Desires, norms and constraints

Laurence Cholvy
ONERA Toulouse
2 av. édouard Belin
31055 Toulouse Cedex 4
cholvy@cert.fr

Christophe Garion
SUPAERO
10 av. édouard Belin
31055 Toulouse Cedex 4
garion@supaero.fr

Abstract

This paper deals with modeling mental states of a rational agent, in particular states based on agent's desires.

It shows that the world the agent belongs to forces it to restrict its desires. More precisely, desires of a rational agent are restricted by the constraints that exist in the world and which express what is possible or necessary. Furthermore, if the agent is law-abiding, its desires are restricted by the regulations that are defined in the world and which express what is obligatory, permitted or forbidden.

This paper characterizes how desires are restricted depending on the fact that the agent is law-abiding or not.

This work considers the general case when the agent orders its own desires according to a preference order.

The solution is based on modeling desires, regulations and constraints in an unique formal system which is a logic of conditional preferences.

1. Introduction

The present work belongs to the domain of modeling mental states of a rational agent. More precisely, we aim at modeling the desires of an agent who belongs to a multi-agent context and determining among its desires those which can be chosen to be achieved. We intend to show that the agent's desires are restricted by some external elements. More precisely, if the agent is rational, its desires are restricted by domain constraints; furthermore, if it is law-abiding, then its desires are restricted by norms.

Domain constraints, for instance laws of the nature, express situations that are possible (or in other terms, discard situations that are impossible). The impact of constraints on the desires of a rational agent is obvious: this reflects the fact that a rational agent will restrict its desires to achievable ones. For instance, even if an agent desires to live more than 150 years, this will not be achievable.

On the other hands, norms are social elements which prescribe agents' behavior in any multi-agent context. They state what is permitted, obligatory or forbidden. For instance, paying taxes is obligatory.

The main difference between constraints and norms is that norms can be violated. Thus, an agent can be law-abiding or not.

In the literature, the most famous framework for modeling mental states of agents is the BDI model [16] which considers that Beliefs, Desires and Intentions are the basic elements for the agent when taking a decision about how to act on the world. *Desires* are internal motivational states expressing situations that the agent prefers and wants to bring about. *Intentions* are the deliberative states expressing desired and reachable situations that the agent has selected and committed to achieve.

Based on the idea that social concepts like *obligations* or more generally *norms* are important to "glue" autonomous agents in a Multi-agent System, the BDI model has recently been extended in order to take into account *obligations* and *norms* [5], [11].

Agreeing with these works, we show that norms may restrict desires of an agent if he is law-abiding. Furthermore, we add that domain constraints restrict desires of an agent if he is rational.

Notice that our model focuses on desires. In comparison to BDI model, it does not take into account beliefs nor intentions. This is mainly because this work was motivated by an application in Requirement Engineering in which considering these notions is not meaningful.

However, the originality of the present work is that we consider a case which is rather general, in which we allow the agent to order its own desires according to a preference order.

Notice that a similar problem is studied in [1] and addressed in the possibilistic logic framework. However, as far as we know, this work does not consider norms.

As said previously, the main application we foresee to this work belongs to Requirement Engineering. Indeed, in

Requirement Engineering, requirements express the properties an agent expects the artefact will satisfy [12], [14]. In other terms, requirements are, for the agent, its desires about the artefact to be build. As soon as the requirements are expressed, it is important to take into account the physical constraints that exist on the real world and which state what is possible. Besides, it is also important to take into account the regulations that exist in the domain and which state what is permitted or obligatory or forbidden. Indeed, before entering the design phase and the manufacturing phase, one must check if the properties expressed by the agent's requirements do not characterize an artefact which cannot be built (due to the constraints) or which will violate a regulation.

In this present work, we show that an unique formalism (Boutilier's logic of conditional preferences [3]) can be used to model desires of the agent, domain constraints and norms as well. Using a single formalism allows us to easily define the influence of constraints and norms on the restriction of desires.

This paper is organized as follows. Section2 presents the conditional preference logic CO. Sections 3, 4 and 5 show how desires, norms and constraints are modeled in CO. Section 6 presents the general mechanism according to which desires are restricted by constraints and norms. Section 7 applies this general mechanism to the cases when the agent is or is not law-abiding. Section 8 describes an example. Finally, section 9 is devoted to a discussion.

2. The CO logic

The CO logic is a logic developed by Craig Boutilier to represent and reason on conditional preferences [3, 2].

2.1. Semantics

Boutilier considers a bimodal propositional language L_B based on a set of atomic propositional variables $PROP$ with the usual connectives and two modal operators $\Box$ et $\overleftarrow{\Box}$.

CO's semantics is characterized by Kripke models of the form $\langle W, \leq, val\rangle$ where:

- W is a set of possible worlds.
- $\leq$ is a total *preference* preorder on W (a reflexive and transitive relation on W^2). If w and w' are two worlds of W, then $w \leq w'$ means that w is at least as preferred as w'.
- val is a valuation function on W[1]. For any formula φ of W, $val(\varphi)$ is the set of worlds of W which classically satisfy φ. As usually, we will denote $val(\varphi)$ by $\|\varphi\|$ (cf. [6]).

1 I.e. $val : PROP \to 2^W$ and val is such that $val(\neg\varphi) = W - val(\varphi)$ and $val(\varphi_1 \wedge \varphi_2) = val(\varphi_1) \cap val(\varphi_2)$.

