Distribution of Goals Addressed to a Group of Agents

Laurence Cholvy
ONERA Centre de Toulouse
2 av. Édouard Belin
31055 Toulouse, France
cholvy@cert.fr

Christophe Garion
SUPAERO
10 av. Édouard Belin
31055 Toulouse, France
garion@supaero.fr

ABSTRACT

The problem investigated in this paper is the distribution of goals addressed to a group of rational agents. Those agents are characterized by their ability (i.e. what they can do), their knowledge about the world and their commitments.

The goals of the group are represented by conditional preferences. In order to deduce the actual goals of the group, we determine its ability using each agent's ability and we suppose that the agents share a common knowledge about the world. The individual goals of an agent are deduced using its ability, the knowledge it has about the world, its own commitments and the commitments of the other agents of the group.

Categories and Subject Descriptors

I.2.4 [**Knowledge Representation Formalisms and Methods**]: Modal Logic; I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems

General Terms

Theory

Keywords

modal logic, qualitative decision theory, multiagent systems, goal selection and theories of rational agency

1. INTRODUCTION

Reaching a complex goal often needs many agents. This group of agents depends on the goal's characteristics, like for instance its complexity, the time needed to achieve it etc. But it also depends on the agents' characteristics, like their abilities, their capacities, their desires etc.

In such a context, several problems arise: we may for instance wonder how to build an agents coalition when a single agent is not sufficient to achieve a given task [17]. We may also try to optimize the number of agents.

AAMAS'03, July 14–18, 2003, Melbourne, Australia.

In this paper, we address a different problem. It assumes that a goal, describing more or less preferred situations, is allocated to a given group of agents. It also assumes a model of agents, describing each agent of that group by its knowledge about the current situation, its abilities and its commitments. It defines a characterization of the most preferred situation which can be achieved by the group and it also defines a way to assign each agent of the group with its individual goals in order to reach that situation.

Let us note that this problem is not to determine the goals of the group given the goals of each agent (cf. [15]).

In the mono-agent case, the derivation of a rational agent's goals depends on notions like belief, preference, capacity... Two main approaches to reason with rational agents have emerged those last ten years: BDI approaches (cf. for instance [16, 20, 11, 18, 9]) and qualitative versions of classical decision theory [19].

The solution described here is based on the solution provided by Boutilier in [2], who addresses these questions in the case of a single agent in a qualitative decision theory framework. In his paper, Boutilier assumes a set of conditional preferences expressing a goal for a single agent. He then describes a way to define the actual goals of the agent, given what it knows (or more exactly, what it believes) and given what it controls.

Boutilier justifies the use of a logic of conditional preferences to express and reason about goals: *A goal is typically taken to be some proposition that we desire an agent to make true. [...] Unfortunately, goals are not always achievable. My robot's goal to bring me coffee may be thwarted by a broken coffee maker. [...] Furthermore, goals may be defeated for reasons other than inability. It is often natural to specify general goals, but list exceptional circumstances that make the goal less desirable than the alternatives. [...] Rather than a categorical distinction between desirable and undesirable situations, we will rank worlds according to their degree of preference. The most preferred worlds correspond to goal states in the classical sense. However, when such states are unreachable, a ranking on alternatives becomes necessary.*

Our aim is to extend Boutilier's formalism to the multi-agent case in order to determine the goals of each agent knowing the conditional preferences imposed to the group.

In this work, we take into account three features to model the agents: their beliefs about the current world, assuming that they share the same beliefs, their capacity to change the truth value of a proposition (we extend the controllability model proposed by Boutilier) and their commitments.

The goal distribution process described in the following

must be viewed as managed by a central authority which knows all those data. This authority must allocate to each agent some tasks which validate a part of the group's goals. Those tasks are constrained by the agent's capacities, but also by its commitments and the other agents' commitments. We do not study here the possible communication problems between the agents and particularly the problem of negotiation in a group of agents (cf. [14, 12]).

The notion of commitments, which does not appear in Boutilier's work, is modeled here with sets of propositional formulas, although it can be modeled with modal logic [18]. We choose this representation in order to focus on the goal distribution problem.

