

ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Modelling Ceteris Paribus Preferences with Deontic Logic

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Availability:
This version is available at: https://hdl.handle.net/11585/872774 since: 2022-02-28

Published:

DOI: http://doi.org/10.1093/logcom/exab088

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

28 March 2024

This is the final peer-reviewed accepted manuscript of:

Andrea Loreggia, Emiliano Lorini, Giovanni Sartor, Modelling Ceteris Paribus Preferences with Deontic Logic, *Journal of Logic and Computation*, Volume 32, Issue 2, March 2022, Pages 347–368

The final published version is available online at: **https://doi.org/10.1093/logcom/exab088**

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/)

When citing, please refer to the published version.

Modeling Ceteris Paribus Preferences with Deontic Logic

Andrea Loreggia[2][0000−0002−9846−0157], Emiliano Lorini[3][0000−0002−7014−6756], and Giovanni Sartor[1,2][0000−0003−2210−0398] ⋆ ⋆⋆

[1] CIRSFID - Alma AI, University of Bologna, Italy
[2] European University Institute, Florence, Italy
[3] IRIT, CNRS, Toulouse University

Abstract. We present a formal semantics for deontic logic based on the concept of ceteris paribus preferences. We introduce notions of unconditional obligation and permission as well as conditional obligation and permission that are interpreted relative to this semantics. We show that these notions satisfy some intuitive properties and, at the same time, do not encounter some problems and paradoxes that have been extensively discussed in the deontic logic literature. We show that the fragment of our logic in which the content of a deontic operator is a literal has an equivalent representation based on CP-nets.

Keywords: Deontic Logic · Ceteris paribus preferences · CP-net.

1 Introduction

Artificial agents are used to automate tasks in many scenarios. They are so pervasive and fast that it is almost impossible for humans to monitor them in order to prevent illegal or unethical behaviour. A possible solution is to embed (aspects of) normative governance into such agents. This requires translating legal and ethical requirements into computable representations of legal knowledge and reasoning. The basic component of normative knowledge consists in obligations and permissions. Obligations impose requirements while permissions confer allowances.

Our approach to model obligations and permission connects deontic logic and preference models.

Deontic logic has been viewed as a key component of logical models of normative knowledge and reasoning. It provides a set of formal tools, usually based on modal logic [2, 13] to capture normative notions, which can be compositionally integrated with other logical formalism, such as predicate logic, logic programming, defeasible logics, logics of action [5, 29].

⋆ A. Loreggia and G. Sartor have been supported by the H2020 ERC Project "CompuLaw" (G.A. 833647). E. Lorini would like to gratefully acknowledge the support from the ANR-3IA Artificial and Natural Intelligence Toulouse Institute.

⋆⋆ Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Preferences are central to decision processes and thus are implemented in many frameworks to drive, assist, or influence individuals (e.g., recommender system [52], sentiment analysis [17, 18], and in deep learning ensemble methods [10]). By comparing agents' preferences, it is possible to assess the similarity of their evaluative attitudes, [31, 39]. Similarly, agents' behaviour can be compared with exogenous priorities, based on moral principles, legal requirements, guidelines or expected business processes. In this way it is possible to assess the extent to which the agent deviates from what is desired [45, 32, 33, 35].

A usual criticism of standard deontic logic is that it focuses on ideal worlds, namely on a set of worlds in which all obligations are satisfied, and each permission can be implemented. Unfortunately real life situations do not meet this description: even if some obligations are not complied with, still the involved agents have to determine what they should do, relatively to the other obligations at stake. A ceteris-paribus based deontic logic provides for this situation, since it focuses on partial improvements, rather than on perfect worlds, namely in those changes in the given worlds which could make it better, all remaining aspects being unchanged (e.g., possibly imperfect as before).

To connect deontic logic and preferences, we provide semantics for a deontic logic based on the idea that obligations and permissions consist in ceteris paribus preferences over worlds: strict ones for obligations and weak ones for permissions.

Such preferences are *ceteris paribus* in the sense that they only concern worlds that are equal in all remaining circumstances, namely, in all aspects except for those contributing to the states of affairs that are affirmed to be obligatory or permitted. Thus, deontic propositions are to be evaluated against model-theoretical frames consisting of sets of worlds over which ceteris paribus preferences are defined.

In this work, we are also interested in making a connection between the area of deontic logic and artificial intelligence. We think cooperation among researchers from the two areas can be very beneficial, and possibly lead to a new workable approach to model norms and reason about them. There have been various attempts at basing a deontic logic on the idea of preferences (see for instance [23, 2, 24]). The notion of ceteris paribus preferences is employed in literature for modeling deontic concepts such as contrary-to-duties (e.g., [4, 9, 15, 37]) and they also provide the intuition at the basis of a preference framework named conditional preference networks or CP-nets [6]. CP-nets, and their variants, are emerging preference frameworks that allow representing conditional preferences in a compact way. The framework is adopted in many different scenarios. For instance, in recommender systems, they are employed to improve the accuracy of personalized search [52]. A first attempt to model deontic notions with CP-nets has been done in [14].

By modeling the normative system of a multi-agent system with a CP-net we obtain a compact representation that would allow us to check important properties of the modeled normative system, like for instance its consistency [36]. When preferences of agents are also modeled through CP-nets, it is possible to check their compliance against exogenous preferences (e.g., normative systems,

laws, ethical principles) [40, 35, 44]. In this regard, recently, new approaches define metric spaces over structured preferences [31, 39, 34], providing feasible ways for preferences comparison.

Our formalisation does not allow for the derivation of deontic paradoxes, but supports some deontic inferences. In future work, we plan to complement our work with a dynamical analysis of the way in which prescriptive acts generate ceteris paribus preferences, so determining the truth-values of primary and inferred deontic propositions. This paper expands results published in [37, 38], and discusses some limitations of our current definitions, which we hope we will be able to address in our future research.

The paper is organized as follows: we first provide a formal account of the idea of ceteris paribus preferences and of the corresponding semantic structures. We then formalise conditional and unconditional obligations and permissions as ceteris paribus preferences, and study their basic logical properties. We illustrate and discuss both ceteris paribus preferences and deontic operators with the help of extensive examples. We show that the satisfiability problem for the proposed languages is P-complete. This is done by providing a polynomial procedure that allows to transform our language into the one proposed in [19]. We also put the basis for a syntax independence of the ceteris paribus deontic logic. Finally, we show how existing AI formalisms (such as CP-nets [6]) can be used to represent our idea of ceteris paribus deontic logic.

2 Related work

In this section we shall provide some background for our proposal, by referring first to deontic logic and then to preferences.

2.1 Deontic logic

An intuitive semantics and a simple axiomatisation for deontic notions are provided by the so-called standard deontic logic or SDL [11], built on the basis of the so-called old system of deontic logic by [48], (for a discussion of SDL, see [27, 28]). In SDL, to be obligatory means to be true in all ideal (perfectly good) worlds, and to be permissible means to be true in at least one ideal world. This idea is captured by a serial accessibility relation R over possible worlds, to be understood as an ideality relation: for every world u there exists at least one world v, such that uRv (v is ideal, relatively to u). In such a semantics frame, φ is obligatory in a world u if and only if φ is true in every world v such that uRv, and φ is permitted in a world u if and only if there exists a world v such that φ is true in at v and uRv. A corresponding axiomatisation can be obtained by adding to propositional logic the deontic **K** principle $\mathsf{O}(\varphi \to \psi) \to (\mathsf{O}\varphi \to \mathsf{O}\psi)$, the principle of deontic consistency $\mathsf{O}\varphi \to \mathsf{P}\varphi$, and the rule of necessitation $\frac{\varphi}{\mathsf{O}\varphi}$ according to which if φ is a logical necessity then it is obligatory.

As it has been often remarked, SDL gives rise to several apparently counter-intuitive implications, the so-called deontic paradoxes (for a discussion, see [2],

for an analysis from the perspective of legal theory, see also [53, 21]). First of all, as legal theorist Alf Ross critically observed [43], the obligation of a certain proposition should not entail the obligation concerning the disjunction of that proposition and any other arbitrary proposition. For instance, the obligation that a letter is posted ($\mathsf{O}p$), should not entail the obligation that the letter is posted or burned, $\mathsf{O}(p \vee b)$. In SDL, however, $\mathsf{O}\varphi$ entails $\mathsf{O}(\varphi \vee \psi)$. It is easy to see that this entailment is sound according to the semantics of SDL: if in all ideal worlds it is the case that φ in all such worlds it must also be the case that $\varphi \vee \psi$. Other paradoxes have to do with the so-called contrary to duty obligations, namely, with obligations that emerge when other obligations are violated, and whose content may contradict obligations holding when no violation takes place. The classical examples are the paradoxes by Forrester (the gentle murderer), Roderick Chishol [8] and by Marek Sergot and Henry Prakken [41]. Various solutions have been proposed to address contrary to duty obligations, often involving technical complexities and sometimes giving rise to additional problems [7].

A further problematic aspect of SDL concerns conditional or rather contextual obligations, namely, assertions to the effect that a certain proposition φ is obligatory under a certain condition ψ. Neither a conditional of classical propositional logic. i.e., $\varphi \rightarrow \mathsf{O}\psi$, nor the embedment of a such a conditional within a deontic operator. i.e., $\mathsf{O}(\varphi \rightarrow \psi)$, appear to provide fully convincing solutions. This issue has spawned the development of dyadic deontic logic, which captures deontic conditionality through a special conditional operator. Dyadic deontic logic was initiated by Georg Henrik Von Wright [50], while a semantics for it was first proposed by Bengt Hansson [22].

Technical solutions have been proposed to deal with deontic conditionals and contrary to duty obligations (see [41, 7]). These solutions, however, generally require a more complex logical framework and a less intuitive semantics, in comparison with SDL (see [26]).

