



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna Archivio istituzionale della ricerca

How to model contrary-to-duty with GCP-nets

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

Published Version:

Loreggia A., Calegari R., Lorini E., Rossi F., Sartor G. (2022). How to model contrary-to-duty with GCP-nets. *INTELLIGENZA ARTIFICIALE*, 16(2), 185-198 [10.3233/IA-221057].

Availability:

This version is available at: <https://hdl.handle.net/11585/912872> since: 2023-05-16

Published:

DOI: <http://doi.org/10.3233/IA-221057>

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)

How to model contrary-to-duty with GCP-nets

Andrea Loreggia^{a,*}, Roberta Calegari^b, Emiliano Lorini^c, Francesca Rossi^d and Giovanni Sartor^{b,e}

^a*University of Brescia - Brescia, Italy*

^b*University of Bologna - Bologna, Italy*

^c*IRIT, CNRS, Toulouse University - France*

^d*IBM Research - Yorktown Heights, NY, USA*

^e*European University Institute - Florence, Italy*

Abstract. Preferences are ubiquitous in our everyday life. They are essential in the decision making process of individuals. Recently, they have also been employed to represent ethical principles, normative systems or guidelines. In this work we focus on a ceteris paribus semantics for deontic logic: a state of affairs where a larger set of respected prescriptions is preferable to a state of affairs where some are violated. Conditional preference networks (CP-nets) are a compact formalism to express and analyse ceteris paribus preferences, with some desirable computational properties. In this paper, we show how deontic concepts (such as contrary-to-duty obligations) can be modeled with generalized CP-nets (GCP-nets) and how to capture the distinction between strong and weak permission in this formalism. To do that, we leverage on an existing restricted deontic logic that will be mapped into conditional preference nets.

Keywords: Deontic logic, GCP-nets, ceteris-paribus semantics, contrary-to-duty, strong and weak permission

1. Introduction

Deontic logic allows us to model legal and ethical prescriptions such as obligations and permissions. Its computational implementation may provide a framework to govern the functioning of AI systems, assessing their behaviour in different situations and scenarios, detecting violations and ensuring compliance with ethical and legal requirements [2].

Modelling deontic notions through preferences [7, 25, 46] has the advantage of linking deontic notions to the rich research on preferences, in multiple disciplines, such as philosophy, mathematics, economics and politics. Preferences are also central in artificial intelligence. In this area, researchers study the computational aspects of classical social choice problems,

preference aggregation and preference reasoning [15, 37, 42]. Applications of preferences can be found in many different contexts: from multi-agent and recommender systems [39, 44]—where preferences are employed to drive the making decision process of agents or to describe the profile of users—to sentiment analysis techniques [21, 22] and machine learning ensemble approaches [9, 10, 52]—where they are adopted to improve the quality of predictions.

In his seminal paper, Makinson [34] discussed the traditional preference based deontic logics and observed that “The definition of a “betterness” relation between possible worlds suggested above, in terms of the set of explicit promulgations of the code that are violated in each world, is perfectly precise; and such a relation permits a semantic account of conditional obligation.”

As already done in literature [7, 27, 28, 38], we model deontic notions through ceteris-paribus pref-

*Corresponding author: Andrea Loreggia, University of Brescia - Brescia, Italy. E-mail: andrea.loreggia@unibs.it.

erences, namely, conditional preferences for a state of affairs over other states of affairs, all the rest being equal. Conditional or rather contextual obligations have been captured in the development of dyadic deontic logic, which captures deontic conditionality through a special conditional operator. The idea of *ceteris-paribus* preferences was originally introduced by the philosopher and logician Georg Henrik Von Wright [49, 50], and after that, a semantics able to represent it was first proposed by Bengt Hansson [23]. In particular, we focus on the *ceteris-paribus* preference for a proposition over its complement. This idea provides the intuition at the basis of CP-nets [5], a formalism which enables representing preferences in a compact way and reasoning about them.

Their application can be effective for the multi-agent pervasive systems of artificial intelligence (AI). Indeed, AI agents are pervasive and they need to be flexible, adaptive, and creative, but at the same time, they have to comply with many ethical and legal requirements [3, 6, 36]. Whether an AI agent is autonomous or collaborating with humans, it should be aware and follow appropriate ethical principles and should thus exhibit properties such as fairness or other virtues [4, 40, 41]. Based on these needs, during the last decade, researchers propose languages to model and reason with norms in multi-agent systems, enabling methodologies and organization models for checking conflicts between the norms at design time [13, 16].

By modelling a normative system of a multi-agent system through a CP-net we obtain a compact representation that shall allow us to check important properties of the modelled normative system, like for instance consistency. Moreover, preferences of agents can also be modelled through CP-nets [35]. This would enable to check agents' preferences against the exogenous preferences expressed by various normative systems, such as laws, ethical principles, feasibility constraints, or safety regulations [29, 31]. In this regard, recently, new approaches define metric spaces over structured preferences [26, 30], providing feasible ways for preferences' comparison. Whilst modeling deontic concepts with *ceteris paribus* preferences has become a standard approach (see for instance, [24, 45, 46]), we found just one first attempt to model deontic notions with CP-nets in [17]. Unfortunately, this work did not recognise CP-nets limitations (in their original formalization). Indeed, they do not allow for the representation of incomplete preference orders over the values of one or more variables. This limitation can be removed by

adopting a recent model generalization called GCP-net (aka generalized CP-net) [18]. To the best of our knowledge, no one else investigated the relationship between deontic concepts and GCP-nets. We think cooperation among researchers from deontic logic and GCP-nets can be very beneficial, and possibly lead to a new workable approach to modelling norms and reasoning about them.

We assume that norms establishing obligations can be viewed as expressing social preferences for situations in which the obligation has complied with other situations in which it is violated. Similarly, we assume that norms establishing liberties (bilateral permissions, to do and not to do) express indifference over complementary situations. In this regard, we discuss and address two topics in deontic logic, namely contrary-to-duty obligations [7] and the distinction between strong and weak permissions.

We show that a consistent GCP-net can be built that captures both an obligation and the prescriptions specifying what should be done in case the obligation is violated. The latter prescriptions may be incompatible with prescriptions on what should be done in case of compliance [8]. We illustrate contrary-to-duty obligations with a well-known example discussed in the literature. However, realistic examples of contrary-to-duty obligations can be found in various domains, such as in commercial contracts, where repair obligations can be established for delay or non-fulfilment: to inform, compensate the damage, replace defective goods, pay penalties, etc. (for a logical model see [20]). It can also be found in civil liability, where harmful behaviour is met with obligations to repair, mitigate, and pay punitive damages in case of persistence.

We also show that a GCP-net model can capture concepts of strong and weak permission, a distinction that has been long discussed in deontic logic [46, 48] and even earlier in legal theory, in connection with issues concerning the completeness of legal systems [53]. We indeed distinguish the case in which an action is positively permitted, through a permissive norm, and the case in which the action is rather unregulated, as no norm, and in particular, no prohibition addresses it [47].

Contribution. We propose to leverage GCP-net, an existing preference framework, to model contrary-to-duty in a compact way. We define different levels of incomparability to differentiate between two situations: one where two outcomes are incomparable because nothing is said about all the values of a

feature, and another, where two or more variables are independent and thus preferences are described over the domains of these variables no matter the assignment of the others. We show that a normative system represented through a restricted deontic logic for contrary-to-duty is isomorphic to preferences represented compactly with GCP-nets.

2. A running example

In this section, we introduce a running example, concerning the presence of cats, dogs, and fences in beach houses (developing the example from [38]). The example is used throughout the paper to explicate new notions when they are introduced.