For any CO-model $\mathcal{M} = \langle W, \leq, val\rangle$, the truth conditions for the modal connectives $\Box$ and $\overleftarrow{\Box}$ are:

- $\mathcal{M} \models_w \Box\varphi$ iff $\forall w' \in W$ such that $w' \leq_P w$ then $\mathcal{M} \models_{w'} \varphi$.
- $\mathcal{M} \models_w \overleftarrow{\Box}\, \varphi$ iff $\forall w' \in W$ such that $w' \not\leq_P w$ then $\mathcal{M} \models_{w'} \varphi$.

$\Box\varphi$ is true at a world w if and only if φ is true at all worlds at least as preferred as w (including w). $\overleftarrow{\Box}\, \varphi$ is true at world w if and only if φ is true at all the worlds less preferred than w. Boutilier then defines two dual modal operators : $\Diamond\varphi \equiv_{def} \neg\Box\neg\varphi$ means that φ is true at some equally or more preferred world and $\overleftarrow{\Diamond}\, \varphi \equiv_{def} \neg \overleftarrow{\Box}\, \neg\varphi$ means that φ is true at some less preferred world. $\overleftrightarrow{\Box}\, \varphi \equiv_{def} \Box\varphi \wedge \overleftarrow{\Box}\, \varphi$ and $\overleftrightarrow{\Diamond}\, \varphi \equiv_{def} \Diamond\varphi \vee \overleftarrow{\Diamond}\, \varphi$ correspond respectively to classical necessity and possibility (cf. [6]).

The validity of a formula φ is defined as follows: let $\mathcal{M} = \langle W, \leq, val\rangle$ be a CO-model. A formula φ is valid in $\mathcal{M}$ (noted $\mathcal{M} \models \varphi$) iff $\forall w \in W$ $\mathcal{M} \models_w \varphi$. φ is CO-valid (noted $\models_{CO} \varphi$) iff for any CO-model $\mathcal{M}$, $\mathcal{M} \models \varphi$. φ is satisfiable iff $\neg\varphi$ is not valid.

For instance, figure 1 presents CO-model $\mathcal{M}$ such that $\mathcal{M} \models \overleftrightarrow{\Box}\, \alpha$ (because all the worlds satisfy α) and $\mathcal{M} \models_{w_2} \Box\beta$ (because all the worlds at least as preferred as w_2 satisfy β).

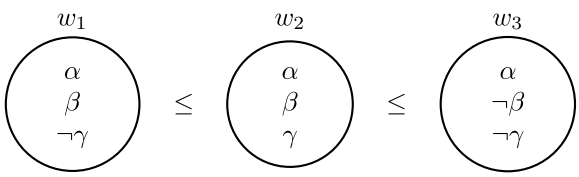

Figure 1. A CO-model

Finally, the consequence relation is defined by:

Definition 1 *Let $\Sigma = \{\varphi_1, \ldots, \varphi_n\}$ be a set of CO formulas. A formula ψ is a consequence of Σ, denoted by $\Sigma \models \psi$ iff for all CO-model $\mathcal{M}$, $\mathcal{M} \models \bigwedge_{i \in \{1,\ldots,n\}} \varphi_i \Rightarrow \mathcal{M} \models \psi$.*

2.2. Conditional preferences in CO

According to Boutilier, preferences are formulas of the form $I(\beta|\alpha)$ which signify that "ideally, if α is true, then β is true". Formally, the I operator is defined by:

$$I(\beta|\alpha) \equiv_{def} \overleftrightarrow{\Box}\, \neg\alpha \vee \overleftrightarrow{\Diamond}\, (\alpha \wedge (\Box\alpha \to \beta))$$

Thus, if we consider a CO-model $\mathcal{M}$, $I(\beta|\alpha)$ will be satisfied in $\mathcal{M}$ iff:

- either α is not true in every world of W;
- either there is a world w which satisfies α and such that all the worlds at least as preferred as w satisfy $\alpha \rightarrow \beta$.

$I(\beta)$ is defined as $I(\beta|\top)$. Moreover, the dual notion of *tolerance* is denoted by $T(\beta|\alpha) \equiv_{def} \neg I(\neg\beta|\alpha)$. Notice that there are other approaches to conditional preferences, based on ideality [15, 18] or for instance *ceteris paribus* [17, 4].

3. Desires representation

3.1. The notion of position

Starting from Cholvy and Hunter work on requirements representation [10], we consider that the desires of an agent are propositional formulas (of a given language $PROP$) on which the agent expresses a priority order. The definition of the agent's desires is then a tuple, called "agent position" defined as follows:

Definition 2 *The* position *of an agent is a tuple of propositional formulas of a given language $PROP$ $\Gamma = [\alpha_1, \ldots, \alpha_n]$ such that $\{\alpha_1, ...\alpha_n\}$ is consistent. Each α_i is a desire of the agent. Moreover, α_i precedes α_j in the tuple Γ iff the agent considers α_i to be more important than α_j.*

The position $[\alpha_1, \alpha_2]$ intuitively means that, for the agent, the most important desire is to bring it about that $\alpha_1 \wedge \alpha_2$ holds. But, if it is not possible (due to external elements), then its second more important desire is to bring it about that $\alpha_1 \wedge \neg\alpha_2$. If it still impossible, then its third more important desire is to bring it about that $\neg\alpha_1 \wedge \alpha_2$. Finally, the worst case for the agent is when the only possible desire is to bring it about that $\neg\alpha_1 \wedge \neg\alpha_2$, i.e. a formula which violates its primary desires.