2.2 Preferences

The idea of ceteris paribus preference was originally introduced by Von Wright [49, 51].

Von Wright observes that our preferences over states of affairs (as opposed to preferences over objects) are usually "holistic", in the sense that they address the compared states of affairs—denoted by Boolean combination of atoms—in the context of the circumstances accompanying such states of affairs: "What makes a person at one time prefer a state of affairs p to a state of affairs q, and at another time perhaps prefer q to p, can be the fact that the circumstances, i.e., the other states beside p and q themselves, which obtain or are taken into account at the first time, are different from the circumstances obtaining or being taken into account at the other time" [51].

However, there is a sense in which we may claim that a certain state of affairs is preferred to another state of affairs without making this preference conditional on particular circumstances. This is the case in which a state of affairs is preferred

to another in every possible context. In such a case we say that the first state of affairs is preferred ceteris paribus (all the rest being equal).

To capture the idea of a holistic preference, von Wright considers a set of atoms $Atm = \{p_1, \ldots, p_n\}$, each describing an elementary and independent state of a complete situation, or world. Since each atom can be true or false in any possible world, Atm originates a powerset $W = 2^{Atm}$ of distinct possible worlds. Von Wright [51] observes that the set $W = 2^{Atm}$ does not need to account for all states that can exist in the real world. It is rather limited to the "*preference horizon* of a given subject at a given time", namely, to the "states which the subject takes into consideration as constituting accompanying circumstances when he contemplates his preference or not-preferences between states".

Ceteris paribus preferences have recently been the object of renewed interest by logicians, who have developed ceteris paribus semantics for action and preference [19, 46]. It has also been remarked [5] that preference logics and preference representation languages are both concerned with reasoning about preferences over combinatorial domains and that in both areas the ceteris paribus principle plays a key role in the interpretation of preference statements [46].

Our notion of an obligation can be connected to the notion of goal introduced in [54], in which a goal represents a partition of possible states, those which satisfy the goal being ceteris paribus preferable to those that do not satisfy the goal.

3 The ceteris paribus Deontic Logic-CPDL

The basic idea of this paper is that the semantics of obligations and permissions can be captured by viewing obligations as strict ceteris paribus preferences and permissions as weak ceteris paribus preferences. We call ceteris paribus Deontic Logic (CPDL) the resulting deontic logic of obligation and permission.

In this section, we provide a formal definition of the relevant concepts. In the next section we shall discuss them and exemplify their application.

3.1 Ceteris paribus Preferences

We first present the basic elements of the formal semantics, namely, the concepts of preference model and ceteris paribus preference.

Let Atm be a non-empty finite set of atomic propositions and let $Lit = Atm \cup \{\neg p : p \in Atm\}$ be the corresponding set of literals.

Definition 1 (Preference model). *A preference model is a tuple $M = (W, \preceq)$ such that: $W = 2^{Atm}$ is the set of worlds, and $\preceq$ is a complete preorder on W.*[4]

Elements of W are denoted by $w, v, \ldots$. Note that we assume that the set W is complete, i.e., that it includes all possible words (relative to atoms Atm). We write $w \preceq v$ meaning that v is at least as good/ideal as w. As usual, we

[4] That is a binary relation on W which is reflexive, transitive and complete.

define $w \prec v$, meaning that world v is better/more ideal than world w, to be an abbreviation of $w \preceq v$ and $v \not\preceq w$. Moreover, we define $w \approx v$, meaning that v is equivalent to w, to be an abbreviation of $w \preceq v$ and $v \preceq w$. The class of preference models is denoted by $\mathcal{P}$.

A weak preference model differs from a preference model as it does not necessarily include all valuations of propositional variables. Specifically, W is a subset of the set of all the possible worlds. In particular:

Definition 2 (Weak preference model). *A weak preference model is a tuple $M = (W, \preceq)$ such that $W \subseteq 2^{Atm}$ is the set of worlds, and $\preceq$ is a complete preorder on W.*

In the following, unless differently specified, we shall only make use of preference model tout court. We shall consider weak preference models only in section 4.2, to deal with contrary-to-duty obligations.

Let us introduce the following concepts of *circumstantial* indistinguishability and *circumstantial* preference.

Definition 3 (Circumstantial Indistinguishability). *Let $M = (W, \preceq)$ be a preference model, let $w, v \in W$ and let X be a finite set of atomic propositions. We say that $w \equiv_X v$ iff $\forall p \in X : p \in w$ iff $p \in v$.*

According to definition 3, $w \equiv_X v$ means that w and v are indistinguishable, with regard to the circumstances (the atoms) in X.

Definition 4 (Circumstantial Preference). *Let $M = (W, \preceq)$ be a preference model, let $w, v \in W$ and let X be a finite set of atomic propositions. We introduce the following abbreviations: $w \preceq_X v$, iff $w \equiv_X v$ and $w \preceq v$, $w \prec_X v$, iff $w \equiv_X v$ and $w \prec v$ respectively.*

According to definition 4, $w \preceq_X v$ means that v is at least as good as w, the two worlds being indistinguishable relative to X. Correspondingly, $w \prec_X v$ means that v is better than w, the two worlds being indistinguishable relative to X. On the basis of the notions of circumstantial equivalence and preference, we can characterise the notions of ceteris paribus (all-the-rest-being equal) preference relative to an atom set X.

Definition 5 (Ceteris Paribus Preference). *A world w is* ceteris paribus at least as good as *or* ceteris paribus better than *a world v apart from Y, if $v \preceq_{Atm \setminus Y} w$ or $v \prec_{Atm \setminus Y} w$ respectively.*

The former definition concerns indistinguishability and preference relative to all atoms not in Y, i.e., relatively to $Atm \setminus Y$.

3.2 Unconditional obligations and permissions: formalisation

On the basis of the notions introduced in the previous section, we shall now address obligations and permission.

Definition 6. *$\mathcal{L}_{\mathrm{CPDL}}(Atm)$ is the modal language which consists of atomic propositions $p, q, \ldots \in Atm$, standard Boolean operators and the modal operators $\mathsf{O}, \mathsf{P}, \mathsf{U}$. More precisely, it is the smallest set such that:*

- *if $p \in Atm$, then $p \in \mathcal{L}_{\mathrm{CPDL}}$*
- *if $\varphi, \psi \in \mathcal{L}_{\mathrm{CPDL}}$, then $\neg\varphi, \varphi \wedge \psi \in \mathcal{L}_{\mathrm{CPDL}}$*
- *if $\varphi, \psi \in \mathcal{L}_{\mathrm{CPDL}}$, then $\mathsf{O}\varphi, \mathsf{P}\varphi, \mathsf{U}\varphi \in \mathcal{L}_{\mathrm{CPDL}}$.*

We note by $\mathcal{L}_{\mathrm{CPDL-Prop}}(Atm)$ the fragment of the language $\mathcal{L}_{\mathrm{CPDL}}(Atm)$ in which formulas in the scope of deontic operators O and P can only be propositional.

Boolean constructions $\top$, $\bot$, $\vee$, $\rightarrow$ and $\leftrightarrow$ are defined from p, $\neg$ and $\wedge$ in the standard way.

Formulas $\mathsf{O}\varphi$ and $\mathsf{P}\varphi$ have to be read, respectively, "φ is obligatory" and "φ is permitted". Formula $\mathsf{U}\varphi$ has to be read "φ is universally true". The truth conditions for the formulas in the language $\mathcal{L}_{\mathrm{CPDL}}(Atm)$ are defined as follows:

Definition 7 (Truth Conditions). *Let $M = (W, \preceq)$ be a preference model and let $w \in W$. Then:*

$$
\begin{aligned}
M, w \models p &\iff p \in w \\
M, w \models \neg\varphi &\iff M, w \not\models \varphi \\
M, w \models \varphi \wedge \psi &\iff M, w \models \varphi \text{ and } M, w \models \psi \\
M, w \models \mathsf{O}\varphi &\iff \exists v \in W \text{ such that } M, v \models \varphi, \text{ and} \\
&\qquad \forall v, u \in W : \text{ if } M, v \models \varphi \text{ and } v \preceq_{Atm \setminus Atm(\varphi)} u \text{ then } M, u \models \varphi \\
M, w \models \mathsf{P}\varphi &\iff \exists v \in W \text{ such that } M, v \models \varphi, \text{ and} \\
&\qquad \forall v, u \in W : \text{ if } M, v \models \varphi \text{ and } v \prec_{Atm \setminus Atm(\varphi)} u \text{ then } M, u \models \varphi \\
M, w \models \mathsf{U}\varphi &\iff \forall v \in W : M, v \models \varphi
\end{aligned}
$$

where $Atm(\varphi)$ denotes the set of atoms from Atm occurring in φ.

It is worth noting that the interpretation of the obligation and permission operator could be equivalently stated as follows:

$$
\begin{aligned}
M, w \models \mathsf{O}\varphi &\iff \exists v \in W \text{ such that } M, v \models \varphi, \text{ and} \\
&\qquad \forall v, u \in W : \text{ if } M, u \models \varphi, M, v \models \neg\varphi \text{ and } v \equiv_{Atm \setminus Atm(\varphi)} u \text{ then} \\
&\qquad v \prec u \\
M, w \models \mathsf{P}\varphi &\iff \exists v \in W \text{ such that } M, v \models \varphi, \text{ and} \\
&\qquad \forall v, u \in W : \text{ if } M, u \models \varphi, M, v \models \neg\varphi \text{ and } v \equiv_{Atm \setminus Atm(\varphi)} u \text{ then} \\
&\qquad v \preceq u
\end{aligned}
$$

In other words, $\mathsf{O}\varphi$ means that φ is possible and, for every two possible worlds that are $Atm \setminus Atm(\varphi)$-indistinguishable and that disagree about the truth value of φ, the world in which φ is true is better than the world in which φ is false. $\mathsf{P}\varphi$ means that φ is possible and, for every two possible worlds that are $Atm \setminus Atm(\varphi)$-indistinguishable and that disagree about the truth value of

φ, the world in which φ is true is at least as good as the world in which φ is false.