Example 1. (Running example). Mary is the mayor of Cattown, the city of cats lovers. She knows that her community, with few exceptions, dislikes dogs and likes cats. Cattown people also dislike the sight of fences around houses and generally prefer to be able to move around and visit each other without the obstacle of fences. However, there was a problem with dogs entering other people's property and causing fear and sometimes harm. Moreover, people are fussy about the way in which their town looks, they like that all houses and street furniture are white, as indeed they are. Mary believes that her role consists in defining policies that fit the preferences of her constituency. Thus, she has enacted the following, making them mandatory in Cattown:

1. dogs are forbidden;
2. people may have a cat or not;
3. if there is a dog, then there must be a fence;
4. if there is a fence, then there must be white;
5. if there is no dog, then there must be no fence.

Though most people in her constituency like these policies, some do not. In particular, there are few dog lovers in the city, who do not share the approach to dogs prevailing in Cattown. However, Mary believes that – given the preferences of the majority and the history and peculiar spirit of Cattown – her regulation articulates a social perspective, namely, a view of the community on what it prefers and a view that can justifiably be imposed on all its citizens. If a dog lover has decided to move into Cattown, he should have known what he would find there.

An issue has recently emerged concerning bobcats. It is being debated whether, in the absence of a provision prohibiting or allowing bobcats, John is

allowed to keep a bobcat in his garden. Mary thinks that he is not, he should have asked for it, while John believes that he is. Can their conflict of opinion be explained away as merely concerning a conceptual misunderstanding (on the notion of permission)?

3. Background

To formally express rules such as those enacted by Mary, a restricted deontic language is sufficient. In what follows, we adopt the language proposed and defined in a recent paper [28]. Let Atm be a countable set of atomic propositions and let $Lit_{Atm} = Atm \cup \{\neg p : p \in Atm\}$, be the corresponding set of literals. Under this assumption, we can identify propositional valuations (or worlds) with maximal consistent conjunctions of literals. In the following, we will use interchangeably \neg and $\bar{}$ to specify the negation.

Definition 1. (Ceteris Paribus Deontic Language [28]). $\mathcal{L}_{CPDL+}(Atm)$ is a modal language which consists of atomic propositions $p, q, \dots \in Atm$, standard Boolean operators and the modal operators **O** and **P**. More precisely, it is the smallest set such that:

- if $p \in Atm$, then $p \in \mathcal{L}_{CPDL+}$
- if $\varphi, \psi \in \mathcal{L}_{CPDL+}$, then $\neg\varphi, \varphi \wedge \psi \in \mathcal{L}_{CPDL+}$
- if $\varphi, \psi \in \mathcal{L}_{CPDL+}$, then $\mathbf{O}(\psi|\varphi)$ and $\mathbf{P}(\psi|\varphi) \in \mathcal{L}_{CPDL+}$.

In particular, $\mathbf{O}(\psi|\varphi)$ models an obligation and $\mathbf{P}(\psi|\varphi)$ models a permission, and they have to be read respectively “under condition ψ , φ is obligatory” and “under condition ψ , φ is permitted.”

Notice that unconditional obligation and permission do not need to be added as primitives in the language of the logic as they are definable from conditional obligation and permission. We also do not need a primitive for the bilateral permission, or liberty, which consists in the permission both of a proposition and of its complement. For the sake of readability let us introduce the following definition:

Definition 2. For all $\varphi \in \mathcal{L}_{CPDL+}(Atm)$:

$$\begin{aligned} \mathbf{O}(\varphi) &=_{def} \mathbf{O}(\top|\varphi) \\ \mathbf{P}(\varphi) &=_{def} \mathbf{P}(\top|\varphi) \\ \mathbf{L}(\psi|\varphi) &=_{def} \mathbf{P}(\psi|\varphi) \wedge \mathbf{P}(\psi|\neg\varphi) \\ \mathbf{L}(\varphi) &=_{def} \mathbf{L}(\top|\varphi) \end{aligned}$$

Using this language and taking into account previous definitions, we can model obligations and permissions depicted in Example 1. We use the following abbreviations: d for “there is a dog”, c for “there is a cat”, f for “there is a fence”, w for “the fence is white”, b for “there is a bobcat”. We can express the norms enacted by Mary as follows (enumeration corresponds to the Example 1): (1) $\mathbf{O}(\neg d)$ (2) $\mathbf{L}(c)$ (3) $\mathbf{O}(d|f)$ (4) $\mathbf{O}(f|w)$ (5) $\mathbf{O}(\neg d|\neg f)$.

3.1. Contrary-to-duty obligations

The norms in our running example include a “contrary-to-duty” obligation, namely, obligations that are triggered by the violation of another obligation. According to the norm $\mathbf{O}(d|f)$, having a dog, and thus violating norm $\mathbf{O}(\neg d)$ (the prohibition to have dogs), triggers the obligation to have a fence. Contrary-to-duty obligations cannot be captured by standard deontic logic: according to those semantics, a proposition is obligatory if and only if it is true in every perfect (ideal) world, a world in which everything is as it should be. But the world in which a contrary-to-duty obligation is triggered is a sub-ideal one, since in it another obligation is violated, and it remains sub-ideal even if the contrary-to-duty obligation is complied with. In all perfect Cattowns there are no dogs, and in such cities, there are also no fences. So, according to the semantics for standard deontic logic, there cannot be an obligation to have fences in Cattown. Different approaches captured the idea of contrary-to-duty (see for instance [7, 19, 46]). A semantics that just distinguishes between ideal and non-ideal worlds cannot capture contrary-to-duty obligations: compliance with a contrary-to-duty obligation takes us to a world that is better – ceteris paribus – than a world in which the contrary-to-duty is not complied with, but which is still imperfect. A world with dogs and fences is ceteris paribus better than a world which has dogs and does not have fences, but the former is still worse than a world in which, ceteris paribus, there are no dogs. To capture contrary-to-duty obligations we need a semantics that distinguishes different levels of preferability so that we can distinguish the imperfect situation in which the main duty is violated, but the contrary-to-duty obligation is complied, from the even more imperfect situation in which both the main duty and the contrary-to-duty obligation are violated.

3.2. Strong and weak permissions

Our running example also shows the difference between strong (explicit) and weak (tacit) permissions, an issue much debated within deontic logic. On the one hand, cats are regulated: there is a norm that deals with cats ($\mathbf{L}(c)$), stating that both having and not having cats is allowed. On the other hand, nothing is said about bobcats. In other terms, while the normative system expresses an equal preference for having and not having cats (both are OK), it expresses no attitude towards bobcats. This distinction too cannot be captured by the semantics for standard deontic logic, where $\mathbf{P}(\phi)$ means that there is at least one perfect world in which ϕ is true, in which case $\neg\mathbf{O}(\neg\phi)$ is true. In legal logic, the most popular perspective to distinguish strong and weak permissions is an inferential one. To say that ϕ is strongly permitted relative to a normative system means that the normative entails $\mathbf{P}(\phi)$. To say that ϕ is weakly permitted means that the normative system under consideration does not entail a corresponding prohibition $\mathbf{O}(\neg\phi)$. Here we shall provide a different way to capture the difference between strong and weak permission, based on the distinction between indifference (equal ceteris-paribus preference) and incomparability (absence of any preference).