More formally, Cholvy and Hunter have shown that a position $[\alpha_1, ..., \alpha_n]$ induces a preorder on possible worlds defined by the lexicographic order on the set $\{\| \ \beta_1 \wedge ... \wedge \beta_n \ \| \ : \ \beta_i \in \{\alpha_i, \neg\alpha_i\}\} - \emptyset$.

Example 1 *Consider an agent who desires to buy a house. He wants the house not to be close to a subway station (because it is noisy) or to be well soundproofed and it also wants the house to be near to a subway station if it is not downtown. The first part of the desire is more important for the agent than the second. Let us consider the propositional variables P (the house is soundproofed), S (the house is near to a subway station) and D (the house is downtown). We can express the agent's desires by the position: $\Gamma = [\neg S \vee P, \neg D \rightarrow S]$. This position corresponds to the following preorder on possible worlds:*

$\|(\neg S \vee P) \wedge (D \vee S)\| \leq \|(\neg S \vee P) \wedge (\neg D \wedge \neg S)\| \leq \|(S \wedge \neg P) \wedge (D \vee S)\| \leq \|(S \wedge \neg P) \wedge (\neg D \wedge \neg S)\|$

Thus, worlds satisfying $(\neg S \vee P) \wedge (D \vee S)$ are more preferred than the worlds satisfying $(\neg S \vee P) \wedge (\neg D \wedge \neg S)$, which are themselves more preferred than the worlds satisfying $(S \wedge \neg P) \wedge (D \vee S)$, which are themselves more preferred than the worlds satisfying $(S \wedge \neg P) \wedge (\neg D \wedge \neg S)$.

This means that the most preferred desire for the agent is to bring it about that its house satisfies $\neg S \vee P$ and $\neg D \rightarrow S$. If this is impossible, then the most important desire will be to bring it about that its house verifies $\neg S \vee P$ and not $\neg D \rightarrow S$. If it is not possible yet, then the most important desire will be to bring it about that the house satisfies $\neg D \rightarrow S$ and not $\neg S \vee P$. Finally, in the worst case, the remaining desire is to bring it about that the house does not satisfy neither $\neg D \rightarrow S$ nor $\neg S \vee P$ (notice that this is inconsistent here).

3.2. Representing desires in CO

In this section, our aim is to show that every position can be translated into a set of CO formulas. In order to do that, we first build the CO formulas associated to each "cluster" of worlds ordered by the preorder corresponding to a position.

Definition 3 *Let $\{\alpha_1, \ldots, \alpha_n\}$ be a consistent set of propositional formulas. Let $f_{[\alpha_1, \ldots, \alpha_n]} : N \rightarrow PROP$ be the function which makes correspond each integer $i \in \{0, \ldots, 2^n - 1\}$ to the propositional formula $\alpha_1(i) \wedge \ldots \wedge \alpha_n(i)$ in the following way:*

- *i is broken down into $i = \sum_{k=0}^{n-1} c_k(i) * 2^{n-1-k}$ with $\forall k \in \{0, \ldots, n-1\}$ $c_k(i) = 0$ or $c_k(i) = 1$;*
- $\forall k \in \{1, \ldots, n\}$ $\alpha_k(i) = \begin{cases} \alpha_k & \text{if } c_{k-1}(i) = 0 \\ \neg\alpha_k & \text{if } c_{k-1}(i) = 1 \end{cases}$

Example 2 *In the case of example 1, the result is:*

$$\begin{aligned} f_\Gamma(0) &= (P \vee \neg S) \wedge (S \vee D) \\ f_\Gamma(1) &= (P \vee \neg S) \wedge \neg(S \vee D) \equiv \neg S \wedge \neg D \wedge P \\ f_\Gamma(2) &= \neg(P \vee \neg S) \wedge (S \vee D) \equiv \neg P \wedge S \wedge D \\ f_\Gamma(3) &= \neg(P \vee \neg S) \wedge \neg(S \vee D) \equiv \bot \end{aligned}$$

We now restrict function f in order to obtain only satisfiable propositions:

Definition 4 *Let $\{i_0, \ldots, i_m\} \subset \{0, \ldots, 2^n - 1\}$ be the set of integers such that $\forall j \in \{i_0, \ldots, i_m\}$ $f_\Gamma(j)$ is satisfiable. f'_Γ is such that $\forall j \in \{0, \ldots, m\}$ $f'_\Gamma(j) = f_\Gamma(i_j)$*

Example 3 *Continuing 2, we obtain:*

$$\begin{aligned} f'_\Gamma(0) &= (P \vee \neg S) \wedge (S \vee D) \\ f'_\Gamma(1) &= (P \vee \neg S) \wedge \neg(S \vee D) \equiv \neg S \wedge \neg D \wedge P \\ f'_\Gamma(2) &= \neg(P \vee \neg S) \wedge (S \vee D) \equiv \neg P \wedge S \wedge D \end{aligned}$$

Let us note that function f' allows to find the "clusters" of ordered worlds corresponding to the preorder build from the agent's position, i.e. $\|f'_\Gamma(0)\| \leq \|f'_\Gamma(1)\| \leq \|f'_\Gamma(2)\|$.

We now build a set of CO formulas from the set of formulas defined by f'. The models of those CO formulas will correspond to the previous preorder.