Note that the valuation of the obligation operator does not depend on the actual world. Either $\mathsf{O}\varphi$ is true at all worlds of the model or at none of them. This is justified by the fact that we are not providing a dynamic model of obligations and permissions whereby norms may differ from world to world. Our model is meant to capture a normative system at a single point in time and no space.

We say that the formula $\varphi \in \mathcal{L}_{\mathrm{CPDL}}(Atm)$ is valid relative to the class of preference models $\mathcal{P}$, denoted by $\models_{\mathcal{P}} \varphi$, iff, for every preference model M and for every world w in M, we have $M, w \models \varphi$. We say that the formula $\varphi \in \mathcal{L}_{\mathrm{CPDL}}(Atm)$ is satisfiable relative to the class of preference models iff, there exists a preference model M and a world w in M, such that $M, w \models \varphi$.

3.3 Conditional obligations and permissions: formalisation

In this section we extend the logic CPDL by operators of conditional obligation and conditional permission. We call CPDL$^+$ the resulting logic.

Definition 8. *$\mathcal{L}_{\mathrm{CPDL}^+}(Atm)$ is a modal language which consists of atomic propositions $p, q, \ldots \in Atm$, standard Boolean operators and the modal operators $\mathsf{O}, \mathsf{P}, \mathsf{U}$. More precisely, it is the smallest set such that:*

- *if $p \in Atm$, then $p \in \mathcal{L}_{\mathrm{CPDL}^+}$*
- *if $\varphi, \psi \in \mathcal{L}_{\mathrm{CPDL}^+}$, then $\neg\varphi, \varphi \wedge \psi \in \mathcal{L}_{\mathrm{CPDL}^+}$*
- *if $\varphi, \psi \in \mathcal{L}_{\mathrm{CPDL}^+}$, then $\mathsf{O}\varphi, \mathsf{P}\varphi, \mathsf{O}(\psi|\varphi), \mathsf{P}(\psi|\varphi), \mathsf{U}\varphi \in \mathcal{L}_{\mathrm{CPDL}^+}$.*

We note by $\mathcal{L}_{\mathrm{CPDL}^+ - \mathrm{Prop}}(Atm)$ the fragment of the language $\mathcal{L}_{\mathrm{CPDL}^+}(Atm)$ in which formulas in the scope of deontic operators O and P can only be propositional.

Formulas $\mathsf{O}(\psi|\varphi)$ and $\mathsf{P}(\psi|\varphi)$ have to be read, respectively, "under condition ψ, φ is obligatory " and "under condition ψ, φ is permitted". The truth conditions for the formulas in the language $\mathcal{L}_{\mathrm{CPDL}^+}(Atm)$ are the ones given in Definition 7 *plus* the following two extra truth conditions for the conditional obligation operator and the conditional permission operator:

Definition 9 (Truth conditions (cont.)). *Let $M = (W, \preceq)$ be a preference model and let $w \in W$. Then:*

$$
\begin{aligned}
M, w \models \mathsf{O}(\psi|\varphi) \iff & \exists v \in W \text{ such that } M, v \models \varphi \wedge \psi, \text{ and} \\
& \forall v, u \in ||\psi||_M : \text{ if } M, v \models \varphi \text{ and} \\
& v \preceq_{Atm \setminus Atm(\varphi)} u \text{ then } M, u \models \varphi \\
M, w \models \mathsf{P}(\psi|\varphi) \iff & \exists v \in W \text{ such that } M, v \models \varphi \wedge \psi, \text{ and} \\
& \forall v, u \in ||\psi||_M : \text{ if } M, v \models \varphi \text{ and} \\
& v \prec_{Atm \setminus Atm(\varphi)} u \text{ then } M, u \models \varphi
\end{aligned}
$$

where $||\psi||_M = \{w \in W : M, w \models \psi\}$ is the truth set of ψ relative to the preference model M.

The definitions of validity and satisfiability for the formulas in $\mathcal{L}_{\text{CPDL}^+}(Atm)$ relative to preference models are analogous to the definitions of validity and satisfiability for the formulas in $\mathcal{L}_{\text{CPDL}}(Atm)$ relative to preference models.

4 Discussion and examples

In this section, we describe several properties of our logic as well as showing that our model does not produce several well-known paradoxes in deontic logic. We give more insights by using detailed examples. We shall first discuss the general properties of our logic.

4.1 Properties of CPDL

In this section we shall illustrates some basic validities of our logic, which enable corresponding deontic inferences.

Proposition 1. *For all* $\varphi \in \mathcal{L}_{\text{CPDL}^+}(Atm)$*:*

$$\models_{\mathcal{P}} \mathsf{O}\varphi \leftrightarrow \mathsf{O}(\top|\varphi)$$
$$\models_{\mathcal{P}} \mathsf{P}\varphi \leftrightarrow \mathsf{P}(\top|\varphi)$$

This highlights that unconditional obligation and permission do not need to be added as primitives in the language of the logic CPDL$^+$, as they are definable from conditional obligation and permission.

The following proposition highlights that if φ is obligatory, then it is also permitted:

Proposition 2. *For all* $\varphi \in \mathcal{L}_{\text{CPDL}^+}(Atm)$*:*

$$\models_{\mathcal{P}} \mathsf{O}\varphi \rightarrow \mathsf{P}\varphi$$

Before dealing with deontic paradoxes, let us consider how factual detachment is represented in the context of the logic CPDL$^+$:

Proposition 3. *For all* $\varphi \in \mathcal{L}_{\text{CPDL}^+}(Atm)$*:*

$$\models_{\mathcal{P}} (\mathsf{O}(\psi|\varphi) \wedge \mathsf{U}\psi) \rightarrow \mathsf{O}\varphi$$
$$\models_{\mathcal{P}} (\mathsf{P}(\psi|\varphi) \wedge \mathsf{U}\psi) \rightarrow \mathsf{P}\varphi$$

This means that if the condition of a conditional obligation/permission is necessarily true, then the obligation/permission is detached and becomes unconditional.

Let us consider the well-known Ross's paradox [43]. In standard deontic logic (SDL), an obligation to mail a letter (i.e., $\mathsf{O}m$) implies the obligation to mail a letter or to burn it (i.e., $\mathsf{O}(m \vee b)$), something that goes against intuition. As the following preference model highlights, our logic CPDL does not encounter this problem.

Example 1. Let $Atm = \{m, b\}$ with $w_1 = \{m, b\}$, $w_2 = \{m\}$, $w_3 = \{b\}$ and $w_4 = \emptyset$. Let us suppose the following preference ordering over the worlds in W: $w_3 \prec w_1 \prec w_4 \prec w_2$. We clearly have $M, w_1 \models \mathsf{O}m \wedge \neg\mathsf{O}(m \vee b)$.

More generally, it is worth noting that the ceteris paribus obligation operator is not normal as it does not satisfy Axiom K. In particular, there exists $\varphi, \psi \in \mathcal{L}_{\text{CPDL}^+}(Atm)$ such that the formula $(\mathsf{O}\varphi \wedge \mathsf{O}(\varphi \rightarrow \psi)) \rightarrow \mathsf{O}\psi$ is not valid in CPDL. To show this, it is sufficient to consider the preference model in Example 1. We have $M, w_1 \models (\mathsf{O}m \wedge \mathsf{O}(m \rightarrow (m \vee b)) \wedge \neg\mathsf{O}(m \vee b)$.

An interesting property of the ceteris paribus operators for obligation and permission concerns aggregation over conjunction. First of all, it is worth noting that, in the general case, ceteris paribus obligation and permission do not aggregate over conjunction. More precisely, there exists $\varphi, \psi \in \mathcal{L}_{\text{CPDL}^+}(Atm)$ such that: $\not\models_{\mathcal{P}} (\mathsf{O}\varphi \wedge \mathsf{O}\psi) \rightarrow \mathsf{O}(\varphi \wedge \psi), \not\models_{\mathcal{P}} (\mathsf{P}\varphi \wedge \mathsf{P}\psi) \rightarrow \mathsf{P}(\varphi \wedge \psi)$

To show this it is sufficient to consider the following example. (The latter is proved in an analogous way.)

Example 2. Let $Atm = \{p, q\}$ and $w_1 = \{p, q\}, w_2 = \{q\}, w_3 = \{p\}, w_4 = \emptyset$. Moreover, let us consider the following preference order $\preceq$: $w3 \prec w1 \prec w4 \prec w2$.

It is routine exercise to check that $M, w_1 \models \mathsf{O}(p \rightarrow q)$, $M, w_1 \models \mathsf{O}q$, but $M, w_1 \not\models \mathsf{O}((p \rightarrow q) \wedge q)$. To verify the latter it is sufficient to observe that $w_1 \preceq w_4$.

Nonetheless, if φ and ψ are conjunctive clauses (i.e., finite conjunctions of literals from Lit) whose sets of atoms have empty intersection (i.e., φ and ψ are independent formulas), then the obligation/permission that φ and the obligation/permission that ψ aggregate over conjunction.