3.3. Ceteris-Paribus Preferences

We assume a set of variables (also called features or attributes) $V = \{X_1, \dots, X_n\}$ each one with a binary domain, i.e. the domain of each $X_i \in V$ is such that $Dom(X_i) = \{x_i, \bar{x}_i\}$. An instantiation $Ass(X)$, where $X \subseteq V$, is an assignment of values to X . When $X = V$ then we call it a *complete assignment* or *outcome* or *world*, otherwise we call it a *partial assignment*.

A preorder is a reflexive and transitive binary relation \leq over the set of all outcomes $U = 2^V$, which are usually denoted by o, w, \dots, v . In what follows, $w \leq v$ means that v is at least as good/ideal as w , $w \approx v$ means that w is equivalent to v which is an abbreviation for $w \leq v$ and $v \leq w$. Moreover, $w < v$ means that v is better/more ideal than w to be an abbreviation of $w \leq v$ and $v \not\leq w$. Given two possible outcomes $w, v \in O$, if neither $w \leq v$ nor $v \leq w$ are valid, then we say that w and v are incomparable, denoted with $w \mathbb{Z} v$. The notation $o[V_i]$ returns the value of the variable V_i in the outcome o , while $o[-V_i]$ returns the value of all the variables but V_i in the outcome o .

We borrow some notations from [5] in order to define *ceteris paribus* semantics. In our analysis of the semantics of preferences, we assume that the set of variables V is partitioned into three sets: A set X which includes the variables over whose values the preference is expressed, a set Y which includes the variables whose values are irrelevant to the preference, a set Z which includes the variables that are assumed to remain constant, i.e., that provide the context for the preference.

Definition 3. (Ceteris-Paribus Preference). Let V be a set of variables and X, Y, Z partitions of V with X be nonempty set. Let $z \in \text{Ass}(Z)$ and $x_1, x_2 \in \text{Ass}(X)$, we say that x_2 is *ceteris-paribus at least as good as* (resp. *better than*) x_1 given z if and only if for all $y \in \text{Ass}(Y)$, $x_1 y z \preceq x_2 y z$ (resp. $x_1 y z < x_2 y z$).

We use $z : x_1 \preceq_Y x_2$ ($<_Y$ resp.) as an abbreviation for it (we will omit the subscript when it is clear from the context).

As the preference for x_1 over x_2 , in the context of z , is independent from the values of Y , we may say that X is preferentially independent of Y if $Z = \emptyset$, otherwise we say that X is conditionally preferentially independent of Y given the assignment of Z .

Moreover, from Definition 3 we derive the following notation regarding ceteris-paribus relation between two outcomes u, v .

Definition 4. Given two outcomes $u, v \in U$, we say that u is *ceteris-paribus at least as good as* (resp. *better than*) v relative to X_i , if and only if they differ only in the value of the variable X_i and we abbreviate it with $v \preceq_{(-X_i)} u$ (resp. $v <_{(-X_i)} u$).

The previous definition introduces a notation that is useful in those situations where u and v are two complete assignments that differ only in the value for the variable X_i , that is $u = x_i y z$ and $v = x'_i y z$, with $y \in \text{Ass}(Y)$, $z \in \text{Ass}(Z)$, $Y \cup Z = V \setminus \{X_i\}$ and $x_i, x'_i \in \text{Dom}(X_i)$. Then adopting Definition 3 this is similar to say that $z : x'_i \preceq_Y x_i$ (resp. $z : x'_i <_Y x_i$).

Let us refer to Example 1. Mary's community preferences are over a set of five features, specifically $V = \{C, D, F, W, B\}$, with the following domains: $\text{Dom}(C) = \{c, \bar{c}\}$, $\text{Dom}(D) = \{d, \bar{d}\}$, $\text{Dom}(F) = \{f, \bar{f}\}$, $\text{Dom}(W) = \{w, \bar{w}\}$ and $\text{Dom}(B) = \{b, \bar{b}\}$. The meaning of variables is the same introduced in Section 3 for atoms: for instance, variable C concerns cats, its value is c or \bar{c} depending on whether there is a cat or not.

Since V has 5 elements, the set of all outcomes $U = 2^V$ contains 32 possible outcomes. For the sake of readability, in this section, we do not consider variables W and C . Examining a reduced number of outcomes will allow for making the partial orders limited in size. Thus, considering the subset $\{B, D, F\}$, the set of outcomes is denoted by the complete assignments to the contemplated variables:

$$U = \{bdf, \bar{b}df, b\bar{d}f, b\bar{d}\bar{f}, \bar{b}\bar{d}\bar{f}, \bar{b}\bar{d}f, \bar{b}\bar{d}\bar{f}\}$$

The different attitudes of the community can be represented with a set of ceteris paribus preferences. In particular, with respect to \neg dogs-assignments (i.e., $D = \bar{d}$) and dog-assignments (i.e. $D = d$), the negative attitude to dogs is represented by the preference for \neg dogs-assignments over ceteris-paribus dog-assignments (i.e. $d <_{V \setminus D} \bar{d}$):

$$bdf < \bar{b}\bar{d}f, \bar{b}df < \bar{b}\bar{d}\bar{f}, b\bar{d}\bar{f}, b\bar{d}\bar{f} < \bar{b}\bar{d}\bar{f}$$

Preference is ceteris-paribus (about d) in the sense that it only concerns the comparison between dog-assignments and \neg dog-assignments that are equal in all the rest (bobcats and fences). These are all pairs of outcomes u and v such that the evaluations of u and v coincide in all values except for D .

Preference for having fences when there are dogs in the house is captured by preferences for fence-assignments over the ceteris-paribus \neg fence-assignments, taking only dog-assignments into consideration (i.e. $d : \bar{f} <_B f$):

$$b\bar{d}\bar{f} < bdf, \bar{b}\bar{d}\bar{f} < \bar{b}df$$

Instead, unawareness of bobcats can be modelled through incomparability between bobcat-assignments and \neg bobcat-assignments, which expresses unfamiliarity to the presence of bobcats, i.e. for all $x_1, x_2 \in \text{Ass}(D)$ and for all $y_1, y_2 \in \text{Ass}(F) : b x_1 y_1 Z \bar{b} x_2 y_2$.

At last, let us include again the variable C in our analysis, so we can have cats and not have them. This can be represented through the indifference for cat-assignments over ceteris paribus \neg cat-assignments (i.e. $c \approx_{V \setminus C} \bar{c}$). These are all the pairs of outcomes u and v such that v is ceteris-paribus at least as good as u and vice-versa (i.e., $(u \preceq_{(-C)} v) \wedge (v \preceq_{(-C)} u)$), that is for instance:

$$c\bar{b}\bar{d}f \approx \bar{c}\bar{b}\bar{d}f, c b d \bar{f} \approx \bar{c} b d \bar{f}$$

3.4. GCP-nets

In order to be able to represent the described scenario, we propose to leverage existing preference frameworks from AI, namely GCP-nets (generalized CP-nets) [18]. GCP-net is an extension of CP-net [5] that allows incomplete preference orders over some values of the features. GCP-nets and related extensions of the formalism [11, 12, 18] are often on multi-valued variables, this means that variables are not strictly boolean. Nevertheless, in this work, we assume that all the variables have binary domains. We start by introducing and defining some useful notions.

We borrow the following definition from [18]:

Definition 5. (Generalized CP-net). A generalized CP-net C (for short, a GCP-net) over V is a set of conditional preference rules. For $x \in V$, we define $p^+(x)$ and $p^-(x)$, as follows: $p^+(x)$ is equal to the disjunction of all p such that there exists a rule $p : \neg x < x$ in C ; $p^-(x)$ is the disjunction of all p such that there exists a rule $p : x < \neg x$ in C . We define the associated directed graph G_C (the dependency graph) over V to consist of all pairs (y, x) of variables such that y appears in either $p^+(x)$ or $p^-(x)$.