Definition 5 *We note :*
$\alpha <_E \beta \equiv_{def} \overleftrightarrow{\Diamond} (\alpha \wedge \overleftarrow{\Box} \neg\alpha \wedge \Box\neg\beta)$

Definition 6 *Let $\Gamma = [\alpha_1, \ldots, \alpha_n]$ be the agent's position. This position is represented by the following set of CO formulas noted Γ^{CO}:*

$$\Gamma^{CO} = \bigcup_{i \in \{0,\ldots,m-1\}} \{f'_\Gamma(i) <_E f'_\Gamma(i+1))\}$$

Definition 7 *We say that $\alpha <_\Gamma \beta$ iff $\Gamma^{CO} \models \alpha <_E \beta$.*

Theorem 1 *$<_\Gamma$ is an order on $PROP$ formulas.*

Proof 1 *There are three properties to verify:*

- *$<_\Gamma$ is an irreflexive relation*

  The proof is easy: let us suppose that $<_\Gamma$ is not an irreflexive relation, then there are two propositional formulas α and β such that $\alpha <_\Gamma \beta$ and $\alpha \equiv \beta$ (in the logical sense). Thus $\Gamma^{CO} \models \overleftrightarrow{\Diamond} (\alpha \wedge \overleftarrow{\Box} \neg\alpha \wedge \Box\neg\beta)$. As $\alpha \equiv \beta$, $\Gamma^{CO} \models \overleftrightarrow{\Diamond} (\alpha \wedge \overleftarrow{\Box} \neg\alpha \wedge \Box\neg\alpha)$.

  Let $M = \langle W, \leq, val\rangle$ be a model of Γ^{CO}. There is a world w_0 of W such that $M, w_0 \models \alpha \wedge \Box\neg\alpha$, so $M, w_0 \models \alpha \wedge \neg\alpha$ which is impossible.

- *$<_\Gamma$ is an antisymmetric relation*

  Let us suppose that $<_\Gamma$ is a symmetric relation, then there are two propositional formulas α and β such that $\Gamma^{CO} \models \alpha < \beta$ and $\Gamma \models \beta < \alpha$. Let $M = \langle W, \leq, val\rangle$ be a model of Γ^{CO}. In this case:

  1. *$\exists w_\alpha$ such that $M, w_\alpha \models \alpha \wedge \Box\neg\beta$*
  2. *$\exists w_\beta$ such that $M, w_\beta \models \beta \wedge \Box\neg\alpha$*

  Let us suppose that $w_\alpha \leq w_\beta$, then in this case, $M, w_\alpha \models \alpha \wedge \neg\alpha$, which is impossible. The proof is the same for the case where $w_\beta \leq w_\alpha$.

- *$<_\Gamma$ is a transitive relation*

  Let α, β and γ three propositional formulas such that $\Gamma^{CO} \models \alpha < \beta$ and $\Gamma^{CO} \models \beta < \gamma$. Thus $\alpha <_\Gamma \beta$ and $\beta <_\Gamma \gamma$.

  Let $M = \langle W, \leq, val\rangle$ be a model of Γ. Then :

  1. *$\exists w_\alpha$ such that $M, w_\alpha \models \alpha \wedge \Box\neg\beta \wedge \overleftarrow{\Box} \neg\alpha$*
  2. *$\exists w_\beta$ such that $M, w_\beta \models \beta \wedge \Box\neg\gamma \wedge \overleftarrow{\Box} \neg\beta$*

  Let us suppose that $w_\beta \leq w_\alpha$, then $M, w_\beta \models \beta \wedge \neg\beta$ because $M, w_\alpha \models \Box\neg\beta$, so $w_\alpha \leq w_\beta$.

  As $M, w_\beta \models \Box\neg\gamma$, $M, w_\alpha \models \Box\neg\gamma$. Thus $M, w_\alpha \models \alpha \wedge \Box\neg\gamma \wedge \overleftarrow{\Box} \neg\alpha$, and $\Gamma^{CO} \models \alpha < \gamma$.

Theorem 2 *Let $\Gamma = [\alpha_1, \ldots, \alpha_n]$ be the agent's position. The CO-models $\langle W, \leq, val\rangle$ of Γ^{CO} are such that: $\|f'_\Gamma(0)\| \leq \ldots \leq \|f'_\Gamma(m)\|$.*

Proof 2 *Let $\Gamma = [\alpha_1, \ldots, \alpha_n]$ be the agent's position. Let $M = \langle W, \leq, val\rangle$ be a model of Γ^{CO}. Let $j \in \{0, \ldots, m-1\}$, we denote by w_j the world of W such that $M, w_j \models f'_\Gamma(j) \wedge \overleftarrow{\Box} \neg f'_\Gamma(j) \wedge \Box\neg f'_\Gamma(j+1)$.*

First, let us prove that $\forall j \in \{1, \ldots, m-1\}$ $w_{j-1} \leq w_j$. Let $j \in \{0, \ldots, m-1\}$, then $M, w_{j-1} \models \Box\neg f'_\Gamma(j)$. But $M, w_j \models f'_\Gamma(j)$ so $w_{j-1} \leq w_j$.

Let $j \in \{0, \ldots, m_a - 1\}$, there are two cases:

- *either $j \neq 0$ ans in this case $M, w_j \models f'_\Gamma(j) \wedge \overleftarrow{\Box} \neg f'_\Gamma(j)$. Thus $\forall w \in \|f'_\Gamma(j)\|$ $w \leq w_j$.*

  Moreover, $M, w_{j-1} \models \Box\neg f'_\Gamma(j)$, so $\forall w \in \|f'_\Gamma(j)\|$ $w_{j-1} \leq w$.