Proposition 4. *If φ, ψ are conjunctive clauses and $Atm(\varphi) \cap Atm(\psi) = \emptyset$ then:*

$$\models_{\mathcal{P}} (\mathsf{O}\varphi \wedge \mathsf{O}\psi) \rightarrow \mathsf{O}(\varphi \wedge \psi)$$
$$\models_{\mathcal{P}} (\mathsf{P}\varphi \wedge \mathsf{P}\psi) \rightarrow \mathsf{P}(\varphi \wedge \psi)$$

Proof. We only prove the former as the latter is proved in an analogous way. We prove it by reductio ad absurdum. Let us suppose that (i) $M, w \models \mathsf{O}\varphi \wedge \mathsf{O}\psi$ and (ii) $M, w \not\models \mathsf{O}(\varphi \wedge \psi)$. Item (i) means that:

(A) $\forall v, u \in W :$ if $M, v \models \varphi$ and $v \equiv_{Atm \setminus Atm(\varphi)} u$ and $v \preceq u$ then $M, u \models \varphi$, and $\forall v, u \in W :$ if $M, v \models \psi$ and $v \equiv_{Atm \setminus Atm(\psi)} u$ and $v \preceq u$ then $M, u \models \psi$, AND
(B) $\exists z, z' \in W$ such that $M, z \models \varphi$ and $M, z' \models \psi$.

Item (ii) means that:

(C) $\exists v, u \in W : M, v \models \varphi \wedge \psi$ and $v \equiv_{Atm \setminus Atm(\varphi \wedge \psi)} u$ and $v \preceq u$ and $M, u \models \neg\varphi \vee \neg\psi$, OR
(D) $\forall z \in W$, $M, z \models \neg\varphi \vee \neg\psi$.

Let us suppose that (C) holds. We consider three possible cases.

Case 1: $M, u \models \neg\varphi \wedge \psi$. Since φ and ψ are conjunctive clauses we have $v \equiv_{Atm \setminus Atm(\varphi)} u$. But this is in contradiction with item (i) above.

Case 2: $M, u \models \varphi \wedge \neg\psi$. Since φ and ψ are conjunctive clauses we have $v \equiv_{Atm \setminus Atm(\psi)} u$. But this is in contradiction with item (i) above.

Case 3: $M, u \models \neg\varphi \wedge \neg\psi$. There exists $w \in W$ such that $M, w \models \neg\varphi \wedge \psi$ and $v \equiv_{Atm \setminus Atm(\varphi)} w$ and $w \equiv_{Atm \setminus Atm(\psi)} u$. It is sufficient to consider the world w such that $\forall p \in Atm(\varphi) : p \in w$ iff $p \in u$, $\forall p \in Atm(\psi) : p \in w$ iff $p \in v$, and $\forall p \in Atm \setminus (Atm(\varphi) \cup Atm(\psi)) : p \in w$ iff $p \in v$. Such a world w exists since $Atm(\varphi) \cap Atm(\psi) = \emptyset$. By item (i) above and the fact that $\preceq$ is complete, we have that $w \preceq v$. Hence, by the transitivity of $\preceq$, we have $w \preceq u$. The latter is in contradiction with item (A) above.

Now, let us suppose that (D) holds. Since φ and ψ are assumed to be conjunctive clauses such that $Atm(\varphi) \cap Atm(\psi) = \emptyset$ and M is a preference model such that $W = 2^{Atm}$, item (B) above implies that $\exists z \in W$ such that $M, z \models \varphi \wedge \psi$. The latter is in contradiction with item (D). □

As shown in Example 2, the formula $(\mathsf{O}(p \to q) \wedge \mathsf{O}q) \to \mathsf{O}((p \to q) \wedge q)$ is not valid in our logic. Notice that $((p \to q) \wedge q)$ is logically equivalent to q. This highlights a more general property of the logic CPDL, namely, the fact that the obligation operator is not closed under logical equivalence. The same property holds for permissions. On the one hand, this might be seen as a limitation of Von Wright's approach to ceteris paribus preferences extended here to ceteris paribus permissions and obligations. Indeed, closure under logical equivalence is the minimal property that any classical modal logic has to satisfy. Thus, our logic of obligations is not a classical modal logic. On the other hand, not meeting closure under logical equivalence might be acceptable from the point of view of the imperative theory of norms defended, among the others, by Von Wright. It has been argued that an obligation, seen as an imperative or a command that a certain fact ought to be the case, does not necessarily imply that all logically equivalent facts are obligatory as well ([30], [21]).

In our logic the principle of the substitution of a formula φ with a logically equivalent formula ψ in a deontic expression $\mathsf{O}(\chi|\varphi)$ or $\mathsf{P}(\chi|\varphi)$ only holds if exactly the same atoms occur in φ and ψ.

However, we agree that in many contexts closure under logical equivalence may be a desirable feature of a deontic logic. In Section 6 we introduce a variant of our logic that provides such feature.

It also worth noting that the obligation operator does not satisfy weakening. Indeed, the formula $\mathsf{O}(p \wedge q) \to \mathsf{O}p$ is not valid in the class of preference models for $p \neq q$. Nonetheless, as the following proposition highlights, we have a weaker property.

Proposition 5. *For all $p, q \in Atm$ such that $p \neq q$:*

$$\models_{\mathcal{P}} \mathsf{O}(p \wedge q) \to \neg\mathsf{O}\neg p$$

Proof. Let us suppose that $M, w \models \mathsf{O}(p \wedge q)$. The latter implies that:

(A) $\forall v, u \in W$: if $M, v \models p \wedge q$ and $v \equiv_{Atm \setminus Atm(p \wedge q)} u$ and $v \preceq u$ then $M, u \models p \wedge q$.

Since M is a preference model such that $W = 2^{Atm}$, we have:

(B) $\exists z, z' \in W$ such that $M, z \models p \wedge q$, $M, z' \models \neg p \wedge q$ and $z \equiv_{Atm \setminus Atm(p)} z'$.

Items (A) and (B) together imply that:

(C) $\exists z, z' \in W$ such that $M, z \models p$ and $M, z' \models \neg p$ and $z \equiv_{Atm \setminus Atm(p)} z'$ and $z' \preceq z$.

The latter means that $M, w \models \neg \mathsf{O} \neg p$. □

Before concluding, we would like to discuss a validity concerning nesting of obligations. In our logic, if φ is obligatory then it is obligatory that φ is obligatory, that is:

$$\models_{\mathcal{P}} \mathsf{O}\varphi \rightarrow \mathsf{OO}\varphi$$

This is due to the fact that (i) if φ is obligatory then it is necessarily the case that φ is obligatory (i.e., $\mathsf{O}\varphi \rightarrow \mathsf{UO}\varphi$) and (ii) if φ is necessary then φ is obligatory (i.e., $\mathsf{U}\varphi \rightarrow \mathsf{O}\varphi$). To prevent such an entailment, a further condition should be added to the definition of obligation, namely, the fact that for φ to be obligatory, $\neg\varphi$ has to be possible (i.e., $\exists v \in W$ such that $M, v \models \neg\varphi$). By adding this negative condition to the definition of the obligation operator, obligations about tautologies become impossible (i.e., $\neg\mathsf{O}\top$ becomes valid), which is often considered a requirement for deontic logic [47]. Also, thanks to the previous negative condition, $\mathsf{O}\varphi \rightarrow \neg\mathsf{OO}\varphi$ becomes valid. Indeed, $\mathsf{O}\varphi$ still implies $\mathsf{UO}\varphi$ and, thanks to the negative condition, $\mathsf{OO}\varphi$ implies $\neg\mathsf{UO}\varphi$.

4.2 The Forrester's paradox and weak models

We conclude this section by considering the well-known Forrester's gentle murderer paradox [12]. Let us assume the following facts: (1) it is obligatory that you do not kill; (2) if you kill you ought to kill gently, and (3) it is necessarily the case that if you kill gently then you kill. Let us assume that the action of killing is captured by the atom k, while the action of killing gently is captured by the atom kg.

A first formalisation may be obtained in the following way: fact 1 is expressed by the formula $\mathsf{O}\neg k$; fact 2 and is expressed by the formula $k \rightarrow \mathsf{O}kg$; while fact 3 is expressed by the formula $\mathsf{U}(kg \rightarrow k)$.

As emphasized in the introduction, under the assumption that you kill, fact 1, fact 2 and fact 3 are together inconsistent in SDL (expanded with a necessity operator U), i.e., $k \wedge \mathsf{O}\neg k \wedge (k \rightarrow \mathsf{O}kg))$ is an inconsistent SDL formula. The problem is that in SDL from k and $k \rightarrow \mathsf{O}kg$ we can infer $\mathsf{O}kg$ which, in combination with $\mathsf{U}(kg \rightarrow k)$, implies $\mathsf{O}k$. Since the SDL obligation operator O satisfies Axiom D, it is not possible to have $\mathsf{O}k$ and $\mathsf{O}\neg k$.

This formalisation is not meaningful in our preference model, since in these models all logically possible words have to be taken into account. Therefore, the formula $\mathsf{U}(kg \rightarrow k)$ (which would exclude from a model all worlds satisfying $kg \wedge \neg k$) is unsatisfiable in the class of preference models.

To capture the Forrester paradox, we need to use weak preference models. The formula $k \wedge \mathsf{O}\neg k \wedge (k \rightarrow \mathsf{O}kg) \wedge \mathsf{U}(kg \rightarrow k)$ is satisfiable for the class of weak preference models. Indeed, there exists a weak preference model which concomitantly satisfies the three formulas $\mathsf{O}\neg k$, $\mathsf{O}kg$ and $\mathsf{U}(kg \rightarrow k)$. The following is an example of such a weak preference model: $W = \{w_1, w_2, w_3\}$, $w_1 = \{kg, k\}$, $w_2 = \{k\}$ and $w_3 = \emptyset$ with $w_2 \prec w_1$, $w_2 \prec w_3$ and $w_1 \approx w_3$.