For each variable $V_i \in V$, a $CPT(V_i)$ is a set of cp-statements, each one represents an ordering over the specific domain $Dom(V_i)$ given the assignment to the parents of V_i , which are denoted with $Pa(V_i)$. For instance, $CPT(F) = \{d : \bar{f} < f, \bar{d} : f < \bar{f}\}$. In its original formulation, each $CPT(V_i)$ reports strict linear order over the values of the domain of V_i given the partial assignment to $Pa(V_i)$. In a GCP-net, CP-tables may be incomplete (i.e., for some values of its parents, the preferred value of a variable X may not be specified) and/or locally inconsistent (i.e., for some values of its parents, the table may both contain the information that $x \in Dom(X)$ is preferred and the information that $\bar{x} \in Dom(X)$ is preferred).

The semantics is connected with the induced preference graph over all the outcomes. A directed edge between a pair of outcomes (o_i, o_j) , 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.

The semantics of a GCP-net is a preorder over all the outcomes, i.e. a reflexive and transitive binary relation over the set of complete assignment of values to the variables of the GCP-net. Thus, for any two outcomes o, u which differs only on the value of one variable $X_i \in V$ we have:

- $o \preceq u$ if $o[X_i] \preceq u[X_i]$ given $o[Pa(X_i)]$
- $o \not\preceq u$ if there is not a cp-statements for $o[X_i], u[X_i]$ given $o[Pa(X_i)]$

The partial order induced by the GCP-net is denoted as $Ord_{\mathcal{N}}$.

The GCP-net formalism provides a qualitative compact representation that is useful to represent scenarios similar to the one depicted by the reduced deontic logic introduced in Section 3.

4. Deontic Language and GCP-nets: bridging the gap

A set of obligations and liberties (bilateral permissions) expressed in the restricted language defined in Section 3 can be represented using a GCP-net, we will call it a prescriptive CP-net. Note that we do not provide for the representation of unilateral permission. In fact, we assume that unilateral permissions must be implied either by an obligation (in case the complement is forbidden) or by liberty (in case the complement is also permitted). Indeed, the restriction on obligations and liberties where antecedents are conjunctions of literals and consequent is a single literal is compatible with the syntax of generalized CP-nets.

4.1. Incomparability

In the previous section, we introduced the notion of incomparability. In the ceteris paribus semantics, the notion of incomparability does not allow to differentiate between two different states of affairs: one where two outcomes are incomparable because nothing is said about some values of a feature, and another, where two or more variables are independent and thus preferences are described over the domains of these variables no matter the assignment of the others. Both cases induce a preference graph where some outcomes are incomparable because there does not exist a path between them. But, while the former case describes a lack of information, possibly because the individual does not know or does not have any information about some features (for instance, in our example the lack of information about bobcats), the latter describes a situation where preferences are reported on all the features but not on all combinations of them (for instance, in the running example norms do not tell whether the situation in which there is not a dog and no fence—whatever the color—is bet-

ter or worse than the situation where there is not a dog and there is a white fence). In this last situation, the issue is connected with the color of the fence and the lack of dependencies between the fence's presence and the color. This seems to be simpler to fix: the individual already has preferences over values of the single features, thus the incomparability can be overcome by adding a contextual dependency over some values. For instance, forcing the choice of white color only when a fence is present and the liberty of choosing the color whenever the fence is not present.

Definition 6. (Strong incomparability). Given a pre-order P , two outcomes $w, v \in P$ are said to be *strongly incomparable* if they belong to two different components of P .

Strong incomparability describes the relation among outcomes of different components. They are incomparable because some information is missing: specifically, preferences are not reported for all the values in the domain of some features. This lack of information can be hard to fix since it entails a lack of knowledge about the scenario.

The next definition refers to *weakly connected* components: a directed graph is said to be weakly connected if the undirected graph resulting from removing the orientation of the edges is connected, i.e. if any pair of vertexes w, v in the undirected graph has a path from w to v .

Definition 7. (Weak incomparability). Given a pre-order P , two outcomes $w, v \in P$ are said to be *weakly incomparable* if they belong to the same weakly connected component of P but there does not exist a path from w to v or vice-versa.

Weak incomparability describes a less problematic situation. In this case, preferences are reported for values in the domain, but the incomparability is due to some missing dependencies among features. We can reduce the number of incomparable outcomes by adding dependencies among features.

4.2. Modeling Deontic Language with GCP-nets

In this section, we define the GCP-net that is induced by a given set of prescriptions. We call this extension a *Prescriptive CP-net*.

Definition 8. Let $C \subseteq \mathcal{L}_{\text{CPDL}^+}(\text{Atm})$ be a set of statements. A prescriptive CP-net $\mathcal{N}_C = (G, \text{CPT})$ is a generalized CP-net (with G a directed graph over

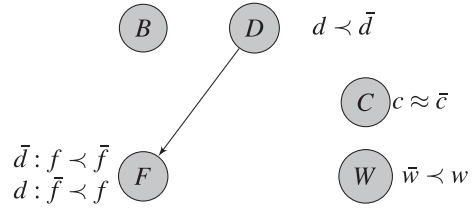


Fig. 1. The GCP-net which represents Mary's community preferences of Example 1.

a set of variables V) defined as follows:

- for each atomic proposition $v_i \in \text{Atm}$ corresponds a variable $V_i \in V$ with $\text{Dom}(V_i) = \{v_i, \bar{v}_i\}$
- for each conditional obligation $\mathbf{O}(\psi|\phi) \in C$ and each conditional liberty $\mathbf{L}(\psi|\phi) \in C$, dependencies are introduced in G . Specifically, for each literal $x_j \in \psi$, a directed edge is introduced between X_j and Φ in G , thus X_j becomes a parent of Φ
- each obligation $\mathbf{O}(\psi|\phi) \in C$ induces a strict order over $\text{Dom}(\Phi)$ given the assignment ψ such that $\text{CPT}(\Phi) = \{\psi : \bar{\phi} < \phi\}$ and each liberty $\mathbf{L}(\psi|\phi) \in C$ induces a weak order over $\text{Dom}(\Phi)$ given the assignment ψ such that $\text{CPT}(\Phi) = \{\psi : \bar{\phi} \approx \phi\}$.

Notice that variables with empty CP-tables or partially empty CP-tables may exist. These variables will induce incomparability in the preorder.

Based on preferences described in Example 1, Fig. 1 reports the dependency graph and CP-tables of the prescriptive CP-net which represents Mary's community preferences. The prescriptive CP-net is over the set of variables $V = \{B, C, D, F, W\}$. Notice that: (a) Variable B does not have a CP-table because the community did not express any preferences about bobcats. (b) The indifference on the values of variable C describes the liberty to have or not cats.

The strong attitude about dogs induces strict order over the domain of the variable D , similarly, orders over the variable W depend on the strong preference for the white color. Orders over the variable F depend on whether or not there is a dog.