  As $\forall j \in \{1, \ldots, m-1\}$ $w_{j-1} \leq w_j$, we can write that $M, w_{j-1} \models \bigwedge_{l \in \{0,\ldots,j-1\}} \overleftarrow{\Box} \neg f'_\Gamma(l)$ and that $M, w_j \models \bigwedge_{l \in \{j+1,\ldots,m\}} \Box\neg f'_\Gamma(l)$. Thus, for all $w \in W$, if $w \leq w_j$ and $w_{j-1} \leq w$, $M, w \models \bigwedge_{\substack{l \in \{0,\ldots,m\} \\ l \neq j}} \neg f'_{\Gamma_a}(l)$.

  Let $w \in W$ such that $w \leq w_j$ and $w_{j-1} \leq w$. If for all $l \in \{0, \ldots, m\}$, we write $f'_\Gamma(l) = \alpha_1(l) \wedge \ldots \wedge \alpha_n(l)$, then $\forall l \in \{1, \ldots, m\}$ $l \neq j \Rightarrow M, w \models \neg\alpha_1(l) \vee \ldots \vee \neg\alpha_n(l)$.

  By building of f'_Γ, $\forall l \in \{1, \ldots, m\}$ $l \neq j \Rightarrow \exists l_j \in \{1, \ldots n\}$ such that $\alpha_{l_j}(l) \equiv \neg\alpha_{l_j}(j)$.

  As $\forall l \in \{1, \ldots, m\}$ $l \neq j \Rightarrow M, w \models \neg\alpha_1(l) \vee \ldots \vee \alpha_n(l)$, $M, w \models \alpha_1(j) \wedge \ldots \wedge \alpha_n(j)$. Thus $M, w \models f'_\Gamma(j)$.

  The models of $f'_\Gamma(j)$ are then the worlds $w \in W$ such that $w \leq w_j$ and $w_{j-1} \leq w$.

- *let $j = 0$, $M, w_0 \models f'_\Gamma(0) \wedge \overleftarrow{\Box} \neg f'_\Gamma(0)$ so $\forall w \in \|f'_\Gamma(0)\|$, $w \leq w_0$.*

  Moreover, we can write that $M, w_0 \models \bigwedge_{l \in \{1,\ldots,m\}} \Box\neg f'_\Gamma(l)$ from the previous proof, $\forall w \in W$, $w \leq w_0 \Rightarrow M, w \models f'_\Gamma(0)$.

Finally, $M, w_{m-1} \models \Box\neg f'_\Gamma(m)$ thus $\forall w \in \|f'_\Gamma(m)\|$ $w_{m-1} \leq w$. Moreover, $M, w_{m-1} \models \bigwedge_{l \in \{0,\ldots,m-1\}} \overleftarrow{\Box} \neg f'_\Gamma(l)$, so $\forall w \in W$ $w_{m-1} \leq w \Rightarrow M, w \models f'_\Gamma(m)$.

Notice that the coding of the agent's position in CO allows to retrieve the same preorder on possible worlds as Cholvy and Hunter.

Example 4 *Let us resume 1: $\Gamma = [\neg M \vee I, \neg C \rightarrow M]$. Thus $\Gamma^{CO} = \{(I \vee \neg M) \wedge (M \vee C) < \neg M \wedge \neg C \wedge I, \neg M \wedge \neg C \wedge I < \neg I \wedge M \wedge C\}$.*

As a conclusion, we have shown in this section that every ordered set of desires can be modeled into a set of CO formulas. This set of formulas is such that its models respects the intuitive idea we can give on ordered worlds.

In the next section, we take an interest in norms representation.

4. Representing norms with CO

We showed in [7, 8, 9, 13] how to model normative sentences with CO formulas. Those sentences can be simple obligations, permissions, prohibitions (like for instance, *it is forbidden to build a house in a non permitted area* , *it is allowed to paint shutters green*), but also complex normative sentences like norms with exceptions (like for instance *it is forbidden to paint shutters green, unless the house is not near to an historical building*) or Contrary-to-Duties (like for instance *it is forbidden to paint shutters colored, but if shutters are paint colored, it should be a light shade*).

In the following table, we list the translations of normative sentences into CO formulas. We do not detail why we choose such translations nor the remaining representation problems (cf. previously cited papers).

it is obligatory that α is true	$I(\alpha)$
it is allowed that α is true	$T(\alpha) \equiv \neg I(\neg\alpha)$
it is forbidden that α is true	$I(\neg\alpha)$
it is normally forbidden that α bu if β is true, then α is allowed	$I(\neg\alpha) \wedge \neg I(\neg\alpha\|\beta)$
it is normally forbidden that α but if β is true then α is obligatory	$I(\neg\alpha) \wedge I(\alpha\|\beta)$
Contrary-To-Duty : (RP) it is forbidden that α is true (CTD) but if α is true, then it is obligatory that β is true	 $I(\neg\alpha)$ $I(\beta\|\alpha)$

A *regulation* is a set of normative sentences or norms, of the previous form.

5. Representing domain constraints with CO

Domain constraints are for instance physical constraints. They represent what is necessarily true in the real world. Expressing domain constraints in the CO logic is easy, because it amounts to restrict the set of possible worlds to worlds verifying the constraints. This is formally expressed by the following definitions.

Definition 8 *A domain constraint of the form "proposition α is always true" is represented by the CO formula $\overleftrightarrow{\Box}\alpha$.*

Thus, the constraints "downtown, houses always cost more than 100,000 euros" can be modeled into the formula $\overleftrightarrow{\Box}(D \rightarrow \neg L)$ (D represents the fact that a house is downtown and L the fact that it costs less than 100,000 euros). The CO-models of this formula are such that the world $\{D, L\}$ is not a possible world.