An alternative solution to the Forrester's gentle murderer paradox, widely explored in the literature (see, e.g., [41]) consists in reformulating condition (2) above as the conditional obligation of killing gently under the condition of killing. Specifically, we again represent fact 1 by the formula $\mathsf{O}\neg k$ and fact 3 as $\mathsf{U}(kg \rightarrow k)$, while we now represent fact 2 by the formula $\mathsf{O}(k|kg)$. Let us also assume that k is true. It is easy to verify that the formula $\mathsf{O}\neg k \wedge \mathsf{O}(k|kg) \wedge k \wedge \mathsf{U}(kg \rightarrow k)$ is satisfiable in the class of weak preference models.

5 Computational Complexity

In this section, we prove that the satisfiability problems for the fragments of $\mathcal{L}_{\mathrm{CPDL}}(Atm)$ and $\mathcal{L}_{\mathrm{CPDL}^+}(Atm)$ in which obligations and permissions are only about propositional facts are P-complete. In order to do that, we provide a translation from these languages to the language proposed in [19], namely the language of the ceteris paribus logic CP. This was designed and developed with the purpose of embedding other formal logics, such as for instance the atemporal version of the logic of "seeing to it that" STIT [3] and dynamic logic of propositional assignments DL − PA [25], and highlighting common relationships and structures employed in different formalisms.

In order to provide a formal translation, we borrow some useful notations from [19] to define the CP logic and the corresponding language $\mathcal{L}_{\mathrm{CP}}$.

Given the finite set of atomic propositions Atm, the language $\mathcal{L}_{\mathrm{CP}}$ is such that:

$$\mathcal{L}_{\mathrm{CP}}(Atm) : \varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid \langle X\rangle\varphi$$

where $p \in P$ and $X \subseteq Atm$. Formula $\langle X\rangle\varphi$ has to be read "φ is possible in some state which is X-equivalent to the current one (or also *all things expressed in X being equal*)". The dual of the operator $\langle X\rangle$ is defined in the usal way as follows:

$$[X]\varphi =_{def} \neg\langle X\rangle\neg\varphi.$$

The language $\mathcal{L}_{\mathrm{CP}}$ is interpreted relative to a *simple model* $W \subseteq 2^{Atm}$ and a point $v \in W$ as follows. (We omit interpretation of atomic propositions and boolean operators since they are defined in the usual way.)

$$W, v \models \langle X\rangle\varphi \Longleftrightarrow \exists u \in W \text{ such that } v \equiv_X u \text{ and } W, u \models \varphi,$$

where $\equiv_X$ is the notion of circumstantial indistinguishability defined in Definition 3.

5.1 Translating Unconditional Obligations and Permissions

In this section, we provide a polynomial satisfiability preserving translation from the fragment of CPDL in which deontic operators can only be about propositional formulas to the CP logic. The idea of the translation is to exploit special atomic formulas for utility of type u_k which allow us to simulate the preference ordering of the original preference model. In particular, the meaning of u_k is that the actual world has a utility value equal to k where k ranges over $Util = \{0, \ldots, n\}$. Since the set of atomic propositions Atm is finite, a weak preference model includes at most $|2^{Atm}|$ elements while a preference model includes exactly $|2^{Atm}|$ elements. Therefore, it suffices to suppose that $n = |2^{Atm}|$. Our translation $tr : \mathcal{L}_{\mathrm{CPDL}}(Atm) \to \mathcal{L}_{\mathrm{CP}}\big(Atm \cup \{u_k : k \in Util\}\big)$ is defined as follows:

$$
\begin{aligned}
tr(p) &= p \text{ for } p \in Atm,\\
tr(\neg\varphi) &= \neg tr(\varphi),\\
tr(\varphi \wedge \psi) &= tr(\varphi) \wedge tr(\psi),\\
tr(\mathsf{O}\varphi) &= \langle\emptyset\rangle tr(\varphi) \wedge\\
&\qquad [\emptyset] \bigwedge_{k \in Util} \Big(\big(u_k \wedge tr(\neg\varphi)\big) \to\\
&\qquad [Atm \setminus Atm(\varphi)]\big(tr(\varphi) \to u_{>k}\big)\Big),\\
tr(\mathsf{P}\varphi) &= \langle\emptyset\rangle tr(\varphi) \wedge\\
&\qquad [\emptyset] \bigwedge_{k \in Util} \Big(\big(u_k \wedge tr(\neg\varphi)\big) \to\\
&\qquad [Atm \setminus Atm(\varphi)]\big(tr(\varphi) \to u_{\geq k}\big)\Big),\\
tr(\mathsf{U}\varphi) &= [\emptyset] tr(\varphi),
\end{aligned}
$$

with

$$
\begin{aligned}
u_{>k} &=_{def} \bigvee_{k' \in Util : k' > k} u_{k'}, \text{ and}\\
u_{\geq k} &=_{def} \bigvee_{k' \in Util : k' \geq k} u_{k'}.
\end{aligned}
$$

Note that when the formula in the scope of a deontic operator for obligation or permission is propositional, the previous translation is polynomial in the size of the input formula.

As the following theorem indicates, the previous translation provides an embedding of satisfiability checking for the language $\mathcal{L}_{\mathrm{CPDL}}$ into satisfiability checking for the language $\mathcal{L}_{\mathrm{CP}}$.

Theorem 1. *Let $\varphi \in \mathcal{L}_{\mathrm{CPDL}}(Atm)$. Then,*

- *φ is satisfiable relative to weak preference models if and only if $[\emptyset](\chi_0 \wedge \chi_1) \wedge tr(\varphi)$ is satisfiable relative to simple models, and*

– *φ is satisfiable relative to preference models if and only if $[\emptyset](\chi_0 \wedge \chi_1 \wedge \chi_2) \wedge tr(\varphi)$ is satisfiable relative to simple models,*

where

$$\chi_0 =_{def} \bigwedge_{X \subseteq Atm, k \in Util} \Big(\langle\emptyset\rangle \big(u_k \wedge con(X)\big) \rightarrow [\emptyset]\big(con(X) \rightarrow u_k\big) \Big),$$

$$\chi_1 =_{def} \big(\bigvee_{k \in Util} u_k \wedge \bigwedge_{k,k' \in Util : k \neq k'} (u_k \rightarrow \neg u_{k'}) \big),$$

$$\chi_2 =_{def} \bigwedge_{X \subseteq Atm} \langle\emptyset\rangle con(X),$$

with

$$con(X) =_{def} \bigwedge_{p \in X} p \wedge \bigwedge_{p \in Atm \setminus X} \neg p.$$

Proof. We only prove the first item, as the proof of the second item is analogous.

($\Rightarrow$) Let W be a simple model and $v \in W$ such that $W, v \models [\emptyset](\chi_0 \wedge \chi_1)$. We build the weak preference model $M = (W', \preceq)$ where W' is defined as follows:

$$W' = \big\{ w \setminus \{u_k : k \in Util\} : w \in W \big\}.$$

The fact $W, v \models [\emptyset](\chi_0 \wedge \chi_1)$ guarantees that we can build a bijection f from W to W' such that, for every $w \in W$, $\big(w \cap Atm\big) = \big(f(w) \cap Atm\big)$. For notational convenience, for each $w \in W$, we denote $f(w)$ by s_w. We moreover define $\preceq$ in M as follows:

$\forall s_w, s_{w'} \in W' : s_w \preceq s_{w'}$ iff $\exists k, k' \in Util$ such that $k \leq k', u_k \in w$ and $u_{k'} \in w'$.

By induction on the structure of φ, we can prove that $W, v \models tr(\varphi)$ iff $M, s_v \models \varphi$. The atomic case and the boolean cases are straightforward. We just prove the case $\varphi = \neg\psi$ as an example. $M, s_v \models \neg\psi$ is equivalent to $M, s_v \not\models \psi$. By induction hypothesis, the latter is equivalent to $W, v \not\models tr(\psi)$. The latter is equivalent to $W, v \models \neg tr(\psi)$ which, by the definition of the translation tr, is the same as $W, v \models tr(\neg\psi)$.

Let us prove the modal case $\varphi = \mathsf{O}\psi$.

$$
\begin{aligned}
M, s_v \models \mathsf{O}\psi \iff & \exists s_z \in W' \text{ s.t. } M, s_z \models \psi \text{ and} \\
& \forall s_w, s_u \in W' : \text{ if } M, s_w \models \psi \text{ and } s_w \preceq_{Atm \setminus Atm(\psi)} s_u \text{ then } M, s_u \models \psi, \\
\iff & \exists z \in W \text{ s.t. } W, z \models tr(\psi) \text{ and} \\
& \forall s_w, s_u \in W' : \text{ if } M, s_u \models \neg\psi, M, s_w \models \psi \text{ and } s_w \equiv_{Atm \setminus Atm(\psi)} s_u \\
& \text{then } s_u \prec s_w, \\
\iff & \exists z \in W \text{ s.t. } W, z \models tr(\psi) \text{ and} \\
& \forall w, u \in W : \text{ if } W, u \models tr(\neg\psi), W, w \models tr(\psi) \text{ and} \\
& w \equiv_{Atm \setminus Atm(\psi)} u \text{ then } s_u \prec s_w \text{ (by induction hypothesis)}, \\
\iff & \exists z \in W \text{ s.t. } W, z \models tr(\psi) \text{ and} \\
& \forall w, u \in W : \text{ if } W, u \models tr(\neg\psi), W, w \models tr(\psi) \text{ and} \\
& w \equiv_{Atm \setminus Atm(\psi)} u \text{ then } \exists k, k' \in \mathit{Util} \text{ such that } k < k', u_k \in u \\
& \text{ and } u_{k'} \in w \text{ (by definition of } \preceq), \\
\iff & \exists z \in W \text{ s.t. } W, z \models tr(\psi) \text{ and} \\
& \forall w, u \in W, \forall k \in \mathit{Util} : \text{ if } W, u \models tr(\neg\psi), u_k \in u, W, w \models tr(\psi) \\
& \text{and } w \equiv_{Atm \setminus Atm(\psi)} u \text{ then } \exists k' \in \mathit{Util} \text{ such that } k < k' \text{ and} \\
& u_{k'} \in w \text{ (by the fact that } W, v \models [\emptyset]\chi_1), \\
\iff & W, v \models \langle\emptyset\rangle tr(\psi) \text{ and} \\
& W, v \models [\emptyset] \bigwedge_{k \in Util} \big((u_k \wedge tr(\neg\psi)) \rightarrow [Atm \setminus Atm(\varphi)](tr(\psi) \rightarrow u_{>k})\big), \\
\iff & W, v \models tr\big(\mathsf{O}\psi\big).
\end{aligned}
$$

The case $\varphi = \mathsf{P}\psi$ is analogous, while the case $\varphi = \mathsf{U}\psi$ is straightforward. Therefore, we do not need to prove them explicitly.