This paper is organized as follows. In section 2, we present the QDT logic and we present our extension to Boutilier's controllability notion. We then turn our interest to the goal derivation process for a single agent. In section 3, we extend those notions to the multi-agent case, by extending the notion of controllability and goals to a group of agents. In section 4, we define the notion of commitment for a single agent and we present how to determine the effective goals of an agent and some properties of our formalism. Section 5 illustrates the goal derivation process on an example. Finally, section 6 is devoted to a discussion.

2. QUALITATIVE DECISION THEORY: REFINEMENT OF BOUTILIER'S APPROACH

Qualitative Decision Theory (QDT) is a qualitative version of classical decision theory [19]. Decision theory is a formalism whose aim is to determine a rational agent's goals given its preferences and its knowledge about the world. In classical decision theory, preferences are represented by utility functions on possible outcomes of actions and knowledge about the world is represented by a probability distribution on possible worlds. Boutilier proposed in [2] to use a modal logic called QDT to represent and reason about those two notions. This is done in a qualitative way, whereas classical decision theory uses numerical probability and utility functions that might not be available.

2.1 The QDT logic

To represent preferences, Boutilier uses a modal logic called QDT based on his logic CO [1]. Boutilier considers a propositional quadrimodal language L_B based on a set of atomic propositional variables $PROP$ with the usual connectives and four modal operators $\Box_P$, $\overleftarrow{\Box}_P$, $\Box_N$ and $\overleftarrow{\Box}_N$. The semantics of QDT is based on Kripke models of the form $\mathcal{M} = \langle W, \leq_P, \leq_N, val\rangle$ where :

- W is a set of possible worlds.
- $\leq_P$ 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 $w \leq_P w'$ means that w is at least as preferred as w'.
- $\leq_N$ is a total *normality* preorder on W. If w and w' are two worlds of W, $w \leq_N w'$ means that w is at least as normal a situation 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.

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

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

The truth conditions for $\Box_N$ and $\overleftarrow{\Box}_N$ are the same substituting $\leq_P$ by $\leq_N$.

$\Box_P\varphi$ (resp. $\Box_N\varphi$) is true at a world w if and only if φ is true at all worlds at least as preferred (resp. normal) as w (including w). $\overleftarrow{\Box}\ \varphi$ (resp. $\overleftarrow{\Box}_N\ \varphi$) is true at world w if and only if φ is true at all the worlds less preferred (resp. normal) than w. Boutilier then defines two dual modal operators : $\Diamond_P\varphi \equiv_{def} \neg\Box_P\neg\varphi$ means that φ is true at some equally or more preferred world and $\overleftarrow{\Diamond}_P\ \varphi \equiv_{def} \neg\ \overleftarrow{\Box}_P\ \neg\varphi$ means that φ is true at some less preferred world. $\Diamond_N$ and $\overleftarrow{\Diamond}_N$ are defined in the same way. $\overleftrightarrow{\Box}_P\ \varphi \equiv_{def} \Box_P\varphi \wedge \overleftarrow{\Box}_P\ \varphi$ and $\overleftrightarrow{\Diamond}_P\ \varphi \equiv_{def} \Diamond_P\varphi \vee \overleftarrow{\Diamond}_P\ \varphi$ correspond respectively to classical necessity and possibility (cf. [4]). Notice that $\overleftrightarrow{\Box}_N$ and $\overleftrightarrow{\Diamond}_N$ defined in the same way are such that $\models \overleftrightarrow{\Box}_N\ \varphi \leftrightarrow \overleftrightarrow{\Box}_P\ \varphi$ and $\models \overleftrightarrow{\Diamond}_N\ \varphi \leftrightarrow \overleftrightarrow{\Diamond}_P\ \varphi$.

The validity of a formula φ is defined as follows: let $\mathcal{M} = \langle W, \leq_P, val\rangle$ 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 satisfaisable iff $\neg\varphi$ is not valid.

If Σ is a set of QDT formulas and φ a QDT formula, then $\Sigma \models \varphi$ means that for every QDT model $\mathcal{M}$ $\mathcal{M} \models \Sigma \Rightarrow \mathcal{M} \models \varphi$.