The partial order induced by the GCP-net is reported in Fig. 2. The binary relation among outcomes is based on dependencies and preferences reported in the CP-tables of the GCP-net. For instance, outcome $\bar{c}\bar{d}f\bar{w}\bar{b}$ is indifferent to outcome $\bar{c}\bar{d}f\bar{w}\bar{b}$, this is due to the indifference on values of the variable C all the rest being equal. In order to

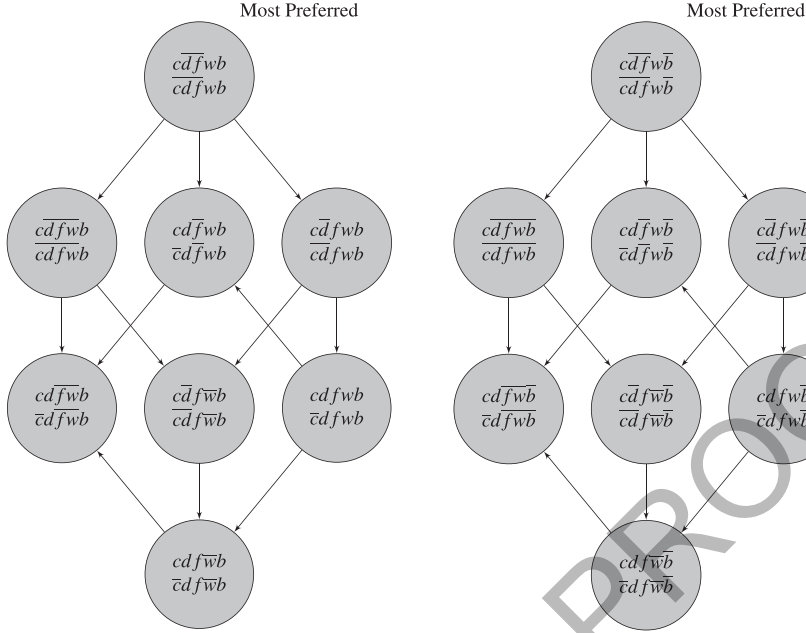


Fig. 2. The partial order induced by the GCP-net in Fig. 1: for the sake of readability, we group into the same nodes some outcomes of the preference model. Outcomes in the same node are indifferent, this is due to the indifference over the values of variable C .

maintain a low level of complexity in the representation of the partial order, we group into the same node outcomes that differ only in the value of the variable C .

Another example is $cd\bar{f}w\bar{b} < cd\bar{f}w\bar{b}$, this is because \neg dog-assignments are ceteris-paribus better than dog-assignments. Instead, following Definition 7, $cd\bar{f}w\bar{b}$ and $cd\bar{f}w\bar{b}$ are weakly incomparable: albeit they belong to the same weakly connected component, there does not exist a path from one outcome to the other nor vice-versa.

Notice that the preference graph has two components. Outcomes in the two components differ on the assignment of the variable B for which no preferences are expressed. Thus each component represents a scenario: one in which we consider the presence of bobcats, and the other in which there are no bobcats. Following Definition 6, $cd\bar{f}w\bar{b}$ and $cd\bar{f}w\bar{b}$ are strongly incomparable.

4.3. Mapping Preference Models to GCP-nets

In this section we shall provide a ceteris paribus semantics for the deontic language provided above. In the next session, we shall show that this semantics identifies models that correspond to GCP-nets. The semantics of the deontic language is defined in terms of preference relations on worlds.

Definition 9. (Preference models of $\mathcal{L}_{\text{CPDL}^+}$). A preference model of $\mathcal{L}_{\text{CPDL}^+}$, $M = (U, \preceq)$ is a pre-order \preceq on the set U of outcomes.

A set of ceteris-paribus conditional obligations and liberties C is *consistent* if it has at least one model, and inconsistent otherwise. A set of ceteris-paribus conditional obligations and liberties C entails another preference formula C' , written $C \models C'$, if every model of C is also a model of C' . In the following definition, an outcome refers to conjunctions of propositional literals referring to each variable in V exactly once, and which are consistent. Thus, for instance, ϕ refers to the valuation of variable Φ . Satisfaction for formulas in the language $\mathcal{L}_{\text{CPDL}^+}(\text{Atm})$ is defined as follows, where $u \models \varphi$ means that the propositional formula φ is true at the valuation corresponding to the outcome u :

Definition 10. (Satisfaction in $\mathcal{L}_{\text{CPDL}^+}$ [28]). Let $M = (W, \preceq)$ be a preference model and let $w \in W$. Then:

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

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

The class of models satisfying a set of norms C is denoted with \mathcal{P}_C . Among all the possible $M \in \mathcal{P}_C$, $rel(\mathcal{P}_C) \subseteq \mathcal{P}_C$ is the subset of models that refers only to consistent conjunctions of propositional literals in C . Then $M_C = (U, \leq_C)$ is the preference model such that for each $M = (U, \leq) \in rel(\mathcal{P}_C)$, $\leq_C \subseteq \leq$.

Proposition 1. As a liberty is a bilateral permission ($\mathbf{L}(\psi|\phi) =_{def} \mathbf{P}(\psi|\phi) \wedge \mathbf{P}(\psi|\neg\phi)$), we have that

$$\begin{aligned} M \models \mathbf{L}(\psi|\phi) &\iff \forall v, u \in U : v \models \psi, u \models \psi, \\ &\text{if } v \models \phi \text{ and } v <_{(-\phi)} u \text{ then } u \models \phi, \text{ and} \\ &\text{if } v \models \neg\phi \text{ and } v <_{(-\phi)} u \text{ then } u \models \neg\phi \end{aligned}$$

Given the previous definitions we now show that a set of ceteris-paribus norms in $\mathcal{L}_{\text{CPDL}^+(Atm)}$ is expressible with a prescriptive CP-net, this means that the preorder induced by the GCP-net is a model which satisfies C .

Theorem 1. Given a set of norms $C \subseteq \mathcal{L}_{\text{CPDL}^+(Atm)}$ and the prescriptive CP-net \mathcal{N}_C , then $Ord_{\mathcal{N}_C} = M_C$.

Proof. This is equivalent to proving that for each norm $N \in C$

$$M_C \models N \text{ iff } Ord_{\mathcal{N}_C} \models N \quad (1)$$

We show how to prove (1) showing how this can be done for obligations. Let start by assuming that $M_C \models \mathbf{O}(\psi|\phi)$, by definition this is equivalent to $\forall u, v \in U : u <_{\psi, \phi} v$, thus for consistency of \mathcal{N}_C with C :

$$\begin{aligned} Ord_{\mathcal{N}_C} \models \psi : \neg\phi < \phi \\ Ord_{\mathcal{N}_C} \models \forall y_1, y_2 \in (V - \Psi - \{\Phi\}) : \\ \neg\phi\psi y_1 < \phi\psi y_2 \end{aligned}$$

by definition, this is equivalent to

$$\begin{aligned} Ord_{\mathcal{N}_C} \models \forall u, v \in U : u, v \models \psi, u \models \neg\phi, v \models \phi \\ Ord_{\mathcal{N}_C} \models \mathbf{O}(\psi|\phi) \end{aligned}$$

This proves \implies direction. We show the other way round (\impliedby) by contradiction. Let us assume that $Ord_{\mathcal{N}_C} \not\models \mathbf{O}(\psi|\phi)$, thus $\mathbf{O}(\psi|\phi) \notin \mathcal{N}_C$ but this means that also $\mathbf{O}(\psi|\phi) \notin C$. But this is a contradiction. \square

On the basis of the GCP-net models described above the following proposition holds.

Proposition 2. A contrary-to-duty framework having the following prescriptions $\mathbf{O}(\phi), \mathbf{O}(\neg\phi|\psi), \mathbf{O}(\phi|\neg\psi)$ is modelled by a consistent prescriptive CP-net.