($\Leftarrow$) Let $M = (W, \preceq)$ be a weak preference model and $v \in W$. We build the following utility ranking inductively:

$$
\begin{aligned}
& U^0 = \{w \in W : \forall u \in W, w \preceq u\}, \\
& U^{k+1} = \Big\{w \in W \setminus \big(\bigcup_{h \leq k} U^h\big) : \forall u \in W \setminus \big(\bigcup_{h \leq k} U^h\big), w \preceq u\Big\}.
\end{aligned}
$$

U^0 is the set of worlds of 0-utility, while U^{k+1} is the set of worlds with $k+1$-utility. We build the simple model W' as follows:

$$W' = \big\{w \cup \{u_k : w \in U^k\} : w \in W\big\}.$$

It is easy to check that $W', s_v \models [\emptyset]\big(\bigvee_{k \in Util} u_k \wedge \bigwedge_{k,k' \in Util : k \neq k'} (u_k \rightarrow \neg u_{k'})\big)$. Moreover, by induction on the structure of φ, we can prove that $W', s_v \models tr(\varphi)$ iff $M, v \models \varphi$. □

The following is a direct corollary of (i) the previous Theorem 1, (ii) the fact that the translation tr applied to the fragment $\mathcal{L}_{\mathrm{CPDL-Prop}}(Atm)$ is polynomial,

(iii) the fact that satisfiability checking for the language $\mathcal{L}_{\mathrm{CP}}$ with finitely many atomic propositions is in PTIME. Item (iii) is a consequence of the fact that, as shown in [19, Section 2.3], there exists a polynomial satisfiability preserving translation of the finite-variable CP logic into the finite-variable modal logic S5. It is known that satisfiability checking for the latter can be done in polynomial time [20].

Corollary 1. *Let $\varphi \in \mathcal{L}_{\mathrm{CPDL-Prop}}(Atm)$. Then, checking satisfiability of φ relative to the class of preference models (resp. weak preference models) can be done in polynomial time.*

5.2 Translating Conditional Obligations and Permissions

All the observations done in the previous section are valid also for the case of conditional obligations and permissions. Notice that a condition simply enforces a preference relation on a smaller set of state-of-affairs. This means that the class of equivalence is smaller in the sense that the states that belong to it are all of those that have the partial assignment specified in the condition. In order to capture this idea, we enrich the translation with the following rules:

$$\begin{aligned} tr\big(\mathsf{O}(\psi|\varphi)\big) = {} & \langle\emptyset\rangle tr(\varphi \wedge \psi) \wedge \\ & [\emptyset]\Big(tr(\psi) \rightarrow \bigwedge_{k \in Util} \Big(\big(u_k \wedge tr(\neg\varphi)\big) \rightarrow \\ & [Atm \setminus Atm(\varphi)]\big((tr(\psi) \wedge tr(\varphi)) \rightarrow u_{>k}\big)\Big)\Big), \\ tr\big(\mathsf{P}(\psi|\varphi)\big) = {} & \langle\emptyset\rangle tr(\varphi \wedge \psi) \wedge \\ & [\emptyset]\Big(tr(\psi) \rightarrow \bigwedge_{k \in Util} \Big(\big(u_k \wedge tr(\neg\varphi)\big) \rightarrow \\ & [Atm \setminus Atm(\varphi)]\big((tr(\psi) \wedge tr(\varphi)) \rightarrow u_{\geq k}\big)\Big)\Big). \end{aligned}$$

The following theorem is a straightforward generalization of Theorem 1.

Theorem 2. *Let $\varphi \in \mathcal{L}_{\mathrm{CPDL}^+}(Atm)$. Then,*

- *φ is satisfiable relative to weak preference models if and only if $[\emptyset]\chi_1 \wedge tr(\varphi)$ is satisfiable relative to simple models, and*
- *φ is satisfiable relative to preference models if and only if $[\emptyset](\chi_1 \wedge \chi_2) \wedge tr(\varphi)$ is satisfiable relative to simple models,*

where

$$\begin{aligned} \chi_1 &=_{def} \Big(\bigvee_{k \in Util} u_k \wedge \bigwedge_{k,k' \in Util : k \neq k'} (u_k \rightarrow \neg u_{k'})\Big), \\ \chi_2 &=_{def} \bigwedge_{X \subseteq Atm} \langle\emptyset\rangle\Big(\bigwedge_{p \in X} p \wedge \bigwedge_{p \in Atm \setminus X} \neg p\Big). \end{aligned}$$

The following is a direct corollary of Theorem 2 and is proved in the same way as Corollary 1.

Corollary 2. *Let $\varphi \in \mathcal{L}_{\mathrm{CPDL}^+-\mathrm{Prop}}(Atm)$. Then, checking satisfiability of φ relative to the class of preference models (resp. weak preference models) can be done in polynomial time.*

6 From Syntax Dependence to Independence

The general idea behind our ceteris paribus notion of obligation is that φ is obligatory if and only if the utility of a world increases in the direction by the formula φ ceteris paribus, i.e., "all else being equal". Following Von Wright (see also [46]), in CPDL we capture this ceteris paribus aspect, by keeping fixed the truth values of the atoms not occurring in φ (i.e., $Atm \setminus Atm(\varphi)$). The fact that the sets of atoms not occurring in two logical equivalent formulas do not necessarily coincide explains why the obligation and permission operators of CPDL are not closed under logical equivalence.

A natural way to obtain obligation and permission operators which are closed under logical equivalence consists in defining the ceteris paribus condition by keeping fixed the truth values of the atoms with respect to which φ is independent (i.e., the atoms which do not affect the truth value of φ). This connects to Rescher's idea that the concept of ceteris paribus should be defined in terms of a concept of independence between formulas [42] (see also [16]). In formal terms, let φ be a propositional formula. Then:

$$
\begin{aligned}
M, w \models \mathsf{O}^i\varphi \iff & \exists v \in W \text{ such that } M, v \models \varphi, \text{ and} \\
& \forall v, u \in W : \text{ if } M, v \models \varphi \text{ and } v \preceq_{Indep(\varphi)} u \text{ then } M, u \models \varphi \\
M, w \models \mathsf{P}^i\varphi \iff & \exists v \in W \text{ such that } M, v \models \varphi, \text{ and} \\
& \forall v, u \in W : \text{ if } M, v \models \varphi \text{ and } v \prec_{Indep(\varphi)} u \text{ then } M, u \models \varphi
\end{aligned}
$$

where $Indep(\varphi) = \{p \in Atm : \forall w \in W, w \cup \{p\} \models \varphi \text{ iff } w \setminus \{p\} \models \varphi\}$ denotes the set of atoms with respect to which φ is independent and $w \models \varphi$ means that the valuation w satisfies the propositional formula φ. We use the notation O^i for independence-based "ceteris paribus" obligation and P^i for independence-based "ceteris paribus" permission. Notice that $Atm \setminus Atm(\varphi) \subseteq Indep(\varphi)$. Thus, a ceteris paribus obligation/permission defined in terms of $Atm \setminus Atm(\varphi)$ implies a ceteris paribus obligation/permission defined in terms of $Indep(\varphi)$, as if two worlds are equivalent with regard to $Indep(\varphi)$ then they are also equivalent with regard to $Atm \setminus Atm(\varphi)$.

The translation given in Section 5.1 can be easily adapted to this new semantics based on the concept of semantic independence. It is sufficient to replace all occurrences of $Atm \setminus Atm(\varphi)$ by $Indep(\varphi)$ in the translation rules for the deontic operators. Therefore, for every formula of the deontic language with the independence-based obligation operator O^i and permission operator P^i, we can find a logically equivalent formula of the ceteris paribus language $\mathcal{L}_{\mathrm{CP}}$.

Note that when φ is propositional computing $Indep(\varphi)$ can be done in polynomial time. Indeed, all atoms in $Atm \setminus Atm(\varphi)$ belong to $Indep(\varphi)$. We simply need to enumerate the atoms in $Atm(\varphi)$ and, for each of them, to verify whether φ is independent with respect to it. The latter problem is reducible to finite-variable SAT which is in PTIME. Therefore, satisfiability checking relative to preference models for the deontic language in which formulas in the scope of deontic operators O^i and P^i are propositional remains polynomial.