Conditional preferences are formulas of the form $I(\beta|\alpha)$ which means that "ideally, if α is true, then β is true". The connective $I(-|-)$ is defined by:

$$I(\beta|\alpha) \equiv_{def} \overleftrightarrow{\Box}_P\ \neg\alpha \vee \overleftrightarrow{\Diamond}_P\ (\alpha \wedge (\Box_P\alpha \rightarrow \beta))$$

Thus, if we consider a CO-model $\mathcal{M}$, $I(\beta|\alpha)$ will be valid in $\mathcal{M}$ iff either α is false at every world of W, either there is some world w which satisfies α and such that every world at least as preferred as w satisfies $\alpha \rightarrow \beta$.

An absolute preference β is expressed as $I(\beta|\top)$ and noted $I(\beta)$.

In the same way, Boutilier introduces a new normative conditional connective $\Rightarrow$ defined in the same way using $\overleftarrow{\Box}$ and $\overleftarrow{\Box}_N$:

$$\alpha \Rightarrow \beta \quad \equiv_{def} \quad \overleftrightarrow{\Box}_N\ \neg\alpha \vee \overleftrightarrow{\Diamond}_N\ (\alpha \wedge \Box_N(\alpha \rightarrow \beta))$$

$\alpha \Rightarrow \beta$ means that β is true at the most normal α-worlds (it can be viewed as a default rule).

2.2 World representation and default knowledge

Boutilier considers KB, a knowledge base, which represents (partially) the state of the actual world.

Definition 1. KB is a finite and consistent set of propositional formulas.

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

As in classical decision theory, Boutilier does not require goals to be based on certain beliefs in KB, but also on reasonable default conclusions. For instance, if the agent knows that the weather is cloudy, then it can reasonably deduces that the situation in which it is raining is more likely (or normal) than the situation in which it is sunny.

Given a set $\mathcal{N}$ of formulas of the form $\alpha \Rightarrow \beta$ and a knowledge base KB, Boutilier defines the default closure of KB :

Definition 2. $Cl(KB) = \{\varphi \in L_{PROP} : \mathcal{N} \models KB \Rightarrow \varphi\}$ where L_{PROP} is the propositional sublanguage based on $PROP$.

$Cl(KB)$ is the agent's set of default conclusions given some representation KB of the actual world. $Cl(KB)$ can be viewed as the closure by $\Rightarrow$ of KB.

2.3 Goals of an agent

Boutilier determines the goal of an agent by using its preferences and its knowledge about the world. More precisely, an agent ought to act as if its default knowledge $Cl(KB)$ were true (not as if only KB were true). Intuitively, the goals of an agent will be some propositions which are true in the most preferred situations.

2.3.1 Ideal goals

Ideal goals are a first approximation : they are the propositions which are true in the most preferred situations where $Cl(KB)$ is true.

Definition 3. Let $\mathcal{P}$ be a set of conditional preferences. An ideal goal derived from $\mathcal{P}$ is a proposition φ such that:

$$\mathcal{P} \models I(\varphi|Cl(KB))$$

Notice that the ideal goals do not provide what the agent should do, because this requires the introduction of capability notion.

Example 1. Let us consider a propositional language whose variables are l (the door is lacquered) and s (the door is sanded). Let $\mathcal{P} = \{I(l), I(\neg l|\neg s)\}$ be a set of preferences meaning that:

- the agent prefers the door to be lacquered.
- if the door is not sanded, the agent prefers it not to be lacquered.

Let us suppose also that $\mathcal{N} = \{sh \Rightarrow s\}$, i.e. that in the most normal situations, if there are shavings on the ground then the door is sanded.

In order to simplify the problem, as preferences do not concern sh, we will only consider the worlds which concern l and s to determine the agent's goals. The possible worlds are $w_1 = \{l, s\}$, $w_2 = \{\neg l, \neg s\}$, $w_3 = \{l, \neg s\}$, $w_4 = \{\neg l, s\}$.

