$ and $S' \preceq_S S^*$), therefore $(S^*, U) \models \alpha$, and thus $(S, U) \models \beta \rhd \alpha$. □

Explicit conditional permissions like $\triangle(\beta \to \alpha)$ do not derive implicit permissions under arbitrary additional assumptions; i.e., in general the following formula is not valid: $(\triangle(\beta \to \alpha) \wedge \Diamond(\beta \wedge \alpha \wedge \varphi)) \to \varphi \rhd \alpha$.

Example 1. Consider the model $(S_1, \{S_1, S_2\})$ where both states share the same permission base $B = \{\beta \to \alpha, \gamma\}$, we draw an arrow from S_1 to S_2 iff $S_1 \prec_{S_1} S_2$:

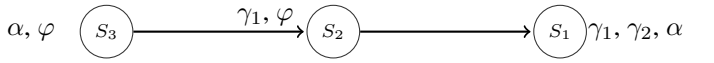


Here, S_1 satisfies $\triangle(\beta \to \alpha) \wedge \Diamond(\beta \wedge \alpha \wedge \varphi)$. Yet S_2, which satisfies all elements in the permission base, is the only element in $Best(\varphi, S_1, U)$. Since S_2 does not satisfy α, we have $(S_1, U) \not\models \varphi \rhd \alpha$.

We have seen that explicit permissions generate implicit ones, but they are not so strong as to entail prohibitions (negated permissions $\neg(\top \rhd \alpha)$) or obligations (duals of permissions $\neg(\top \rhd \neg\alpha)$). This is because an explicit permission of α merely ensures that α holds in some best state if it is possible—i.e., $\Diamond\alpha$—but does not require that all best states satisfy α. Therefore, the following is not valid: $(\triangle\alpha \wedge \Diamond\alpha) \to \neg(\top \rhd \neg\alpha)$

Example 2. Take the model $(S_1, \{S_1, S_2\})$ such that $S = (B, V')$ and $S_2 = (B, V'')$ with $B = \{p, q\}$, $V' = \{p\}$ and $V'' = \{q\}$. In this model, S_1 satisfies $\triangle p \wedge \Diamond p$. The state S_2 is in $Best(\top, S_1, U)$ and satisfies $\neg p$. Therefore, $(S_1, U) \models \top \rhd \neg p$ still holds. This shows that the explicit permission for p does not yield a prohibition for $\neg p$, nor does it entail the obligation of p, i.e., $\neg(\top \rhd \neg p)$.

Note that explicit permissions are stronger than implicit ones, since an unconditional implicit permission does not imply a conditional implicit permission. More specifically, $((\top \rhd \alpha) \wedge \Diamond(\varphi \wedge \alpha)) \to (\varphi \rhd \alpha)$ is not valid as shown by the following model $(S_1, \{S_1, S_2, S_3\})$ where all states share the same permission base $B = \{\gamma_1, \gamma_2\}$:



Here, S_1 satisfies $\top \rhd \alpha$ and $\Diamond(\varphi \wedge \alpha)$. Yet S_2 is the only element in $Best(\varphi, S_1, U)$. As S_2 does not satisfy α, we have $(S_1, U) \not\models \varphi \rhd \alpha$. This example shows that the permission base B grounds the ideality ordering in such a way that explicitly permitted formulas are still satisfied when moving up the order. In the case α is added to B this model is no longer an element of the class $\mathbf{M}$ since the state S' invalidated Def. 3.

Connection with conditional logics. Implicit permissions alone behave like (dual) conditionals in standard conditional logics. We show that formulas in the following triangle-free fragment of our language:

$$\mathcal{L}_{\rhd} \stackrel{\text{def}}{=} \pi ::= p \mid \neg\pi \mid \pi \wedge \pi \mid \pi \rhd \pi,$$

valid w.r.t. $\mathbf{M}$ and $\mathbf{M}^{abs}$ correspond to the theorems of the conditional logics $\mathbb{PCLTU}$ and $\mathbb{PCLTA}$, respectively. Both logics belong to a foundational family [9] of extensions to preferential conditional logic $\mathbb{PCL}$ [5]. Specifically, $\mathbb{PCLTU}$ is a variant of $\mathbb{PCL}$ with models satisfying Total Reflexivity and Uniformity [9], while $\mathbb{PCLTA}$ adds Absoluteness, where all worlds share the same ordering. The latter is well-known in the deontic logic literature as a member of Åqvist's family (i.e. **F** with cautious monotonicity [35]). These logics are based on the following notion of preference models.

Definition 9 (Preference model). *A* preference model *is a tuple $M = \langle W, \preceq, \mathcal{V}\rangle$, where W is a set of worlds, $\mathcal{V} : W \to 2^{Atm}$ a valuation on W, and $\preceq$ is a ternary (world-indexed) preference relation: $\preceq_w$ is a preorder on W for each world w. The satisfaction relation is defined as follows. (Again, boolean cases are omitted as they are defined in the usual way.)*

$$(M, w) \models p \Leftrightarrow p \in \mathcal{V}(w),$$
$$(M, w) \models \pi_1 \rhd \pi_2 \Leftrightarrow \exists u \in Best(\pi_1, w, M), (M, u) \models \pi_2.$$

where $Best(\pi, w, M) = \{v \in W \mid (M, v) \models \pi$ and $\nexists v' \in W : (M, v') \models \pi$ and $v \preceq_w v'$ and $v' \not\preceq_w v\}$.

The definition of preference models in [9] is more general than this one. We use here a simpler version with preference relation being defined over the whole set of worlds — the consequence of Total Reflexivity and Uniformity. Another property standardly assumed for preference models is the limit assumption [23] (also known as stoppering [33] or smoothness [22]): For every $w, u \in W$ if $(M, u) \models \pi$ then either $u \in Best(\pi, w, M)$ or there exists $v \in Best(\pi, w, M)$ such that $u \preceq_w v$ and $v \not\preceq_w u$. This condition is ubiquitous in studies of conditional logics, yet it is non-trivial and depends not only on the model's structure (i.e., the frame) but also on the evaluation of formulas. An alternative to this assumption, used in [9], employs a significantly more complicated truth condition for evaluation of conditionals. These alternatives are known to give rise to the same valid formulas [23]. Here we adhere to the simpler version of the truth condition, since for models in $\mathbf{M}$, smoothness arises naturally as a corollary of the (much simpler) requirement of finiteness of permission bases. Logic $\mathbb{PCLTU}$ is defined by preference models satisfying smoothness, while $\mathbb{PCLTA}$ additionally requires Absoluteness, i.e. preference models where $\preceq_{w_1} = \preceq_{w_2}$ for all $w_1, w_2 \in W$. We say that a formula $\pi \in \mathcal{L}_{\rhd}$ is valid in $\mathbb{PCLTU}$ ($\models_{\mathbb{PCLTU}} \pi$) if it is satisfied in all worlds of all smooth preference models, and is valid in $\mathbb{PCLTA}$ ($\models_{\mathbb{PCLTA}} \pi$) if it is satisfied in all worlds of all smooth preference models satisfying Absoluteness.