The reason why the previous notions of obligation and permission are closed under logical equivalence is that two logical equivalent formulas are independent with respect to the same set of atomic propositions. Moreover, they have the same truth values at all worlds of a preference model. This feature is captured by the following two validities:

$$\models_{\mathcal{P}} \mathsf{U}(\varphi \leftrightarrow \psi) \rightarrow (\mathsf{O}^i\varphi \rightarrow \mathsf{O}^i\psi)$$
$$\models_{\mathcal{P}} \mathsf{U}(\varphi \leftrightarrow \psi) \rightarrow (\mathsf{P}^i\varphi \rightarrow \mathsf{P}^i\psi)$$

This means that if φ and ψ are universally equivalent, then having the obligation (resp. permission) that φ is the same as having the obligation (resp. permission) that ψ. Since logical equivalence (i.e., equivalence relative to all preference models) is stronger than universal equivalence (i.e., equivalence relative to a specific preference model), we also have the following properties:

$$\models_{\mathcal{P}} \varphi \leftrightarrow \psi \text{ then } \models_{\mathcal{P}} (\mathsf{O}^i\varphi \rightarrow \mathsf{O}^i\psi)$$
$$\models_{\mathcal{P}} \varphi \leftrightarrow \psi \text{ then } \models_{\mathcal{P}} (\mathsf{P}^i\varphi \rightarrow \mathsf{P}^i\psi)$$

7 From Deontic Logic to CP-nets

In this section we shall show how the fragment of the logic for complete preference models introduced in Section 3 has an equivalent representation based on CP-nets. Specifically, this is the fragment in which the content of a deontic operator is a literal. In the case of conditional deontic operators the antecedent can be a conjunction of literals.

7.1 CP-nets

CP-nets [6] are a compact representation of conditional preferences over ceteris paribus semantics.

Definition 10. *A CP-net over a set of binary variables V is a tuple $\mathcal{N} = (G, CPT)$, where $G = (V, E)$ is a directed graph and $CPT = \{CPT(V_i) | V_i \in V\}$ is a set of conditional preference tables (CP-tables). An edge $(V_i, V_j) \in E$ represents that preferences over $Dom(V_i)$ depend on the value of V_j.*

For each variable $V_i \in V$, given the assignment to its parents, a CP-table $CPT(V_i)$ represents the preference order over the values of the domain of V_i. For instance, $CPT(A) = \{a \prec \bar{a}\}$ represents the strict preference over the values

of a variable A, i.e., $\bar{a}$ is more preferred than a. Each preference order in a CP-table is also called a CP-statement. A CP-net induces a preference graph over all the possible outcomes: each node corresponds to an outcome, that is, a complete assignment of values to variables. Moreover, a directed edge between a pair of outcomes (o_j, o_i), which differ only in the value of one variable, means that $o_j \preceq o_i$. A *worsening flip* is a change in the value of a variable to a less preferred value according to the CP-statement for that variable. A more recent extension, namely CP-net with indifference [1], takes into account indifference and models lack of information using incomparability. The qualitative compact representation of ceteris paribus scenario and the algorithms developed for inference, make CP-net an interesting and useful tool to represent our model.

7.2 CP-net Representation

Let us restrict our analysis to conditional obligations and permissions whose antecedent is a consistent conjunction of literals $l_1 \wedge \ldots \wedge l_n$ and whose consequent is a literal l such that $Atm(l_1 \wedge \ldots \wedge l_n) \cap Atm(l) = \emptyset$. We note $\mathcal{L}^{Frag}_{\text{CPDL}^+}$ such a fragment of the language $\mathcal{L}_{\text{CPDL}^+}$. We can show that the induced preference model can be represented compactly by a CP-net with indifference [1].

Proposition 6. *Let C be a set of obligations and permissions from the language $\mathcal{L}^{Frag}_{\text{CPDL}^+}$. Let $M = (W, \preceq)$ be the minimal preference model induced by C and $\mathcal{N} = (G, P)$ be the CP-net induced by C. Then, M and $\mathcal{N}$ are isomorphic.*

Proof. The minimal preference model induced by C is the one which satisfies C and has the less restrictive constraints. First, for each permission in C, introduce a weak order among worlds that differ only for the consequent of the permission, ceteris paribus the antecedent of the permission. For each obligation, introduce a strict order over the worlds that differ only for the consequent of the obligation ceteris paribus the antecedent of the obligation. For all the worlds that are not explicitly compared we introduced a weak order among them. The induced CP-net is built as follows: to each atom $v_i \in Atm$ there is a corresponding variable $V_i \in V$ such that $Dom(V_i) = \{v_i, \bar{v_i}\}$.

Each conditional obligation $O(v_i|v_j) \in N$ introduces a directed edge (V_i, V_j) in the dependency graph G, such that V_i becomes a parent of V_j. It induces a strict order over $Dom(V_j)$ given the assignment of V_i such that $CPT(V_j) = \{v_i : \bar{v_j} \prec v_j, \bar{v_i} : v_j \prec \bar{v_j}\}$. Similarly, each conditional concession $P(v_i|v_j) \in N$ induces a weak order over $Dom(V_j)$ such that $CPT(V_j) = \{v_i : \bar{v_j} \preceq v_j, \bar{v_i} : v_j \preceq \bar{v_j}\}$. Notice that in our model, as well as in SDL, everything that is not explicitly forbidden is permitted, i.e., in general $\neg O(v_i) \rightarrow P(\bar{v_i})$. Thus, a variable with indifference over its domain is introduced for each atom that is not explicitly a consequent of any obligation/permission. To show the isomorphism, consider the bijection between the set of worlds W and the set of all the outcomes in the partial orders. We can show that there is a bijection between edges of the partial order and the ordering relations among worlds in preference model. From the subset of worlds that satisfy all the obligations we can move to less preferred

worlds by changing one literal at a time until we visit all the possible worlds. This corresponds to visit all the outcomes in the partial order starting from the subset of optimal outcomes using the definition of worsening flip of CP-net. □

Example 3 (Running example). Let us introduce a running example concerning the presence of cats, dogs, and fences in beach houses (developing the example from [41]). The set of atoms is $Atm = \{c, d, f\}$, where: c represents whether there is a cat; d represents whether there is a dog and f represents whether there is a fence. Mary is the mayor of the town. For safety reasons, she has ordered that there should be fences when there are dogs, and that, on the contrary, there should be no fences when there are no dogs. Cats are allowed with no restrictions. It is easy to check that the following preferences verify $\mathsf{P}(c)$, $\mathsf{O}(d|f)$ and $\mathsf{O}(\neg d|\neg f)$: $w_{\{c,d,f\}} \approx w_{\{d,f\}}, w_{\{c,d\}} \approx w_{\{d\}}, w_{\{c,f\}} \approx w_{\{f\}}, w_{\{\}} \approx w_{\{c\}}, w_{\{c,d\}} \prec w_{\{c,d,f\}}, w_{\{d\}} \prec w_{\{d,f\}}, w_{\{c,f\}} \prec w_{\{c\}}, w_{\{f\}} \prec w_{\{\}}$

Moreover, these preferences are compatible with $\mathsf{O}\neg d$, since this obligation only concerns the ceteris paribus preference for worlds that do not have dogs over worlds that have them. Following Proposition 6, Mary's obligations and concessions can be represented using the CP-net with indifference depicted in Figure 1a, which compactly represents the partial order depicted in Figure 1b. For the sake of readability, we group into the same nodes some worlds of the preference model. Worlds in the same node are indifferent, this is due to the indifference over the values of variable C.

Following Proposition 6, to each atom there is a corresponding variable, thus we have $V = \{C, D, F\}$ representing respectively whether there is cat, a dog and a fence, $Dom(C) = \{c, \bar{c}\}$, $Dom(D) = \{d, \bar{d}\}$ and $Dom(F) = \{f, \bar{f}\}$. Due to obligations and permission, variables C, D are independent while variable F depends on D. Moreover, obligations define the strict orders over $Dom(D), Dom(F)$. From Proposition 1, the unconditional permission $P(c)$ is defined as $\mathsf{P}(\top|c)$ and introduces indifference over $Dom(C)$.

8 Conclusion and perspectives

We have presented a new approach to deontic logic, based on ceteris paribus preferences, which provides a fresh foundation to the logical analysis of deontic concepts. We have introduced the idea of ceteris paribus preferences and on this basis we have built the semantics of a deontic logic, named CPDL (ceteris paribus deontic logic). We have shown that CPDL not only avoids some deontic paradoxes, but also provides an adequate conceptualisation of obligations and permission, conditioned and unconditioned. In particular, CPDL supports formal models of obligations and permission that match common-sense intuitions and legal language.

We have also examined some properties of the resulting logical system showing in particular how it supports for limited aggregation of conjunctions and factual detachment. We have provided variants of our logic that enable for detachment and for closure under logical equivalence.

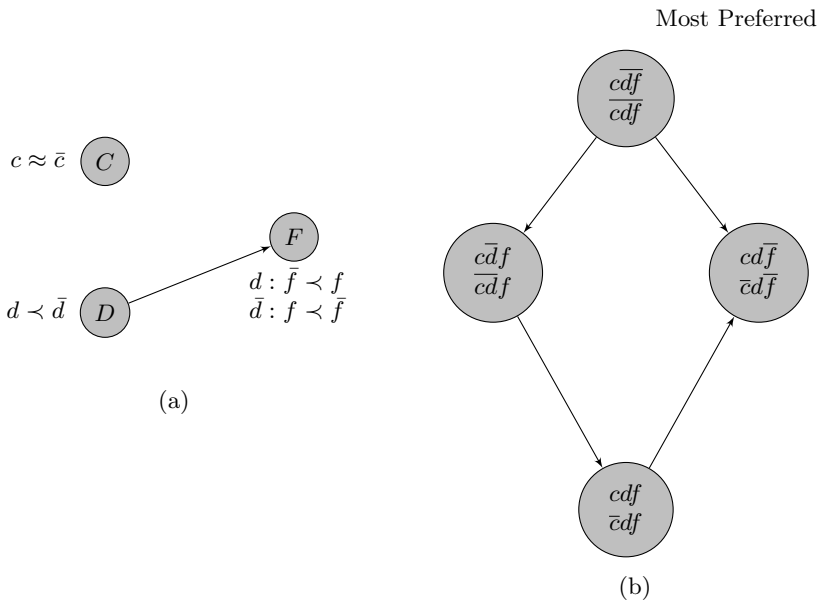


Fig. 1: (a) The CP-net with indifference which represents obligations and permissions in Example 3. (b) The partial order induced by the CP-net in Figure 1a.