The worlds w_2 and w_4 cannot be the most preferred worlds, because of $I(l)$. But, because of $I(\neg l|\neg s)$, w_3 cannot be the most preferred world. So in every model of $\mathcal{P}$, w_1 is the most preferred world and so $w_1 \leq_P w_2$, $w_1 \leq_P w_3$ and $w_1 \leq_P w_4$. More, $w_3 \leq_P w_2$ cannot hold, because $I(\neg l|\neg s) \in \mathcal{P}$. So $w_2 \leq_P w_3$. $I(l)$ and $I(\neg l|\neg s)$ are valid only in the following QDT-model[2]:

$\mathcal{M}_1$	w_1	$\leq_P$	w_2	$\leq_P$	w_3	$\leq_P$	w_4
$\mathcal{M}_2$	w_1	$\leq_P$	w_2	$\leq_P$	w_4	$\leq_P$	w_3
$\mathcal{M}_3$	w_1	$\leq_P$	w_2	$\leq_P$	w_4	$\leq_P$	w_3
$\mathcal{M}_4$	w_1	$\leq_P$	w_4	$\leq_P$	w_2	$\leq_P$	w_3

Let us suppose that $KB_1 = \{sh\}$ (there are shavings on the ground and this is known by the agent). Then $Cl(KB_1) = \{sh, s\}$ because $sh \Rightarrow s$. The ideal goals for the agent are the α such that $\forall \mathcal{M}\ \mathcal{M} \models I(\alpha|sh \wedge s)$. l is then the only ideal goal for the agent: as the door is normally sanded, the agent should lacquer it.

Now, if we consider that $KB_2 = \{\neg s, sh\}$ (the door is not sanded but there are shavings on the ground, which does not contradict the normality rule $sh \Rightarrow s$), only one ideal goal can be deduced: $\neg l$. As the door is not sanded, the agent should not lacquer it.

2.3.2 CK-goals

Ideal goals represent the fact that the agent should reach the best situation given a certain knowledge about the world. But this definition is too restrictive for two reasons:

- to determine the goals of an agent, only the elements of KB which are *fixed* should be used. If an agent can change the truth value of an atom in KB, then this atom should not be taken into account in the goal derivation process. In particular, as for any formula φ and any QDT-model $\mathcal{M}$ $\mathcal{M} \models I(\varphi|\varphi)$ holds, every atom in KB is an ideal goal for the agent.
- ideal goals represent some desired situations that the agent should reach. They do not represent what the agent should *do*. An agent may prefer that it rains for instance. It is intuitively correct to suppose that the agent does not have any control on rain. In this case, it seems correct to consider "it rains" as an ideal situation, but not as a goal for the agent.

In order to refine the goal notion, Boutilier introduces a simple model of action and ability to demonstrate its influence on conditional goals. He suggests partitioning the atomic propositions in two classes : $P = C \cup \overline{C}$, in which C is the set of atomic propositions that the agent can control (i.e. the agent can change the truth value of those propositions) and $\overline{C}$ is the set of atomic propositions that the agent cannot control.

For instance the atomic proposition representing the fact *the agent lacquers the door* can be considered as controllable. The atomic proposition representing the fact *it rains* can reasonably be considered as uncontrollable.

We think that Boutilier's ability model is too restrictive. For instance, if an agent controls s which represents the fact that the door is sanded, then it controls also $\neg s$. The agent which can sand the door can also "unsand" the door. This is a bit controversial.

To solve this problem, we extend Boutilier's model by partitioning the literals of the language. Intuitively, a literal l is controllable by the agent iff:

- if the current situation is such that l is false, then the agent can by one of its actions make l and only l true.
- if the current situation is such that l is true, then the agent can keep l true by one of its actions or by doing nothing.

[2]In fact, we only consider the QDT-model for which $\overrightarrow{\Box}\ \varphi$ holds for any satisfaisable proposition φ.

Definition 4. Let C be the set of literals which are controllable by the agent and $\overline{C}$ the set of literal which are uncontrollable by the agent.

We then extend this definition as follows[3]:

Definition 5. Let w and w' be two worlds of W. Let us note $w' - w = \{l \ : \ w' \models l, \ w \models \neg l$ and l is a literal$\}$. A proposition φ is:

- controllable iff $\forall w \in W \ (w \models \neg\varphi \exists w' \in W \ w' \models \varphi$ and $(w' - w) \subseteq C)$;
- influenceable iff $\exists w \in W \ (w \models \neg\varphi \exists w' \in W \ w' \models \varphi$ and $(w' - w) \subseteq C)$.

  In this case, φ is influenceable in w.
- uninfluenceable iff it is not influenceable.