Proof. The CP-net is reported in Fig. 3a, and the induced partial order is depicted in Fig. 3b. \square

5. Possible applications

Our analysis shows that a set of norms expressed in deontic logic can be translated into a corresponding GCP-net without loss of semantics, thus providing a conditional preference network giving a compact representation.

On the contrary, GCP-nets allow different scenarios to be modelled and interesting properties and information about the set of norms to be inferred. Adopting the GCP-net framework we are able to reason about:

Dominance testing: the problem consists in deciding whether one outcome v dominates another outcome u , i.e. $u < v$.

Consistency testing: the problem consists in finding out whether there is a dominance cycle in the preorder defined by a GCP-net, i.e. whether there is an outcome that dominates (is preferred to) itself.

Both of the aforementioned problems are usually far from easy. In general both dominance and consistency for CP-nets can be NP-complete [5, 18]. But when the CP-net is over a set of binary variables, and the dependency graph is a tree, the dominance testing becomes linear in the number of variables, if it is a poly-tree then dominance testing becomes polynomial in the number of variables [5]. Moreover, in many scenarios, the presence of cycles in the dependency graph means that the network is inconsistent [14]. Thus, consistency testing would allow inconsistencies (contradiction) to be determined in a set of norms while the dominance test allows for testing whether one world is better than another. There are other possible applications that can be enabled with the adoption of a GCP-net, we briefly describe them in the next paragraphs.

Preference comparison.. Another application of this preference framework is in the comparison of prefer-

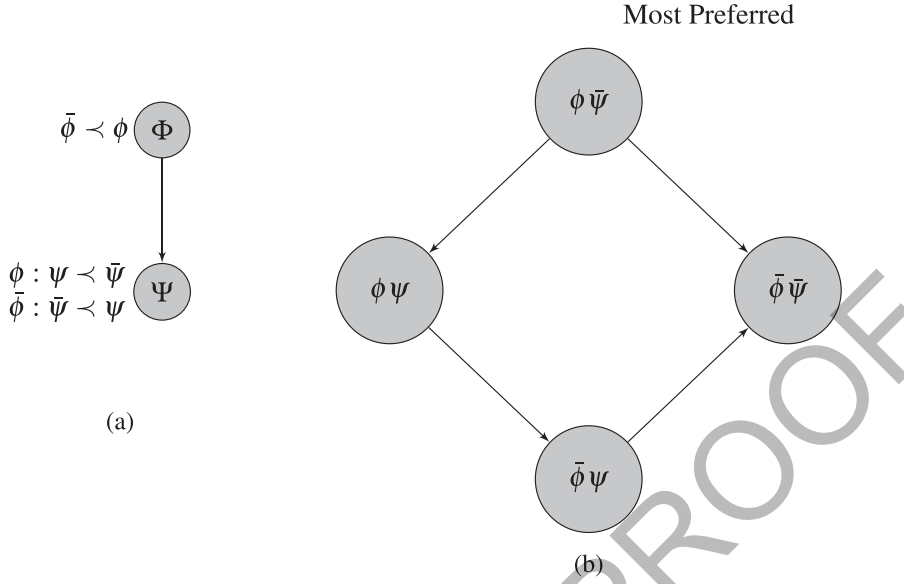


Fig. 3. (3a) The prescriptive CP-net induced by the duty framework of Proposition 2. (3b) The partial order induced by the prescriptive CP-net which satisfies the duty framework of Proposition 2.

ences. We can think of the decision-making process as an evaluation of the resulting state-of-affairs. We choose one among all the available options because we prefer the resulting state-of-affair over the others. Thus, we can think of moral and normative judgments as a form of preferences, driven by moral or normative reasoning [43]. Finding a feasible way to model, reason, and compare these preferences are central in the field of Artificial Intelligence in order to ensure a convergence of AI agents' behavior with humans' values.

One possible solution is to compute the distance among preferences by exploiting the compact representation of such domains. In recent studies, preferences represented with compact representation are used in metric spaces in order to define how similar they are [26, 30]. When both the normative system and individual preferences are represented compactly, preferences comparisons might be useful to understand whether an agent is deviating from the desired behaviour—which is depicted by the legal framework. When this is the case, the system might intervene to find a trade-off which better aligns with the set of values [32].

Let us expand Example 1 in order to depict a situation in which preferences of individuals deviate from the normative system.

Example 2. (Running Example continued). The inhabitants of Cattown are not homogeneous in their

preferences, but they have different opinions on the norms that Mary enforced. Bob, for instance, moved to Cattown because of his job. He is a dog lover and he has a beautiful Golden Retriever with him. At the same time, he is respectful of the law and shares the same worries about fences when a dog is on a property. Bob's preferences are reported in the CP-net of Fig. 4a.

On the other side, a group of activists in Cattown is fighting the normative system and attacking the set of obligations and permissions which should regulate the presence of dogs, cats and fences. They think that anyone—humans and animals—should move freely in the city and thus they prefer to live without fences. Moreover, many of the people in the group are dog lovers and are indifferent to cats. The preferences of this group are represented in Fig. 4b.

The previous example describes the different points of view that might exist in society: different perspectives that people might have over the same set of features, a situation which is very common in practice. Sometimes norms are closed to the preferences of people, some others are not.

We can use a metric space to measure the similarity of preferences in Mary's community [30, 33]. This may help to understand which individual is adhering to the normative system or how to adjust the set of obligations and permissions in order to mitigate differences in society.

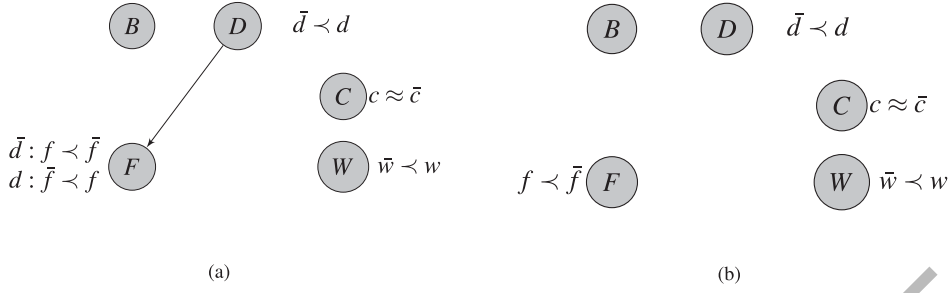


Fig. 4. (4a) The GCP-net which represents Bob preferences of Example 2. (4b) The GCP-net represents preferences of the group of activists as described in Example 2.

A definition of distance over GCP-nets does not exist yet, but existing metrics work on the induced partial orders making the comparison feasible.

These kinds of assessments may be used to study possible ways of changing a normative system in order to get closer to the preferences of individuals in the community. This should be done to reduce divergences between a group of people and the normative system. On the contrary, they can also be used to identify which part of the normative system is more infringed.

Modelling factual and deontic detachment.. The proposed language allows two important principles of detachments related to the conditional nature of the modeled norms to be modelled:

- **Factual detachment:** It represents a situation where a conditional norm $\mathbf{O}(\phi|\psi)$ exists. If ϕ is true in the scenario, then we can conclude that in that specific scenario it ought to be that ψ .
- **Deontic detachment:** It represents a situation where two norms $\mathbf{O}(\phi)$ and $\mathbf{O}(\phi|\psi)$ exist, the first prescribing the state which is the antecedent of the second. We can then conclude that in the scenario it ought to be ψ .