Models in $\mathbf{M}$ are instances of preference models with preference relations given by $\{\preceq_S\}_{S \in U}$. Conversely, we show that an arbitrary preference relation over a finite model can be grounded by some selected finite permission base. Since, as shown in [9], $\mathbb{PCLTU}$ and $\mathbb{PCLTA}$ satisfy the finite model property, this implies that our framework is a conservative extension of the conditional logics $\mathbb{PCLTU}$ and $\mathbb{PCLTA}$.

Theorem 3 (Conservativity). *Let $\pi \in \mathcal{L}_{\rhd}$. Then, i) $\models_{\mathbf{M}} \pi$ iff $\models_{\mathbb{PCLTU}} \pi$; and ii) $\models_{\mathbf{M}^{abs}} \pi$ iff $\models_{\mathbb{PCLTA}} \pi$.*

Proof. The directions from right to left are straightforward, since $(S, U) \in \mathbf{M}$ corresponds to a preference model satisfying the limit assumption (with the set of worlds U and the preference relation by Definition 3), and $(S, U) \in \mathbf{M}^{abs}$ also satisfies Absoluteness. For the opposite directions, we use finite model property for $\mathbb{PCLTU}$ and $\mathbb{PCLTA}$ and show that any preference model $M = \langle W, \{\preceq_w\}_{w \in W}, \mathcal{V}\rangle$ with finite W such that $(M, w_0) \not\models \pi$ for some $w_0 \in W$ can be transformed into a grounded model $(S, U) \in \mathbf{M}$ such that $(S, U) \not\models \pi$, and this transformation preserves Absoluteness. Specifically, let

$Atm(\pi)$ be a set of all atoms appearing in π. Since Atm is infinite, we can take an injective mapping $\chi : W \times W \to (Atm \setminus Atm(\pi))$ that selects some fresh atom for each pair of worlds. Consider the following transformation of a world in W into a grounded state: $\mathcal{S}(w) = (B_w, \mathcal{V}(w) \cup \bigcup_{v \in W} Pr_v(w))$ where $B_v = \{\chi(v,u) \mid u \in W\}$ is a set of fresh explicit allowances in v corresponding to all worlds and $Pr_v(w) = \{\chi(v,u) \mid u \preceq_v w\}$ is the set of such allowances for predecessors of w w.r.t. $\preceq_v$. Then the ideality ordering generated by the mapped state $\mathcal{S}(v)$ (Def. 3) coincides with $\preceq_v$: if $u \preceq_v w$ then $Pr_v(u) \subseteq Pr_v(w) \subseteq B_v$ by transitivity of $\preceq_v$ and so $\mathcal{S}(u) \preceq_{\mathcal{S}(v)} \mathcal{S}(w)$, conversely if $u \not\preceq_v w$ then $\chi(v,u) \notin Pr_v(w)$ while $\chi(v,w) \in Pr_v(w)$ by reflexivity of $\preceq_v$ so $\mathcal{S}(u) \not\preceq_{\mathcal{S}(v)} \mathcal{S}(w)$. This implies $(\mathcal{S}(z), \{\mathcal{S}(w)\}_{w \in W}) \models \pi$ iff $M, z \models \pi$ for $z \in W$ (since $\mathcal{S}(\cdot)$ preserves all orderings and the valuation on $Atm(\pi)$, which fully determine the evaluation of π), therefore $(\mathcal{S}(w_0), \{\mathcal{S}(w)\}_{w \in W}) \models \pi$. Note that this transformation preserves the absoluteness, so the case of $\mathbf{M}^{abs}$ is covered as well. □

3 Case Study

Consider a mobile robot that moves in a space while complying with the rules explicitly specified by the system designer. These include what the robot is allowed to do when they encounter an intersection: a) the robot is allowed to cross an intersection when the traffic light is green, b) the robot is allowed to cross an intersection when the traffic light is green or orange, and no other vehicle is approaching the intersection from the right. These two explicit allowances are expressed as: $\mathsf{all}_1 \stackrel{\text{def}}{=} \triangle(gr \to cr)$, and $\mathsf{all}_2 \stackrel{\text{def}}{=} \triangle\Big(\big((gr \vee or) \wedge \neg ri\big) \to cr\Big)$. We assume that a traffic light at an intersection is either red, green, or flashing orange, and cannot have different colors at the same time. This assumption is captured by the following abbreviation: $\alpha_1 \stackrel{\text{def}}{=} (re \wedge \neg gr \wedge \neg or) \vee (\neg re \wedge gr \wedge \neg or) \vee (\neg re \wedge \neg gr \wedge or)$. Additionally, we assume that it is possible for the robot to cross when no vehicle is approaching from the right and the traffic light is not orange and not red. This assumption is captured by the following abbreviation: $\alpha_2 \stackrel{\text{def}}{=} \Diamond(cr \wedge \neg re \wedge \neg or \wedge \neg ri)$.

In the following, we show what the robot is permitted to do in a given situation. Specifically, we consider, without loss of generality, the permissions of the robot when they are standing in front of a traffic light.

When the robot is at a red traffic light, it checks whether it has permission to cross. It is routine to verify that:

$$\{\alpha_1, \alpha_2\} \not\models_{\mathbf{M}} (\mathsf{all}_1 \wedge \mathsf{all}_2) \to (re \rhd cr). \tag{9}$$

Thus, the robot does not have the permission to cross when the traffic light is red.