Finally, we have established a connection with the CP-nets, which provide for a compact representation and efficient reasoning over sets of ceteris paribus obligations and permissions. We are currently working to develop the framework of CPDL in various directions, such as the integration of our logic with logics of time and actions, and further exploring its translation into different kinds of CP-nets.

We believe that the complexity results presented in Sections 5 and 6 set the basis for an implementation of our deontic logic. Specifically, given the NP-completeness for the satisfiability checking of our logic, we expect to find a reduction of the latter to SAT. This opens up the possibility of using existing SAT solvers for verifying deontic properties and automating deontic reasoning in our setting.

References

1. Allen, T.E.: CP-nets with indifference. In: 2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton). pp. 1488–1495. IEEE (2013)
2. Åqvist, L.: Deontic logic. In: Handbook of philosophical logic, pp. 605–714. Springer (1984)

3. Balbiani, P., Herzig, A., Troquard, N.: Alternative axiomatics and complexity of deliberative STIT theories. Journal of Philosophical Logic **37**(4), 387–406 (2008)
4. van Benthem, J., Grossi, D., Liu, F.: On the two faces of deontics: Semantic betterness and syntactic priority. Tech. rep., Institute for Logic, Language and Computation, University of Amsterdam (2011)
5. Bienvenu, M., Lang, J., Wilson, N.: From preference logics to preference languages, and back. In: Twelfth International Conference on the Principles of Knowledge Representation and Reasoning (2010)
6. Boutilier, C., Brafman, R.I., Hoos, H.H., Poole, D.: Reasoning with conditional ceteris paribus preference statements. In: Proc. of the 15th UAI. pp. 71–80 (1999)
7. Carmo, J., Jones, A.J.: Deontic logic and contrary-to-duties. In: Handbook of philosophical logic, pp. 265–343. Springer (2002)
8. Chisholm, R.M.: Contrary-to-duty imperatives and deontic logic. Analysis **24**(2), 33–36 (1963)
9. Cholvy, L., Garion, C.: An attempt to adapt a logic of conditional preferences for reasoning with contrary-to-duties. Fundamenta Informaticae **48**(2-3), 183–204 (2001)
10. Cornelio, C., Donini, M., Loreggia, A., Pini, M.S., Rossi, F.: Voting with random classifiers (VORACE): theoretical and experimental analysis. Autonomous Agents and Multi-Agent Systems **35**(2), 22 (2021). https://doi.org/10.1007/s10458-021-09504-y
11. Føllesdal, D., Hilpinen, R.: Deontic logic: An introduction. In: Deontic logic: Introductory and systematic readings, pp. 1–35. Springer (1970)
12. Forrester, J.W.: Gentle murder, or the adverbial samaritan. The Journal of Philosophy **81**(4), 193–197 (1984)
13. Gabbay, D., Horty, J., Parent, X., van der Meyden, R., van der Torre, L.: Handbook of deontic logic and normative systems (2013)
14. Garion, C.: Apports de la logique mathématique en ingénierie des exigences. Ph.D. thesis (2002)
15. Garion, C., Van Der Torre, L.: Design by contract deontic design language for multiagent systems. In: International Conference on Autonomous Agents and Multiagent Systems. pp. 170–182. Springer (2005)
16. Girard, P.: Von wright's preference logic reconsidered (2006)
17. Grandi, U., Loreggia, A., Rossi, F., Saraswat, V.: From sentiment analysis to preference aggregation. In: International Symposium on Artificial Intelligence and Mathematics, ISAIM 2014 (2014)
18. Grandi, U., Loreggia, A., Rossi, F., Saraswat, V.: A borda count for collective sentiment analysis. Annals of Mathematics and Artificial Intelligence **77**(3-4), 281–302 (2016)
19. Grossi, D., Lorini, E., Schwarzentruber, F.: The ceteris paribus structure of logics of game forms. Journal of Artificial Intelligence Research **53**, 91–126 (2015)
20. Halpern, J.Y.: The effect of bounding the number of primitive propositions and the depth of nesting on the complexity of modal logic. Artificial Intelligence **75**(2), 361–372 (1995)
21. Hansen, J.: Is there a logic of imperatives. Deontic Logic in Computer Science, Twentieth European summer school in Logic, Language and Information, Germany (2008)
22. Hansson, B.: An analysis of some deontic logics. In: Deontic Logic: Introductory and Systematic Readings, pp. 121–147. Springer (1970)
23. Hansson, S.O.: Preference-based deontic logic (pdl). Journal of Philosophical Logic **19**(1), 75–93 (1990)

24. Hansson, S.O.: The structure of values and norms. Cambridge University Press (2001)
25. Herzig, A., Lorini, E., Moisan, F., Troquard, N.: A dynamic logic of normative systems . In: Walsh, T. (ed.) Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI'11). pp. 228–233. Morgan Kaufmann Publishers (2011)
26. Hilpinen, R.: Deontic logic. In: Goble, L. (ed.) The Blackwell Guide to Philosophical Logic, chap. 8, pp. 159–82. Blackwell (2001)
27. Hilpinen, R.: Deontic logic: Introductory and systematic readings, vol. 33. Springer Science & Business Media (2012)
28. Hilpinen, R., McNamara, P.: Deontic logic: A historical survey and introduction. Handbook of deontic logic and normative systems **1**, 3–136 (2013)
29. Jones, A.J.: Deontic logic and legal knowledge representation. Ratio Juris **3**(2), 237–244 (1990)
30. Kanger, S.: New foundations for ethical theory. In: Hilpinen, R. (ed.) Deontic Logic, pp. 36–58. Reidel ([1957]1971)
31. Li, M., Kazimipour, B.: An efficient algorithm to compute distance between lexicographic preference trees. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18. pp. 1898–1904. Stockholm, Sweden (7 2018)
32. Loreggia, A., Mattei, N., Rossi, F., Venable, K.B.: Preferences and ethical principles in decision making. In: AIES 2018 - Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (2018)
33. Loreggia, A., Mattei, N., Rossi, F., Venable, K.B.: Value alignment via tractable preference distance. In: Yampolskiy, R.V. (ed.) Artificial Intelligence Safety and Security, chap. 16. CRC Press (2018)
34. Loreggia, A., Mattei, N., Rossi, F., Venable, K.: Metric learning for value alignment. vol. 2419 (2019)
35. Loreggia, A., Mattei, N., Rossi, F., Venable, K.: Modeling and reasoning with preferences and ethical priorities in AI systems. Oxford University Press (2020)
36. Loreggia, A., Rossi, F., Venable, K.: Modelling ethical theories compactly. vol. WS-17-01 - WS-17-15, pp. 122–126 (2017)
37. Loreggia, A., Lorini, E., Sartor, G.: A ceteris paribus deontic logic. In: 35th Italian Conference on Computational Logic (CILC 2020). vol. 2710, pp. 248–262. CEUR (2020)
38. Loreggia, A., Lorini, E., Sartor, G.: A novel approach for a ceteris paribus deontic logic. In: Proceedings of the EKAW 2020 Posters and Demonstrations Session. CEUR Workshop Proceedings, vol. 2751, pp. 12–16. CEUR-WS.org (2020)
39. Loreggia, A., Mattei, N., Rossi, F., Venable, K.B.: On the distance between CP-nets. In: Proc. of the 17th AAMAS. pp. 955–963 (2018)
40. Loreggia, A., Mattei, N., Rossi, F., Venable, K.B.: CPMetric: Deep siamese networks for metric learning on structured preferences. In: Artificial Intelligence. IJCAI 2019 International Workshops. pp. 217–234. Springer (2020)
41. Prakken, H., Sergot, M.: Dyadic deontic logic and contrary-to-duty obligations. In: Defeasible deontic logic, pp. 223–262. Springer (1997)
42. Rescher, N.: Semantic foundations for the logic of preference. The logic of decision and action pp. 37–62 (1967)
43. Ross, A.: Imperatives and logic. Philosophy of Science **11**(1), 30–46 (1944)
44. Rossi, F., Loreggia, A.: Preferences and ethical priorities: Thinking fast and slow in AI. In: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2019. pp. 3–4 (2019)

45. Sen, A.: Choice, ordering and morality. In: Körner, S. (ed.) Practical Reason. Blackwell, Oxford (1974)
46. Van Benthem, J., Girard, P., Roy, O.: Everything else being equal: A modal logic for ceteris paribus preferences. Journal of philosophical logic **38**(1), 83–125 (2009)
47. Von Wright, G.H.: Norm and Action: A Logical Inquiry. Routledge (1963)
48. Von Wright, G.H.: Deontic logic. Mind **60**(237), 1–15 (1951)
49. Von Wright, G.H.: The logic of preference (1963)
50. Von Wright, G.H.: A new system of deontic logic. In: Deontic Logic: Introductory and Systematic Readings, pp. 105–120. Springer (1970)
51. Von Wright, G.H.: The logic of preference reconsidered (1972)
52. Wang, H., Shao, S., Zhou, X., Wan, C., Bouguettaya, A.: Preference recommendation for personalized search. Knowledge-Based Systems **100**, 124–136 (2016)
53. Weinberger, C., Weinberger, O.: Logik, semantik, hermeneutik (1980)
54. Wellman, M.P., Doyle, J.: Preferential semantics for goals. In: Proceedings of the 9th National Conference on Artificial Intelligence, Anaheim, CA, USA, July 14-19, 1991, Volume 2. pp. 698–703. AAAI Press / The MIT Press (1991)