The CP-net framework allows us to model factual detachment: from fact ϕ and conditional norm $\mathbf{O}(\phi|\psi)$ we infer ψ . In the CP-net, factual detachment can be computed by considering the partial assignments of variables in P corresponding to ϕ : for each literal p or $\neg p$ in ϕ the corresponding atom p in P is assigned respectively true or false. Then for each CP-table of the descendants of P , we consider only preference orders where atoms in P have the given partial assignments. The process is repeated recursively for each atom in P . Called m the number of atoms in P , if the dependency graph is acyclic, then this can be done in $O(m)$.

On the contrary, deontic detachment cannot be captured by our preference model. Indeed, inferring $\mathbf{O}(\psi)$ from $\mathbf{O}(\phi|\psi)$ when $\mathbf{O}(\phi)$ is the case, corresponds to removing the edges in the dependency graph between variables corresponding to ϕ and to ψ . This move is not allowed since within the CP-net framework it is not possible to remove a dependency between two variables unless it is a fake dependency. A fake dependency is introduced in the dependency graph when it concerns complementary literals p and $\neg p$. This would apply for instance to a pair of obligations $\mathbf{O}(p|q)$ and $\mathbf{O}(\neg p|q)$ (i.e., an obligation that has the same consequent and complementary antecedents).

Modelling disjunctive obligations.. The proposed language does not allow for deontic operators \mathbf{O} and \mathbf{P} to be applied to disjunctions. This means that the consequent of norms cannot contain disjunctions.

This limitation, however, does not really affect the expressiveness of the language, since an obligation to realise at least one of two states of affairs ϕ and ψ can be modelled through the combination of two obligations $\mathbf{O}(\neg\phi|\psi)$ and $\mathbf{O}(\neg\psi|\phi)$, possibly with permissions to realise both, $\mathbf{P}(\phi)$ and $\mathbf{P}(\psi)$, if this is meant to be the case.

Let us exemplify this situation by expanding our running example.

Example 3. (Running example continued). The municipality of Cattown is tolerant towards the ownership of dogs, as long as they are kept in private properties. As already mentioned, for some security reasons, if there is a dog on a property then there should also be a fence. This would allow the dog to be left free on the property. Otherwise, the municipality tolerates the absence of a fence, but in these cases, dogs must be kept on a leash.

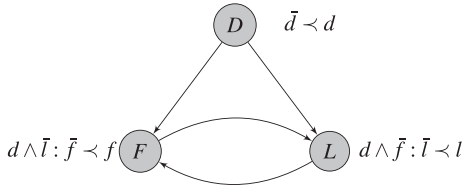


Fig. 5. The prescriptive CP-net which represents the normative system described in the Example 3.

The previous example describes a situation where mutually exclusive norms can be enforced. Using our language, the normative system can be represented with the following obligations:

$$\mathbf{O}(d \wedge \neg f | l) \wedge \mathbf{O}(d \wedge \neg l | f)$$

Focusing only on the norms in Example 3, we can build the prescriptive CP-net represented in Fig. 5. Notice that in this case the dependency graph of the related prescriptive CP-net is not acyclic.

6. Conclusions

GCP-nets, and their variants, are an interesting and emerging preference frameworks that enable the representation of conditional preferences in a compact way. These frameworks are adopted in many different scenarios. For instance, in recommender systems, it is used to improve the accuracy of personalized search [51], and recently it was also employed to capture and represent how the deliberation process of a group of individuals switches to different moral frameworks in order to make a moral decision [1].

In this work, we have shown how to model contrary-to-duty obligations and permissions using GCP-nets. Using this formalism, we have also shown how to capture the distinction between strong and weak permissions. In order to do that, we adopted a restricted deontic logic representing a set of obligations and permissions. The resulting normative system induces a preference order over the possible worlds.

The adoption of these preference representations shall allow the combination of machine learning methods and standard reasoning frameworks to compare preferences from different users [33].

Acknowledgments

Roberta Calegari, Andrea Loreggia, and Giovanni Sartor have been supported by the H2020 ERC Project “CompuLaw” (G.A. 833647).

Authors have no Conflict of Interest (CoI) to share.

References

- [1] E. Awad, S. Levine, A. Loreggia, N. Mattei, I. Rahwan, F. Rossi, K. Talamadupula, J. Tenenbaum and M. Kleiman-Weiner, When is it morally acceptable to break the rules? a preference-based approach. In *12th Multidisciplinary Workshop on Advances in Preference Handling (MPREF 2020)*, 2020.
- [2] T. Bench-Capon and S. Modgil, Norms and value based reasoning: justifying compliance and violation, *Artificial Intelligence and Law* **25**(1) (2017), 29–64.
- [3] G. Boella, L. Van Der Torre and H. Verhagen, Introduction to normative multiagent systems, *Computational & Mathematical Organization Theory* **12**(2) (2006), 71–79.
- [4] G. Booch, F. Fabiano, L. Horesh, K. Kate, J. Lenchner, N. Linck, A. Loreggia, K. Muresan, N. Mattei, F. Rossi, et al. Thinking fast and slow in AI. In *Proceedings of the AAAI Conference on Artificial Intelligence* **35** (2021), 15042–15046.
- [5] C. Boutilier, R.I. Brafman, H.H. Hoos and D. Poole, Reasoning with conditional ceteris paribus preference statements. In *Proc. of the 15th Conf. on Uncert. in Art. Intell., UAI’99*, pages 71–80, 1999.
- [6] R. Calegari, G. Ciatto, V. Mascardi and A. Omicini, Logicbased technologies for multi-agent systems: A systematic literature review, *Autonomous Agents and Multi-Agent Systems* **35**(1) (2021), 1:1–1:67. Collection “Current Trends in Research on Software Agents and Agent-Based Software Development”.
- [7] J. Carmo and A.J. Jones, *Deontic Logic and Contrary-to-Duties*, chapter Handbook of Philosophical Logic: Volume 8, pages 265–343. Springer Netherlands, Dordrecht, 2002.
- [8] R.M. Chisholm, *Perceiving: A Philosophical Study*. Cornell University Press, Ithaca, NY, 1957.
- [9] C. Cornelio, M. Donini, A. Loreggia, M.S. Pini and F. Rossi, Voting with random classifiers (VORACE). In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1822–1824, Richland, SC, 2020.
- [10] C. Cornelio, M. Donini, A. Loreggia, M.S. Pini and F. Rossi, Voting with random classifiers (VORACE): theoretical and experimental analysis. *Auton. Agents and Multi-Agent Systems* **35**(2) (2021), 22.
- [11] C. Cornelio, J. Goldsmith, N. Mattei, F. Rossi and K.B. Venable, Updates and uncertainty in CP-nets. In S. Cranefield and A. Nayak, editors, *AI 2013: Advances in Artificial Intelligence*, pages 301–312, Cham, 2013. Springer International Publishing.
- [12] C. Cornelio, U. Grandi, J. Goldsmith, N. Mattei, F. Rossi and K.B. Venable, Reasoning with PCP-Nets in a multi-agent context. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, AAMAS’15*, pages 969–977, Richland, SC, 2015.