When the traffic light turns green, the robot has permission to cross. Indeed, thanks to the validity (8) in Theorem 2 and Theorem 1, we have:

$$\{\alpha_1, \alpha_2\} \models_{\mathbf{M}} (\mathsf{all}_1 \wedge \mathsf{all}_2) \to (gr \rhd cr). \tag{10}$$

When the traffic light is flashing orange, the robot has no permission to cross. Indeed, we have:

$$\{\alpha_1, \alpha_2\} \not\models_{\mathbf{M}} (\mathsf{all}_1 \wedge \mathsf{all}_2) \to (or \rhd cr). \tag{11}$$

However, the robot has permission to cross when the traffic light is not red and no vehicle is coming from the right:

$$\{\alpha_1, \alpha_2\} \models_{\mathbf{M}} (\mathsf{all}_1 \wedge \mathsf{all}_2) \rightarrow \big((\neg re \wedge \neg ri) \rhd cr\big). \tag{12}$$

Note that the properties (9), (10), (11), and (12) are expressed in terms of logical consequence. By Theorem 1, these can be reduced to validity checking problems. In Section 5, we introduce a QBF-based validity checking procedure that enables the automatic verification of such properties.

4 Defeasible Deontic Operators

Using our semantics, we formalize defeasible obligations, prohibitions, and tacit permissions. We then examine their behavior, along with that of the implicit permission operator, in relation to key deontic paradoxes, understood here as (un)derivable theorems that challenge intuition.

4.1 Obligations and Prohibitions

A defeasible prohibition refers to actions that are not permitted, while their negations are. In contrast, a defeasible obligation concerns actions that are permitted, while their negations are not. Formally, these are defined as:

Definition 10. *Let Σ be a set of formulas:*

- *φ is* defeasibly prohibited *when ψ with respect to Σ iff $\Sigma \not\models_{\mathbf{M}} \psi \rhd \varphi$ and $\Sigma \models_{\mathbf{M}} \psi \rhd \neg\varphi$. We denote this as $\Sigma \mid\sim F^*(\varphi/\psi)$.*
- *φ is* defeasibly obligatory *when ψ with respect to Σ iff $\Sigma \mid\sim F^*(\neg\varphi/\psi)$. We write this as $\Sigma \mid\sim O^*(\varphi/\psi)$.*

We call these operators defeasible[3] because extending the set of assumptions might invalidate previously derived norms. The non-monotonicity of the consequence relation $\mid\sim$ is illustrated by the following example.

Example 3. Take $B = \{p\}$ and $\Gamma = \{\Diamond p\}$. Hence $\triangle B \cup \Gamma \mid\sim O^*(p/\top)$. However, when $B' = B \cup \{\neg p\}$ and $\Gamma' = \Gamma \cup \{\Diamond \neg p\}$, we get that $\triangle B' \cup \Gamma' \models_{\mathbf{M}} \top \rhd p$ and $\triangle B' \cup \Gamma' \models_{\mathbf{M}} \top \rhd \neg p$, and thus $\triangle B' \cup \Gamma' \not\mid\sim O^*(p/\top)$.

Note that omitting the condition $\Sigma \models_{\mathbf{M}} \psi \rhd \neg\varphi$ from the definition of defeasible prohibition would result in undesirable behavior. For instance, when $B = \{\alpha\}$, and set of assumptions $\Gamma = \{\Diamond\alpha, \Diamond\beta\}$ we would obtain that β and $\neg\beta$ are prohibited.

Definition 11. *φ is* tacitly permitted *when ψ with respect to Σ iff $\Sigma \not\mid\sim F^*(\varphi/\psi)$.*

However, since this does not clarify whether the action is implicitly permitted, an agent should not act based solely on this presumed permission.

[3] The term is used in legal reasoning (e.g. [32]) and in Deontic Logic (e.g. [12,34]), with a different meaning, accounting for norms involving possible (prima facie) conflicts and exceptions. Also, there is no obvious connection between our logic and Defeasible Deontic Logic (DDL) [1], as they originate from fundamentally different traditions; ours is preference-based, extending Åqvist's system **F**+(CM), while DDL is rule-based.

4.2 Paradoxes

We evaluate our framework on paradoxes that challenged Åqvist systems.

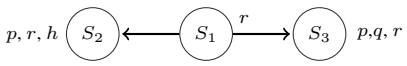
Free-Choice Paradox. It results from the undesired formulas arising when accepting as premise the (formalization of the) sentence [18]: "It is permitted to have tea or coffee implies permitted to have tea and permitted to have coffee". In a dyadic deontic setting, it is formalized as $\top \rhd (tea \vee coffee) \rightarrow (\top \rhd tea) \wedge (\top \rhd coffee)$. In Åqvist's systems, assuming this formula, one can derive the following: I) $\theta \rhd \varphi \rightarrow \theta \rhd \psi$, II) $\theta \rhd \varphi \rightarrow \theta \rhd (\varphi \wedge \psi)$, and III) $\neg(\theta \rhd \neg\varphi) \rightarrow \neg(\theta \rhd \neg(\varphi \wedge \psi))$, for any φ, ψ, θ.

We formalize the free-choice inference in two ways, with $\alpha_1, \alpha_2, \alpha_3$ in $\mathcal{L}_0$:

1. $\triangle(\alpha_3 \rightarrow (\alpha_1 \vee \alpha_2)) \rightarrow \triangle(\alpha_3 \rightarrow \alpha_1) \wedge \triangle(\alpha_3 \rightarrow \alpha_2)$
2. $\triangle(\alpha_3 \rightarrow (\alpha_1 \vee \alpha_2)) \rightarrow \alpha_3 \rhd \alpha_1 \wedge \alpha_3 \rhd \alpha_2$

Ex. 4 shows that the undesired results I)–III) do not follow from assuming 1 and 2.

Example 4. Let $B = \{r \rightarrow (p \vee q), r \rightarrow p, r \rightarrow q, h\}$, where p, q, r, h are atomic formulas. The following model satisfies 1–2 but not I)–III):



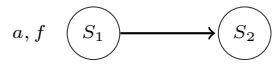
It is easy to see that both formulations 1 and 2 of the free-choice inference are true in the model, while none if the undesired formulas I)-III) is. For I), we see that $(S_i, U) \models r \rhd p$, but $(S_i, U) \not\models \neg(r \rhd (q \wedge \neg p))$, and for II) we have $(S_i, U) \models r \rhd q$, but $(S_i, U) \not\models \neg(r \rhd (q \wedge h))$, for $i \in \{1, 2, 3\}$. For III), we see that $(S_i, U) \models \neg(r \rhd \neg p)$ but $(S_i, U) \not\models \neg(r \rhd \neg(p \wedge q))$, for $i \in \{1, 2, 3\}$.