Definition 9 *A set of domain constraints $\mathcal{C}$ is a consistent set of formulas of the form $\overleftrightarrow{\Box}\alpha$.*

6. Restriction of desires by taking norms and constraints into account

In this section, we study the impact of norms and constraints on the restriction of the agent's desires.

First of all, we have to check that the set of norms is compatible with the constraints. This means intuitively that the regulation represented by the set of norms does not obliges something that is impossible or forbid something that is necessarily true.

6.1. Compatibility between norms and domain constraints

Definition 10 *Let $\mathcal{R}$ be a set of normative sentences. $\mathcal{R}$ is compatible with the domain constraint $\overleftrightarrow{\Box}\varphi$ iff $\forall\psi$ such that $\psi \wedge \varphi$ is satisfiable, then $\mathcal{R} \models \neg I(\neg\varphi|\psi)$. $\mathcal{R}$ is compatible with the set of domain constraints $\mathcal{C} = \{\overleftrightarrow{\Box}\varphi_1, \ldots, \overleftrightarrow{\Box}\varphi_l\}$ iff $\mathcal{R}$ is compatible with $\overleftrightarrow{\Box}(\varphi_1 \wedge \ldots \wedge \varphi_l)$.*

Example 5 *Let us consider the following regulation : every house must be connected to the electrical network, but if a house is not connected, then it must be equipped with a generator. This regulation is modeled by the set $\{I(elec), I(gen|\neg elec)\}$. Let us suppose that the domain constraints are such that it is impossible for the house to be equipped with a generator. In this case, the regulation is not compatible with the domain constraints. Indeed, there are situations (when the house is not connected to the electrical network) in which it is obligatory to equip the house with a generator although it is impossible.*

6.2. Tolerated states

Definition 11 *Let $\mathcal{R}$ be a regulation and $\mathcal{C}$ a set of domain constraints. We note:*

$$T(\mathcal{R}, \mathcal{C}) = \bigvee_{R \models \neg I(\neg\varphi) \text{ and } \models \overleftrightarrow{\Diamond}\varphi} \varphi$$

$T(\mathcal{R}, \mathcal{C})$ is a formula whose models are the states tolerated by the regulation $\mathcal{R}$ under the constraints $\mathcal{C}$.

The following result proves that the set of states tolerated by a regulation which is compatible with the domain constraints cannot be empty.

Theorem 3 *Let $\mathcal{R}$ be a regulation compatible with a set of domain constraints $\mathcal{C}$. $\parallel T(\mathcal{R},\mathcal{C}) \parallel \neq \emptyset$*

In order to prove this theorem, just remark that if $\mathcal{R}$ is compatible with $\mathcal{C} = \{\overleftrightarrow{\Box}\ \varphi_1, \ldots, \overleftrightarrow{\Box}\ \varphi_l\}$, then $\varphi_1 \wedge \ldots \wedge \varphi_l$ is a conjunction of $T(\mathcal{R},\mathcal{C})$.

6.3. Restriction of desires

Definition 12 *Let Γ^{CO} be the CO formulas expressing the agent's desires. Let $\mathcal{R}$ be a regulation and $\mathcal{C}$ a set of domain constraints such that $\mathcal{R}$ is compatible with $\mathcal{C}$. The restriction of the agent's desires taking $\mathcal{R}$ and $\mathcal{C}$ into account is denoted Γ^{CO}_{min}. It is defined by:*
$\Gamma^{CO}_{min} = \{\varphi : \Gamma^{CO} \models I(\varphi | T(\mathcal{R},\mathcal{C}))\}$

The restriction of desires by taking $\mathcal{R}$ and $\mathcal{C}$ into account is the set of the formulas which are true in the minimal worlds for the preorder defined by Γ^{CO} given the tolerated states.

Theorem 4 *There are formulas which are not tautologies and which belong to Γ^{CO}_{min}*

Proof 3 *Let us suppose that Γ^{CO}_{min} contains only tautologies.*

Let $\mathcal{M} = \langle W, \leq, val\rangle$ be a model of Γ^{CO}. There are two cases:

- *either $\mathcal{M} \models \neg T(\mathcal{R},\mathcal{C})$ and then $\mathcal{M} \models I(\varphi|T(\mathcal{R},\mathcal{C}))$ for every formula φ (notice that this case never happens, because $\bigvee_{i\in\{1,\ldots,m\}} f'(i) \equiv \top$ and $T(\mathcal{R},\mathcal{C}) \neq \phi$).*
- *either $\exists (w_o, \ldots, w_l) \in W^n$ such that $\forall i \in \{1,\ldots,l\}$ $\mathcal{M}, w_i \models T(\mathcal{R},\mathcal{C})$.*

  Let $J \subseteq \{1,\ldots,l\}$ such that $\forall i \in J\ \forall j \in \{1,\ldots,l\}\ w_i \leq w_j$ (remember that by definition of Γ^{CO}, all the worlds of W are comparable). In this case, as $\mathcal{M}$ is a model of Γ^{CO}, there is $i_0 \in \{1,\ldots,m\}$ such that $\forall j \in J\ w_j \in \|f'(i_0)\|$ (by definition of Γ^{CO}).

  Moreover, by definition of Γ^{CO} again, considering any model $\mathcal{M}$ of Γ^{CO}, i_0 is unique (i.e. the best $T(\mathcal{R},\mathcal{C})$ are always in the same "cluster" of worlds).

  As all the worlds of $\|f'(i_0)\|$ verify $f'(i_0)$, then $\Gamma^{CO} \models I(f'(i_0)|T(\mathcal{R},\mathcal{C}))$.