- [13] V.T. da Silva, C. Braga and K. Figueiredo, A modeling language to model norms. *Coordination, Organization, Institutions and Norms in Multi-Agent Systems@ AAMAS2010*, page 25, 2010.
- [14] C. Domshlak and R.I. Brafman, CP-Nets: Reasoning and consistency testing. In *Proceedings of the Eighth International Conference on Principles of Knowledge Representation and Reasoning, KR'02*, pages 121–132, San Francisco, CA, USA, 2002.
- [15] C. Domshlak, E. Hüllermeier, S. Kaci and H. Prade, Preferences in AI: An overview, 2011.
- [16] E.S.S. Freire, M.I. Cortés, R.M.D.R. Júnior, Ê.J.T. Gonçalves and G.A.C. De Lima, Normas-ml: Supporting the modeling of normative multi-agent systems, *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal* 8(4) (2019), 49–81.
- [17] C. Garion, *Apports de la logique mathématique en ingénierie des exigences*. Ph.d. thesis, Université de Toulouse, 2002.
- [18] J. Goldsmith, J. Lang, M. Truszczyński and N. Wilson, The computational complexity of dominance and consistency in CP-nets, *Journal of Artificial Intelligence Research* 33(1) (2008), 403–432.
- [19] G. Governatori and A. Rotolo, Logic of violations: A gntzen system for reasoning with contrary-to-duty obligations, *Australasian Journal of Logic* 4 (2006), 193–215.
- [20] G. Governatori and A. Rotolo, Justice delayed is justice denied: Logics for a temporal account of reparations and legal compliance. In J. Leite, P. Torroni, T. Ágotnes, G. Boella and L. van der Torre, editors, *Computational Logic in Multi-Agent Systems*, pages 364–382, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
- [21] U. Grandi, A. Loreggia, F. Rossi and V. Saraswat, From sentiment analysis to preference aggregation. In *Proceedings of the International Symposium on Artificial Intelligence and Mathematics (ISAIM-2014)*, pages 1–5, Florida, USA, 2014. Fort Lauderdale.
- [22] U. Grandi, A. Loreggia, F. Rossi and V. Saraswat, A borda count for collective sentiment analysis, *Annals of Mathematics and Artificial Intelligence* 77(3–4) (2016), 281–302.
- [23] B. Hansson, An analysis of some deontic logics. In *Deontic Logic: Introductory and Systematic Readings*, pages 121–147. Springer, 1970.
- [24] S.O. Hansson, Preference-based deontic logic (pdl), *Journal of Philosophical Logic* 19(1) (1990), 75–93.
- [25] S.O. Hansson, *The Structure of Values and Norms*. Cambridge Studies in Probability, Induction and Decision Theory. Cambridge University Press, Cambridge, 2001.
- [26] M. Li and B. Kazimipour, An efficient algorithm to compute distance between lexicographic preference trees. In *Proc. of the 27th Intern. Joint Conf. on Art. Intelligence, IJCAI-18*, pages 1898–1904, 7 2018.
- [27] A. Loreggia, E. Lorini and G. Sartor, A ceteris paribus deontic logic. In *35th Italian Conference on Computational Logic (CILC 2020)*, volume 2710, pages 248–262. CEUR, 2020.
- [28] A. Loreggia, E. Lorini and G. Sartor, Modelling ceteris paribus preferences with deontic logic, *Journal of Logic and Computation* 32(2) (2022), 347–368.
- [29] A. Loreggia, N. Mattei, F. Rossi and K. Venable, Modeling and reasoning with preferences and ethical priorities in ai systems, *Ethics of Artificial Intelligence*, pages 127–154, 2020.
- [30] A. Loreggia, N. Mattei, F. Rossi and K.B. Venable, On the distance between CP-Nets. In *Proc. of the 17th Intern. Conf. on Auton. Agents and MultiAgent Systems, AAMAS '18*, pages 955–963, 2018.
- [31] A. Loreggia, N. Mattei, F. Rossi and K.B. Venable, Preferences and ethical principles in decision making. In *AIES '18: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics and Society*, page 222, New York, NY, USA, 2018. Association for Computing Machinery.
- [32] A. Loreggia, N. Mattei, F. Rossi and K.B. Venable, Value alignment via tractable preference distance. In R.V. Yampolskiy, editor, *Artificial Intelligence Safety and Security*, chapter 16, pages 249–260. CRC Press, Boca Raton, Florida, USA, 2018.
- [33] A. Loreggia, N. Mattei, F. Rossi and K.B. Venable, CPMetric: Deep siamese networks for metric learning on structured preferences. In *Artificial Intelligence. IJCAI 2019 International Workshops*, pages 217–234, Cham, 2020. Springer International Publishing.
- [34] D. Makinson, On a fundamental problem of deontic logic. In P. McNamara and H. Prakken, editors, *Norms, logics and information systems: New studies in deontic logic and computer science*. IOS Press, 1999.
- [35] M. Mohajeriparizi, G. Sileno and T. van Engers, Preference-based goal refinement in bdi agents. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, pages 917–925, 2022.
- [36] A. Omicini and R. Calegari, Injecting (micro)intelligence in the IoT: Logic-based approaches for (M)MAS. In *Mass. Multi-Agent Systems II*, volume 11422, chapter 2, pages 21–35. Springer, May 2019.
- [37] G. Pigozzi, A. Tsoukiàs and P. Viappiani, Preferences in artificial intelligence, *Annals of Mathematics and Artificial Intelligence* 77(3–4) (2016), 361–401.
- [38] H. Prakken and M. Sergot, Dyadic deontic logic and contraryto- duty obligations. In *Defeasible deontic logic*, pages 223–262. Springer Netherlands, Dordrecht, 1997.
- [39] F. Ricci, L. Rokach and B. Shapira, Introduction to recommender systems handbook. In *Recommender Systems Handbook*, pages 1–35. Springer US, Boston, MA, 2011.
- [40] F. Rossi and A. Loreggia, Preferences and ethical priorities: Thinking fast and slow in AI. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pages 3–4, 2019.
- [41] F. Rossi and N. Mattei, Building ethically bounded AI. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 9785–9789, Honolulu, Hawaii, USA, 2019. AAAI Press.
- [42] F. Rossi, K.B. Venable and T. Walsh, A short introduction to preferences: between artificial intelligence and social choice, *Synthesis Lectures on Artificial Intelligence and Machine Learning* 5(4) (2011), 1–102.
- [43] A. Sen and S. Körner, Choice, ordering and morality. In *Practical Reason*. Blackwell, Oxford, 1974.
- [44] Y. Shoham and K. Leyton-Brown, *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, New York, NY, USA, 2008.
- [45] J. van Benthem, D. Grossi and F. Liu, On the two faces of deontics: Semantic bnterness and syntactic priority. Technical Report, 2011.
- [46] L. van der Torre and Y.-H. Tan, Contrary-to-duty reasoning with preference-based dyadic obligations, *Annals of Mathematics and Artificial Intelligence* 27(1–4) (1999), 49–78.
- [47] L.W. van der Torre, *Reasoning about Obligations: Defeasibility in Preference-Based Deontic Logic*. Ph.d. thesis, Erasmus University of Rotterdam. Tinbergen Institute

- Research Series n° 140. Thesis Publishers: Amsterdam., 1997.
- [48] W. Von and H. Georg, *Norm and Action: A Logical Inquiry*. Routledge & Kegan Paul London, London, 1963.
- [49] G.H. Von Wright. The logic of preference, *Studia Logica* **30** (1963), 159–162.
- [50] G.H. Von Wright, A new system of deontic logic. In *Deontic Logic: Introductory and Systematic Readings*, pages 105–120. Springer, 1970.
- [51] H.Wang, S. Shao, X. Zhou, C. Wan and A. Bouguet-taya, Preference recommendation for personalized search, *Knowledge-Based Systems* **100** (2016), 124–136.
- [52] H. Werbin-Ofir, L. Dery and E. Shmueli, Beyond majority: Label ranking ensembles based on voting rules, *Expert Systems with Applications* **136** (2019), 50–61.
- [53] E. Zitelmann, *Lücken im Recht*. Duncker-Humblot, Leipzig, 1903.

CORRECTED PROOF