A different way of avoiding the formula III) involves the use of the defeasible obligation; we indeed show that if there is a set of formulas Σ such that $\Sigma \mathrel{|\!\sim} O^*(p/r)$, this does not imply $\Sigma \mathrel{|\!\sim} O^*((p \wedge q)/r)$. Consider the set of assumptions $\triangle B \cup \Gamma$, where $\Gamma = \{\Diamond(p \wedge r), \Diamond(q \wedge r)\}$. Then, we have that $\triangle B \cup \Gamma \models_{\mathbf{M}} r \rhd p$, and our model shows that $\triangle B \cup \Gamma \not\models_{\mathbf{M}} r \rhd p$, and thus $\triangle B \cup \Gamma \mathrel{|\!\sim} O^*(p/r)$; and since $\triangle B \cup \Gamma \not\models_{\mathbf{M}} r \rhd (p \wedge q)$, we have $\triangle B \cup \Gamma \not\mathrel{|\!\sim} O^*((p \wedge q)/r)$.

Ross' paradox. In SDL [38] or Åqvist systems, from the sentence: 1. "You are permitted to mail the letter", follows the unintuitive sentence 2. "You are permitted to mail the letter or burn it" [37]. While this is still the case for the implicit permission in our framework – since $(S, U) \vDash \theta \rhd \varphi$ implies that $(S, U) \vDash \theta \rhd (\varphi \vee \psi)$ – this is not for explicit permissions. Namely, $(S, U) \vDash \triangle\alpha$ does not imply $(S, U) \nvDash \triangle(\alpha \vee \beta)$, nor does it imply $(S, U) \vDash \top \rhd (\alpha \vee \beta)$. To see why, take a model with permission base $B = \{p\}$, and one state $S_1 = (B, V)$ with $V = \emptyset$. In that case, we have $(S_1, U) \models \triangle p$, but $(S_1, U) \not\models \top \rhd p$ and $(S_1, U) \not\models \top \rhd p \vee q$, and $(S_1, U) \not\models \triangle(p \vee q)$.

Asparagus Paradox. It consists of: 1. You should not eat with your fingers; 2. When eating asparagus, you should eat with your fingers; 3. If you eat with your fingers, you should wash them. As pointed out in [17], the formalization of 1. and 2. in Kratzer's semantics [21] (and in Åqvist systems [2]) leads to the counterintuitive prohibition to eat asparagus, $\neg(\top \rhd a)$, while in other frameworks (e.g. [7]) the original prohibition of not eating with fingers is somehow canceled (this is called the drowning problem

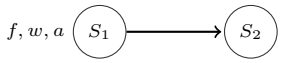
in [4]). We use the defeasible deontic operators of Def. 10 to avoid both undesired consequences. We assume the permission base: $B = \{\neg f,\, a \to f\}$ and the set of formulas $\Gamma = \{\Diamond\neg f, \Diamond(f \wedge a)\}$. The figure below shows that $\triangle B \cup \Gamma \not\models_{\mathbf{M}} a \rhd \neg f$ and $\triangle B \cup \Gamma \not\models_{\mathbf{M}} \top \rhd f$.



From Theorem 2.7 follows $\triangle B \cup \Gamma \models_{\mathbf{M}} \top \rhd \neg f$ and we conclude $\triangle B \cup \Gamma \;\not\sim\; F^*(f/\top)$. Then from Theorem 2.8, it follows that $\triangle B \cup \Gamma \models_{\mathbf{M}} a \rhd f$ and thus $\triangle B \cup \Gamma \;\not\sim\; O^*(f/a)$. Then, $\triangle B \cup \Gamma \models_{\mathbf{M}} \top \rhd a$, and thus we have $\triangle B \cup \Gamma \;\not\sim\; F^*(a/\top)$. Therefore, we do not obtain the counterintuitive prohibition to eat asparagus.

Statements 1. and 3. are contrary-to-duties CTD (the most famous paradox involving CTDs being the Gentle Murder [8] paradox). In contrast with Standard Deontic Logic SDL [38], preference-based logics can correctly handle CTDs. The same applies to our defeasible operators, which enable to derive both $\Sigma \;\not\sim\; O^*(w/f)$ and $\Sigma \;\not\sim\; F^*(\neg f/\top)$, from a set of assumptions Σ.

To show this, let us define $B' = B \cup \{f \to w\}$ and $\Gamma' = \Gamma \cup \{\Diamond(f \wedge w)\}$. Theorem 2. 8 yields that $\triangle B \cup \Gamma \models_{\mathbf{M}} f \rhd w$, and Theorem 2.7 that $\triangle B \cup \Gamma \models_{\mathbf{M}} \top \rhd \neg f$. The following figure shows that $\triangle B \cup \Gamma \not\models_{\mathbf{M}} f \rhd \neg w$ and $\triangle B \cup \Gamma \not\models_{\mathbf{M}} \top \rhd f$:



In the model, we see that when violating the prohibition to eat with your fingers, we do not get an undesired result; we are simply in the suboptimal state S_1. Thus, we can consistently model the sentences 1. and 3. using the defeasible operators.

5 Complexity and Automated Deduction

We show that validity checking w.r.t. $\mathbf{M}^{abs}$ is PSPACE-complete. We achieve this via polynomial-time reductions to the Quantified Boolean Formula (QBF) problem and back, enabling efficient automated deduction using QBF solvers. Note that validity checking in $\mathbb{PCLTA}$ is co-NP-complete. This is a consequence of the *small model property*: every formula in $\mathbb{PCLTA}$ is satisfiable by some preference model of polynomial size. Such model construction for $\mathbb{PCLTA}$ is established in [9] by extending a preorder in a given (finite) model to an arbitrary total order and then selecting a subchain of polynomial size in it. Such transformation is not possible for models with grounded ordering, since $\triangle$-subformulas may constrain certain worlds to be incomparable in any satisfying model. We can use this observation to show that satisfiability w.r.t. $\mathbf{M}^{abs}$ does not adhere to the small model property.