Thus, there is always a set of restricted desires which are compatible with the regulation and the domain constraints. This is because the preorder induced by the agent's position considers all the worlds, even those which contradict all the agent's desires.

7. Attitudes of the agent towards the regulation

As said in the introduction, an agent can have two attitudes towards the law: whether he does not care about obeying the law or he does.

In the first case, the previous general process is used to restrict the agent's desires, by considering an empty regulation, i.e. $\mathcal{R} = \phi$.

In the case of an law-abiding agent, the previous general process is used to restrict the agent's desires with a non empty regulation, i.e. $\mathcal{R} \neq \phi$.

Notice also that this approach is flexible. We could model an agent who wants to obey only a part of the regulation.

8. Example

We consider an agent which wants to build a house. Its desires concerning its future house are:

- *it desires that its house is downtown or near to a subway station,*
- *it desires that the house has white walls,*
- *it desires that the house costs less than 100,000 euros and is situated in a quiet area.*

The position of the agent respects the order given in the previous enumeration.

The city the agent wants to build its house in is such that *the area close to a subway station is not quiet*. Moreover, *houses situated downtown always cost more than 100,000 euros.*

Finally, the city regulation stipulates that *downtown, it is forbidden to paint walls white.*

We consider the following propositional variables: D (the house is downtown), S (the house is near a subway station), W (the walls of the house are white), L (the house costs less than 100,000 euros), Q (the house is a quiet area).

The agent's position, the set of domain constraints and the regulation are respectively:

$$\Gamma = [D \vee S, W, L \wedge Q]$$

$$\mathcal{C} = \{\overleftrightarrow{\Box}(S \rightarrow \neg Q), \overleftrightarrow{\Box}(D \rightarrow \neg L)\}$$

$$\mathcal{R} = \{I(\neg W|D)\}$$

Let us remark that $\mathcal{C}$ is a consistent set of CO formulas and that $\mathcal{R}$ is compatible with $\mathcal{C}$.

The set of CO formulas which models the agent's position is:

$\Gamma^{CO} =$
$\{(D \vee S) \wedge W \wedge (L \wedge Q)\ <\ (D \vee S) \wedge W \wedge (\neg L \vee \neg Q),$
$(D \vee S) \wedge W \wedge (\neg L \vee \neg Q)\ <\ (D \vee S) \wedge \neg W \wedge (L \wedge Q),$

$(D \vee S) \wedge \neg W \wedge (L \wedge Q) \ < \ (D \vee S) \wedge \neg W \wedge (\neg L \vee \neg Q),$
$(D \vee S) \wedge \neg W \wedge (\neg L \vee \neg Q) \ < \ (\neg D \wedge \neg S) \wedge W \wedge (L \wedge Q),$
$(\neg D \wedge \neg S) \wedge W \wedge (L \wedge Q) \ < \ (\neg D \wedge \neg S) \wedge W \wedge (\neg L \vee \neg Q),$
$(\neg D \wedge \neg S) \wedge W \wedge (\neg L \vee \neg Q) \ < \ (\neg D \wedge \neg S) \wedge \neg W \wedge (L \wedge Q),$
$(\neg D \wedge \neg S) \wedge \neg W \wedge (L \wedge Q) \ < \ (\neg D \wedge \neg S) \wedge \neg W \wedge (\neg L \vee \neg Q)\}$

1. case of a law-abiding agent
   - Let us consider the most preferred formula, that is to say $(D \vee S) \wedge W \wedge (L \wedge Q)$. We can verify that its models are not tolerated states, because there are not states compatible with the domain constraints. Indeed, $((D \vee S) \wedge (L \wedge Q)) \wedge ((S \rightarrow \neg Q) \wedge (D \rightarrow \neg L))$ is not satisfiable.
   - Let us now look at the models of the formula just less preferred, i.e. $(D \vee S) \wedge W \wedge (\neg L \vee \neg Q)$.

     Some of its models are not tolerated states. Particularly, the worlds satisfying D are not tolerated because they also verify W (yet regulation forbids $D \wedge W$). The models of $(D \vee S) \wedge W \wedge (\neg L \vee \neg Q)$ which verify $\neg D$ also satisfy S. However, among those models, only those which verify $\neg Q$ are tolerated states (because of the first constraint). Thus, the two models of $(D \vee S) \wedge W \wedge (\neg L \vee \neg Q)$ which are tolerated are: $\{\neg D, S, W, \neg T, Q\}$ and $\{\neg D, S, W, \neg T, \neg Q\}$.

     The formula whose models are $\{\neg D, S, W, \neg T, Q\}$ and $\{\neg D, S, W, \neg T, \neg Q\}$ is $S \wedge \neg D \wedge W \wedge \neg Q$.

   Thus finally, $\Gamma^{CO}_{min} = \{\neg D \wedge S \wedge \neg Q \wedge W\}$.

   The agent's restricted desires are thus: *the house is not downtown; however it is close to a subway station and thus not in a quiet area; finally, it will have white walls*

2. case of a non law-abiding agent

   In the case when the agent does not take into account the regulation, its restricted desires are:
   - *the house is situated downtown, thus it is not less than 100,000 euros; it is near a subway station, thus not in a quiet area; it has white walls*, or
   - *the house is situated downtown, thus it is not less than 100,000 euros; it is not close to a subway station; it has white walls*, or
   - *the house is not situated downtown, but it is near a subway station, thus not in a quiet area; it has white walls*.