Lemma 3. *There is a formula $\varphi \in \mathcal{L}$ satisfiable in $\mathbf{M}^{abs}$ only by models with at least $2^{\Theta(|\varphi|)}$ states.*

Proof. Consider set Φ_n of formulas from $\mathcal{L}$:

$$\{d_0, a_0, b_0\} \cup \{\triangle(d_i \wedge l_i), \triangle(d_i \wedge r_i), \Box(l_i \wedge r_i \to \bot)\}_{i\in\{1,\ldots,n\}} \cup$$
$$\{\neg(a_k \rhd \neg(d_{k+1} \wedge a_{k+1} \wedge b_{k+1} \wedge l_{k+1}))\}_{k\in\{0,\ldots,n-1\}} \cup$$
$$\{\neg(b_k \rhd \neg(d_{k+1} \wedge a_{k+1} \wedge b_{k+1} \wedge r_{k+1}))\}_{k\in\{0,\ldots,n-1\}}$$

Let φ be a conjunction of these formulas. Then $|\varphi| = \Theta(n)$. φ is satisfiable by a model corresponding to a full binary tree of depth (in edges) n, with valuation assigned to each node as follows: d_k is true for nodes at depth at least k, a_k (resp. b_k) is true for nodes at depth exactly k or left (resp. right) immediate children of such nodes; l_i (resp. r_i) is true for nodes at depth at least i such that the i-th edge in the path to this node goes to the left (resp. right). Take the permission base $B = \{d_k \wedge l_k\}_{k\in\{1,\dots,n-1\}} \cup \{d_k \wedge r_k\}_{k\in\{1,\dots,n-1\}}$ to satisfy exactly the $\triangle$-subformulas in φ and take all valuations assigned to nodes as described above, the ideality ordering will reflect the structure of the described binary tree: if V_x, V_y, V_z are valuations assigned to nodes x, y, z then $(B, V_x) \preceq_{(B,V_z)} (B, V_y)$ iff x is a predecessor of y (or y itself). It is easy to check that all formulas in Φ_n are satisfied in such a model.

Now we show that any model M satisfying φ will always, in a sense, contain such a tree inside. Precisely, for every valuation V_x described above, there should exist a state (B', V'_x) in M such that $V_x \subseteq V'_x$. This can be proved by induction on the depth of the tree, with the root mapped to the state where φ is satisfied and left (resp. right) child of any node x at depth k mapped to an arbitrary a_k-best (resp. b_k-best) world preferred to the state corresponding to x. Therefore for every leaf in the tree the subset of $\{l_i\}_{i\in\{1,\dots,n\}} \cup \{r_i\}_{i\in\{1,\dots,n\}}$ encoding a path to it belongs to the valuation of some state in M. But due to satisfaction of $\{\Box(l_i \wedge r_i \to \bot)\}_{i\in\{1,\dots,n\}}$ all these states should be different, therefore M contains at least $\Theta(2^n)$ states.

Thus, extending the language with $\triangle$-modality enables expressing more complex model conditions, increasing the complexity of satisfiability (and hence validity) checking. Nonetheless, we can use the reasoning from [9] to transform an arbitrary satisfying model into a somewhat bounded model. In particular, we can bound polynomially the depth of the model, i.e. the length of the longest strictly ascending chain of worlds.

Lemma 4. *If $\varphi \in \mathcal{L}$ is satisfiable by some $(S, U) \in \mathbf{M}^{abs}$, then it is satisfiable by some $(S', U') \in \mathbf{M}^{abs}$ such that the length m of any ascending chain of states $S'_1 \prec_{S'} \dots \prec_{S'} S'_m$ in U' is at most $(n + 1)$, where n is the number of conidionals in φ.*

Proof. Let $\{\xi_1 \rhd \tau_1, \dots, \xi_n \rhd \tau_n\}$ be all conditionals in φ. Consider the context $U' = S \cup Best(\xi_1, S, U) \cup \dots \cup Best(\xi_n, S, U)$. $(S, U') \models \varphi$ since the evaluation of any conditional inside φ did not change. In any ascending chain in U', every state apart from one (S) belongs to $Best(\xi_i, S, U)$ for some i, and all states later in the chain cannot belong to it (due to the definition of $Best$), so any chain contains at most $(n+1)$ states.

The depth of a model can be used as a recursive parameter in the QBF-encoding of satisfiability of a formula in a model, resulting in a polynomial reduction to QBF. Conversely, we can encode any QBF formula as a set of formulas in $\mathcal{L}$, for which every satisfying model will correspond to a winning strategy in QBF-game on the given formula. Conversely, the tree-model in Lemma 3 already corresponds to a tree of all possible choices of values for variables $\{l_i\}_{i\in\{1,\dots,n\}}$, capturing universal boolean quantification over these variables. We can slightly modify the construction to also incorporate existential quantifiers, thus providing a polynomial reduction from QBF. These back-and-forth reductions imply PSPACE-completeness of validity checking w.r.t. $\mathbf{M}^{abs}$.

Theorem 4. *Validity checking w.r.t.* $\mathbf{M}^{abs}$ *is PSPACE-complete.*

Proof. We show that satisfiability checking w.r.t. $\mathbf{M}^{abs}$ is polynomially reducible to QBF, and vice versa. As usual, satisfiability checking is reducible to validity checking by negating the formula and inverting the output.

PSPACE-Membership. We construct a polynomial QBF formula[4] encoding SAT of a formula φ w.r.t. $\mathbf{M}^{abs}$ by induction on the depth of the model, relying on Lem. 4.

Let $\{p_i\}_{1\le i\le r}$, $\{\triangle\alpha_i\}_{1\le i\le m}$, and $\{\xi_i \rhd \tau_i\}_{1\le i\le n}$ be the atoms, $\triangle$-formulas and conditionals in φ. First, note that satisfaction of any subformula of φ in a state is determined by satisfaction of formulas from these three sets, so we can define a predicate $Sat^{\psi}(\mathcal{A},\mathcal{B},\mathcal{C})$ encoding the satisfaction of a subformula ψ of φ in a given state based on subsets of indices of satisfied atoms ($\mathcal{A}$), $\triangle$-subformulas ($\mathcal{B}$), and conditionals ($\mathcal{C}$).

$$Sat^{p_i}(\mathcal{A},\mathcal{B},\mathcal{C}) = i \in \mathcal{A} \qquad Sat^{\neg\psi}(\mathcal{A},\mathcal{B},\mathcal{C}) = \neg Sat^{\psi}(\mathcal{A},\mathcal{B},\mathcal{C})$$
$$Sat^{\psi_1\wedge\psi_2}(\mathcal{A},\mathcal{B},\mathcal{C}) = Sat^{\psi_1}(\mathcal{A},\mathcal{B},\mathcal{C}) \wedge Sat^{\psi_2}(\mathcal{A},\mathcal{B},\mathcal{C})$$
$$Sat^{\triangle\alpha_i}(\mathcal{A},\mathcal{B},\mathcal{C}) = i \in \mathcal{B} \qquad Sat^{\xi_i\rhd\tau_i}(\mathcal{A},\mathcal{B},\mathcal{C}) = i \in \mathcal{C}$$

We define the predicate $State^{\psi}_{d}(\mathcal{A},\mathcal{B},\mathcal{C})$ encoding that a given valuation (represented by $\mathcal{A}\subseteq\{1,\dots,r\}$) can appear in some state at depth at most d (where the depth of a state is the number of states in the longest ascending chain starting in this state) in a model, still assuming that the subsets $\mathcal{B}$ and $\mathcal{C}$ of the indices of the satisfied $\triangle$-subformulas and conditionals in φ are given. Specifically, we check that no $(\xi_i \rhd \tau_i)$ for $i \notin \mathcal{C}$ is validated in this state. Additionally, we require a given subformula ψ to be false in all strictly preferable states (to be able to ensure bestness). The predicate is defined by induction on d, with the base case $State^{\psi}_{0}(\mathcal{A},\mathcal{B},\mathcal{C}) = \bot$.

$$\begin{aligned} State^{\psi}_{d+1}(\mathcal{A},\mathcal{B},\mathcal{C}) = \forall i \in \{1,\dots,n\}\setminus\mathcal{C}.\ \neg Sat^{\tau_i}(\mathcal{A},\mathcal{B},\mathcal{C}) \vee \neg Sat^{\xi_i}(\mathcal{A},\mathcal{B},\mathcal{C}) \vee \\ \exists \mathcal{A}_i \subseteq \{1,\dots,r\}.\ State^{\psi}_{d}(\mathcal{A}_i,\mathcal{B},\mathcal{C}) \wedge Sat^{\xi_i}(\mathcal{A}_i,\mathcal{B},\mathcal{C}) \wedge \neg Sat^{\psi}(\mathcal{A}_i,\mathcal{B},\mathcal{C}) \wedge \\ \big(\bigwedge_{1\le j\le m} (j \in \mathcal{B} \wedge Sat^{\alpha_j}(\mathcal{A},\mathcal{B},\mathcal{C})) \rightarrow Sat^{\alpha_j}(\mathcal{A}_i,\mathcal{B},\mathcal{C})\big) \end{aligned}$$

We can now encode the satisfability of φ w.r.t. $\mathbf{M}^{abs}$ by first guessing which $\triangle$-subformulas and conditionals in φ are true in a satisfying model, and then checking existence of states at depth at most $(n+1)$ satisfying φ and validating every conditional $(\xi_i \rhd \tau_i)$ guessed to be true (i.e. satisfying τ_i while being one of best states for ξ_i). We encode this with the following closed QBF formula:

$$\begin{aligned} &\exists\mathcal{B}\subseteq\{1,\dots,m\}.\exists\mathcal{C}\subseteq\{1,\dots,n\}. \\ &\quad(\exists\mathcal{A}\subseteq\{1,\dots,r\}.\ State^{\bot}_{n+1}(\mathcal{A},\mathcal{B},\mathcal{C}) \wedge Sat^{\varphi}(\mathcal{A},\mathcal{B},\mathcal{C})) \wedge \\ &\quad\big(\bigwedge_{1\le i\le n} (i \in \mathcal{C} \rightarrow \exists\mathcal{A}_i\subseteq\{1,\dots,r\}.State^{\xi_i}_{n+1}(\mathcal{A}_i,\mathcal{B},\mathcal{C}) \wedge \\ &\qquad\qquad Sat^{\xi_i}(\mathcal{A}_i,\mathcal{B},\mathcal{C}) \wedge Sat^{\tau_i}(\mathcal{A}_i,\mathcal{B},\mathcal{C}))\big) \end{aligned}$$

[4] For simplicity, we use quantification over subsets of predefined finite sets of natural numbers. Every subset of S can be represented as a boolean vector of length $|S|$ (indicator function), making translation to standard Boolean quantification straightforward.

Notice that $|State^{\psi}_{d}(\mathcal{A},\mathcal{B},\mathcal{C})| = \mathcal{O}(d \cdot |\varphi|)$, so this QBF formula has a polynomial size w.r.t. $|\varphi|$. If φ is satisfiable in some $M \in \mathbf{M}^{abs}$, this QBF formula is true: there exist correct guesses of subsets $\mathcal{B}$ and $\mathcal{C}$ and required states at depth at most $(n+1)$ (due to Lem. 4). Conversely, if this QBF formula is true, we can construct $M \in \mathbf{M}^{abs}$ satisfying φ. Namely, each true predicate $State^{\psi}_{d+1}(\mathcal{A},\mathcal{B},\mathcal{C})$ (for some specific values of $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$) yields a tree of valuations of depth at most $(d+1)$ by taking valuation $\{p_i\}_{i\in\mathcal{A}}$ in the root and attaching to it trees corresponding to each true $State^{\psi}_{d}(\mathcal{A}_i,\mathcal{B},\mathcal{C})$ in the definition of this predicate. Satisfaction of the closed QBF formula above provides us with some guessed $\mathcal{B}$ and $\mathcal{C}$ and $(|\mathcal{C}|+1)$ such trees. Notice that with the permission base $\{\alpha_j\}_{j\in\mathcal{B}}$, $(B,V) \preceq_{(B,V'')} (B,V')$ if V' is a successor of V in one of the trees. To ensure that all conditionals outside $\mathcal{C}$ are false in all states we need to make these preference relations inside trees strict, which we can achieve with the same trick as in the proof of Th. 3: add fresh atom $\chi(S)$ for every node S, extend permission base with these atoms and extend all valuations in every state with its fresh atom and fresh atoms for all its predecessors in all trees. This ensures that $(\xi_i \rhd \tau_i)$ is true in the resulting model iff $i \in \mathcal{C}$, and consequently that φ is satisfied in the state corresponding to the root of the first tree. Thus, this QBF encoding indeed constitutes a polynomial reduction of satisfiability w.r.t. $\mathbf{M}^{abs}$ to QBF.