9. Discussion

This work intends to show that the desires of an agent may be restricted by external elements such as norms and constraints. This restriction is a consequence of the attitude of the agent. Any rational agent must take into account constraints in order to restrict its own desires. Furthermore, law-abiding agents must also take into account regulations.

Since we allow the agent to express a preference order on its desire, we guarantee that there will always be a set of restricted desires, if the regulation is, of course, compatible with the constraints.

Modeling desires, norms and constraints in an unique formalism allows us to provide a simple semantical characterization of the restricted desires.

However, this work suffers from a too simple agency model. This is why some results can be discussed. For instance, in the example detailed in section 8 (case of a law-abiding agent), the fact that *the house is not in a quiet area* belongs to the set of restricted desires can be discussed.

The main reason for discussing this is that the agent will have no influence on the fact that the area is quiet or not, once he will decide that the house is downtown. This is due to the constraints.

Thus, we suggest to modify definition 12 as follows:
$\Gamma^{CO}_{min} = \{\varphi : \Gamma^{CO} \models I(\varphi | T(\mathcal{R}, \mathcal{C}))$
and φ is controllable by the agent $\}$

The notion of controllable variables could be based on Boutilier's one [3] by also taking constraints into account.

This extension is currently foreseen.

Another interesting extension of this work is to consider the case of several agents, each of them having its own desires. The problem will then be to merge the agents' desires and characterize the restricted desires of the group of agents, by taking into account constraints and norms.

References

[1] S. Benferhat, D. Dubois, S. Kaci, and H. Prade. Bipolar representation and fusion of preferences on the possibilistic logic framework. In D. Fensel, F. Giunchiglia, D. L. McGuinness, and M.-A. Williams, editors, *Proceedings of the eight International Conference on Principles of Knowledge Representation and Reasoning (KR'02)*, pages 421–448. Morgan Kaufmann, 2002.

[2] C. Boutilier. Conditional logics of normality : a modal approach. *Artificial Intelligence*, 68(1):87–154, 1994.

[3] C. Boutilier. Toward a logic for qualitative decision theory. In J. Doyle, E. Sandewall, and P. Torasso, editors, *Principles of Knowledge Representation and Reasoning (KR'94)*, pages 75–86. Morgan Kaufmann, 1994.

[4] C. Boutilier, R. Brafman, H. Hoos, and D. Poole. Reasoning with conditional ceteris paribus preference statements. In *Proceedings of the fifteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI'99)*, pages 71–80, 1999.

[5] J. Broersen, M. Dastani, and L. van der Torre. Resolving conflicts between beliefs, obligations, intentions and desires. In S. Benferhat and P. Besnard, editors, *Proceedings of the*

6th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty ECSQARU 2001, volume 2143 of *Lecture Notes in Computer Science*, pages 568–579. Springer, 2001.

[6] B. Chellas. *Modal logic. An introduction*. Cambridge University Press, 1980.

[7] L. Cholvy and C. Garion. An attempt to adapt a logic of conditional preferences for reasoning with contrary-to-duties. *Fundamenta Informaticae*, 48(2,3):183–204, November 2001.

[8] L. Cholvy and C. Garion. Collective obligations, commitments and individual obligations: a preliminary study. In J. Horty and A. Jones, editors, *Proceedings of the* 6^{th} *International Workshop on Deontic Logic In Computer Science (*Δ*EON'02)*, pages 55–71, Londres, May 2002.

[9] L. Cholvy and C. Garion. Deriving individual obligations from collective obligations. In J. S. Rosenschein, T. Sandholm, M. Wooldridge, and M. Yokoo, editors, *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 962–964. ACM Press, July 2003.

[10] L. Cholvy and A. Hunter. Merging requirements from different agents. *Knowledge Based Systems*, 16:113–126, 2003.

[11] F. Dignum, D. Kinny, and L. Sonenberg. From desires, obligations and norms to goals. *Cognitive Science Quartely*, 2(3-4):407–430, 2002.

[12] A. Finkelstein, D. Gabbay, A. Hunter, J. Kramer, and B. Nuseibeh. Inconsistency handling in multi-perspective specifications. In *IEEE Transactions on Software Engineering*, volume 20, pages 569–578, August 1994.

[13] C. Garion. *Apports de la logique mathématique en ingénierie des exigences*. PhD thesis, École Nationale Supérieure de l'Aéronautique et de l'Espace, 2002. In French.

[14] A. Hunter and B. Nuseibeh. Managing inconsistent specifications : Reasoning, analysis and action. *Transactions on Software Engineering and Methodology*, 7(4):335–367, October 1998.

[15] J. Lang. Conditional desires and utilities - an alternative logical approach to qualitative decision theory. In W. Wahlster, editor, *Proceedings of the Twelth European Conference on Artificial Intelligence ECAI 96*, pages 318–322. Jonh Wiley and Sons, Ltd., 1996.

[16] A. Rao and M. Georgeff. Modeling rational agents within a BDI architecture. In *Proceedings of the Second International Conference on Konwledge Representation and Reasoning (KR'91)*, pages 473–484. Morgan Kaufmann, 1991.

[17] S.-W. Tan and J. Pearl. Specification and evaluation of preferences under uncertainty. In J. Doyle and al, editors, *Proceedings of the 4th International Conference on Principles of Knowledge Representation and Reasoning (KR'94)*, pages 530–539, Bonn, Germany, 1994. Morgan Kaufmann.

[18] L. van der Torre and E. Weydert. Parameters for utilitarian desires in a qualitative decision theory. *Applied Intelligence*, 14:285–301, 2001.