PSPACE-Hardness. Consider the following QBF-formula F^* in the alternating prefix normal form: $\forall x_1 \exists y_1 \forall x_2 \exists y_2 \ldots \forall x_n \exists y_n . F(x_1, y_1, \ldots, x_n, y_n)$. We extend the set Φ_n from the proof of Lem. 3 with the following formulas:

$$\{\triangle(d_i \wedge y_i), \triangle(d_i \wedge \neg y_i)\}_{i\in\{1,\ldots,n\}} \cup \{\Box(l_i \leftrightarrow x_i),\ \Box(r_i \leftrightarrow \neg x_i)\}_{i\in\{1,\ldots,n\}} \cup$$
$$\{\neg(d_n \rhd \neg F(x_1, y_1, \ldots, x_n, y_n))\}$$

As a result of this extension, the value of y_k in each node t at depth k will be preserved in all predecessors of t, while the value of x_k in the leaves will correspond to the direction of k-th edge in its path. Hence, such a tree will represent one strategy in the QBF-game on the formula F^*: for each choice of values of $\{x_i\}_{i\in\{1,\ldots,k\}}$ some value of y_k is chosen and carried to the leaf. It will be a winning strategy iff a valuation in every leaf satisfies $F(x_1, y_1, \ldots, x_n, y_n)$, i.e. iff the last (dual) conditional is satisfied in such a tree-model. So we can build a satisfying model on the basis of any winning strategy. Conversely, any model satisfying the conjunction of the extended set of formulas contains a tree corresponding to a winning strategy. Thus, we have a polynomial reduction from QBF to satisfiability checking w.r.t. $\mathbf{M}^{abs}$.

6 Conclusions and Perspectives

We have presented a novel logical framework for reasoning about explicit and implicit permissions. It conservatively extends the deontic system **F** + (CM) (i.e., the conditional logic $\mathbb{PCLTA}$). Its behavior is illustrated through a case study and analysis of deontic paradoxes. Additionally, we showed that validity checking in our framework is PSPACE-complete, unlike in $\mathbb{PCLTA}$. Directions for future research include:

Beyond Åqvist. Our analysis was restricted to normatively absolute models in the class $\mathbf{M}^{abs}$. This corresponds to the notion of Absoluteness in conditional logic and in

Åqvist's systems. The model class $\mathbf{M}$ is also worth investigating, as it allows permission bases to vary across states, enabling more complex examples and the representation of higher-order norms, which are central to legal theory [36]. Consider indeed the robots example in Sec. 3. From a policy-maker's perspective, we may wish to engage in meta-reasoning about second-order explicit allowances, i.e., permissions over which first-order allowances are themselves permitted. For example, a second-order allowance of the form: "it is allowed to allow Robot 1 to cross an intersection when the traffic light is flashing orange and Robot 2 is not approaching the intersection from the right" (or, from the left, if the robots are designed in the UK).

Given Friedman and Halpern's EXPTIME-completeness result for $\mathbb{PCLTU}$, it follows that validity checking for the language $\mathcal{L}$ relative to the class $\mathbf{M}$ is EXPTIME-hard. Future work will focus on establishing a tight complexity bound for this problem.

Non-monotonic reasoning. Despite the global monotonicity of the underlying entailment relation, our logic exhibits non-monotonic behavior locally and globally. Locally, our system inherits the non-monotonicity from preference-based deontic logics (and, in particular, of Åqvist's System **F** + (CM)) via the failure of strengthening of the antecedent. This feature is well-known to introduce a form of non-monotonicity in otherwise monotonic systems. More significantly, the defeasible obligations and prohibitions introduced in Section 4 exhibit global non-monotonicity, as the addition of premises can retract previously held conclusions. Future work will be devoted to studying in-depth the axiomatic properties of our non-monotonic entailment relation, also addressing typicality reasoning (thus tackling the problem identified in [16]), along the lines of [3].

Epistemic extension. We also plan to extend the semantics and language introduced in Sec. 2 with an epistemic component, in order to reason about agents' beliefs and knowledge concerning explicit and implicit permissions. In particular, we are interested in modeling scenarios in which agents have incomplete information about the environment and the permission base, and it is important to represent norms with epistemic content, such as the permission to let an agent believe or know something. For example, in the robots' scenario, we may want to represent a robot's uncertainty about the color of the traffic light due to misperception or lack of visibility, as well as the epistemic permission to let a robot know the color of the other robot's traffic light.

Dynamic extension. Last but not least, we intend to add a new family of dynamic modalities, following the style of belief base change modalities in [27,31], to model changes in permissions. In particular, we plan to consider three types of operation on permission bases: permission expansion, retraction, and revision.

Acknowledgments. Work partially supported by the Austrian Science Fund (FWF - 6372-N) in the *Logical Methods for Deontic Explanations* project, by the European Union's Horizon 2020 research and innovation programme under grant agreement No 101034440, and by the TIRIS project CaRe "Caring about Others: AI and Psychology Meet to Model and Automate Colective Reasoning".

References

1. Antoniou, G., Dimaresis, N., Governatori, G.: A system for modal and deontic defeasible reasoning. In: Orgun, M.A., Thornton, J. (eds.) AI 2007: Advances in Artificial Intelligence, 20th Australian Joint Conference on Artificial Intelligence, Proceedings. Lecture Notes in Computer Science, vol. 4830, pp. 609–613. Springer (2007)
2. Åqvist, L.: Deontic logic. In: Gabbay, D., Guenthner, F. (eds.) Handbook of Philosophical Logic: Volume II, pp. 605–714. Springer, Dordrecht (1984)
3. Britz, K., Varzinczak, I.: From klm-style conditionals to defeasible modalities, and back. J. Appl. Non Class. Logics **28**(1), 92–121 (2018). https://doi.org/10.1080/11663081.2017.1397325
4. Broersen, J.M., van der Torre, L.W.N.: Ten problems of deontic logic and normative reasoning in computer science. In: Bezhanishvili, N., Goranko, V. (eds.) Lectures on Logic and Computation - ESSLLI 2010. LNCS, vol. 7388, pp. 55–88. Springer (2011)
5. Burgess, J.P.: Quick completeness proofs for some logics of conditionals. Notre Dame J. Formal Log. **22**(1), 76–84 (1981)
6. Danielsson, S.: Preference and Obligation. Filosofiska Färeningen, Uppsala (1968)
7. Delgrande, J.P.: A preference-based approach to defeasible deontic inference. In: Calvanese, D., Erdem, E., Thielscher, M. (eds.) Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning, KR 2020. pp. 326–335 (2020)
8. Forrester, J.W.: Gentle murder, or the adverbial samaritan. The Journal of Philosophy **81**(4), 193–197 (1984)
9. Friedman, N., Halpern, J.Y.: On the complexity of conditional logics. In: Doyle, J., Sandewall, E., Torasso, P. (eds.) Proceedings of the 4th International Conference on Principles of Knowledge Representation and Reasoning (KR'94). Bonn, Germany, May 24-27, 1994. pp. 202–213. Morgan Kaufmann (1994)
10. Gabbay, D.M.: Theoretical foundations for non-monotonic reasoning in expert systems. In: Apt, K.R. (ed.) Logics and Models of Concurrent Systems. pp. 439–457. Springer Berlin Heidelberg, Berlin, Heidelberg (1985)
11. Gabbay, D., Horty, J., Parent, X., van der Mayden, R., van der Torre, L. (eds.): Handbook of Deontic Logic and Normative Systems, Volume 2. College Publications (2021)
12. Governatori, G., Olivieri, F., Rotolo, A., Scannapieco, S.: Computing strong and weak permissions in defeasible logic. Journal of Phil. Logic **42**(6), 799–829 (2013). https://doi.org/10.1007/s10992-013-9295-1
13. Hansson, B.: An analysis of some deontic logics. Noûs **3**(4), 373–398 (1969), reprinted in [15, pp. 121-147]
14. Hansson, S.O.: The varieties of permission. In: Gabbay, D.M., Horty, J., Parent, X., van der Meyden, R., van der Torre, L. (eds.) Handbook of deontic logic and normative systems, pp. 195–240. College Publications, London (2013)
15. Hilpinen, R. (ed.): Deontic Logic. Reidel, Dordrecht (1971)
16. Horty, J.: Reasons as Defaults. Oxford University Press (2012)
17. Horty, J.: Deontic modals: Why abandon the classical semantics? Pacific Philosophical Quarterly **95**(4), 424–460 (2014)
18. Kamp, H.: Iv*—free choice permission. Proceedings of the Aristotelian Society **74**(1), 57–74 (07 2015). https://doi.org/10.1093/aristotelian/74.1.57
19. Konolige, K.: A deduction model of belief. Morgan Kaufmann Publishers, Los Altos (1986)
20. Kratzer, A.: The notional category of modality. In: Eikmeyer, H.J., Rieser, H. (eds.) Words, Worlds, and Contexts. de Gruyter, Berlin / New York (1981)
21. Kratzer, A.: Modals and Conditionals. New and Revised Perspectives. Oxford University Press (2012)

22. Kraus, S., Lehmann, D., Magidor, M.: Nonmonotonic reasoning, preferential models and cumulative logics. Artificial Intelligence **44**(1), 167–207 (1990)
23. Lewis, D.K.: Counterfactuals. Harvard University Press (1973)
24. de Lima, T., Lorini, E.: Model checking causality. In: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI 2024). pp. 3324–3332. International Joint Conferences on Artificial Intelligence Organization (2024). https://doi.org/10.24963/ijcai.2024/368
25. de Lima, T., Lorini, E., Perrotin, E., Schwarzentruber, F.: A computationally grounded framework for cognitive attitudes. In: Proceedings of the Thirty-Ninth International Joint Conference on Artificial Intelligence (AAAI 2025). pp. 14858–14866. AAAI Press (2025)
26. Lorini, E.: Exploiting belief bases for building rich epistemic structures. In: Proceedings of the Seventeenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK 2019). Electronic Proceedings in Theoretical Computer Science (EPTCS), vol. 297, pp. 332–353. Open Publishing Association (2019). https://doi.org/10.4204/EPTCS.297.20
27. Lorini, E.: Rethinking epistemic logic with belief bases. Artificial Intelligence **282**, 103233 (2020). https://doi.org/10.1016/j.artint.2020.103233
28. Lorini, E.: A rule-based modal view of causal reasoning. In: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI 2023). pp. 3286–3295. International Joint Conferences on Artificial Intelligence Organization (2023). https://doi.org/10.24963/ijcai.2023/366
29. Lorini, E., Perrotin, E., Schwarzentruber, F.: Epistemic actions: Comparing multi-agent belief bases with action models. In: Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning (KR 2022). pp. 236–246. IJCAI Organization (2022). https://doi.org/10.24963/kr.2022/24
30. Lorini, E., Rapion, É.: Logical theories of collective attitudes and the belief base perspective. In: Proceedings of AAMAS 2022. pp. 833–841. International Foundation for Autonomous Agents and Multiagent Systems (2022)
31. Lorini, E., Schwarzentruber, F.: Multi-agent belief base revision. In: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21). pp. 1959–1965 (2021). https://doi.org/10.24963/ijcai.2021/270
32. MacCormick, N.: Defeasibility in law and logic. In: Bankowski, Z., White, I., Hahn, U. (eds.) Informatics and the Foundations of Legal Reasoning, pp. 99–117. Springer Netherlands (1995)
33. Makinson, D.: General theory of cumulative inference. In: Reinfrank, M., de Kleer, J., Ginsberg, M.L., Sandewall, E. (eds.) Non-Monotonic Reasoning. pp. 1–18. Springer Berlin Heidelberg, Berlin, Heidelberg (1989)
34. Nute, D. (ed.): Defeasible Deontic Logic. Kluwer, Dordrecht (1997)
35. Parent, X.: Preference semantics for hansson-type dyadic deontic logic: a survey of results. In: Handbook of Deontic Logic and Normative Systems, Volume 2, pp. 7–70. College Publications (2021)
36. Raz, J.: The Authority of Law: Essays on Law and Morality. Oxford University Press, Oxford, 2nd edn. (2009), first published 1979 by Clarendon Press
37. Ross, A.: Imperatives and logic. Philosophy of Science **11**(1), 30–46 (1944). https://doi.org/10.2307/2268596
38. von Wright, G.H.: Deontic logic. Mind **60**(237), 1–15 (1951). https://doi.org/10.1093/mind/LX.237.1