A world $w \in W$ is a context for some influenceable proposition φ iff φ is influenceable in w or $w \models \varphi$.

The contexts of an influenceable proposition φ are the worlds in which either φ is false but the agent can change the valuations of some controllable literal to make φ true, or the worlds in which φ is already true.

A controllable proposition is a proposition φ the agent is able to make true from a situation in which φ is false by changing only the valuation of some controllable literals. If the current situation is such that φ is true, then the agent can keep this situation by the definition of controllable literals. Every world of W is a context for a controllable proposition.

An influenceable proposition is a proposition φ the agent can make true only from some initial situations. For instance, if $a \in C$ and if $b \in \overline{C}$ then $a \wedge b$ is influenceable, but not controllable. As b is not controllable, if b is false, then the agent cannot make $a \wedge b$ true, so $a \wedge b$ is not controllable. But if b is true and a is false, the agent can make $a \wedge b$ true so $a \wedge b$ is influenceable. The contexts of $a \wedge b$ are in this case $\{a, b\}$ and $\{\neg a, b\}$.

We can now redefine the notion of CK goal (Complete Knowledge goal) introduced by Boutilier in [2]. The CK goals of the agent will be determined from the set of propositions φ which are true in $Cl(KB)$ and such that $Cl(KB)$ is not a context for $\neg\varphi$. The truth value of those propositions cannot be changed by the agent's actions and they will remain true. Moreover, the CK goals will only be propositions φ for which $Cl(KB)$ is a context, because either φ is false in $Cl(KB)$ and the agent can change the truth value of φ, either φ is true in $Cl(KB)$ and the agent can keep the truth value of φ. We first define the non-contextual propositions of KB:

Definition 6. The set on non-contextual propositions of KB is defined by:

$$NC(KB) = \{\varphi \in Cl(KB) \ : \ Cl(KB) \text{ is not a context for } \neg\varphi\}$$

We suppose here that $NC(KB)$ is complete, i.e. that the agent knows the truth value of all the literals for which $Cl(KB)$ is not a context.

[3]We have respected the notations of Boutilier in [2].

Definition 7. Let $\mathcal{P}$ be a set of conditional preferences. φ is a CK goal given $\mathcal{P}$ iff $\mathcal{P} \models I(\varphi|NC(KB))$ and $Cl(KB)$ is a context for φ.

We can also determine the agent's "minimal" atomic actions:

Definition 8. Let $\mathcal{P}$ be a set of conditional preferences. A set of atomic goals is a set of controllable literals $\mathcal{L} = \{l_1, \ldots, l_n\}$ such that:

- $\forall i \in \{1, \ldots, n\}$ $Cl(KB)$ is a context for l_i.
- for all CK goal φ given $\mathcal{P}$, $\mathcal{P} \models NC(KB) \wedge \mathcal{L} \rightarrow \varphi$.

Example 2. Let us resume the previous example. Let us suppose that $\mathcal{P} = \{I(l), I(\neg l|\neg s)\}$ and $\mathcal{N} = \{sh \Rightarrow s\}$.

Assume that $KB = \{\neg s\}$ (the door is not sanded) and that the agent controls both l and s (he can sand the door and lacquer it). $\neg s \in Cl(KB)$ and s is controllable so $Cl(KB)$ is a context for s. So $NC(KB) = \phi$. In this case, the potential CK goals are l and s. As $Cl(KB)$ is a context for s, s is a CK goal. $Cl(KB)$ is also a context for l, because:

- either the door is not lacquered and the agent can lacquer it.
- either the door is lacquered and the agent can do nothing.

Thus, l is a CK goal for the agent. The atomic goals set of the agent is $\{l, p\}$. Let us remark that we find the same results as Boutilier's approach in [2].

Let us suppose now that $\mathcal{P} = \{I(l|s), I(\neg l|\neg s)\}$ (in this case, the ideal world is not a l world in every model). Let us suppose that the agent controls l and s, and that $KB = \{s\}$. In this case, as the agent does not control $\neg s$, $Cl(KB)$ is not context for $\neg s$, so $NC(KB) = \{s\}$. The atomic goals set is $\{l, s\}$: the agent "should" keep the door sanded and lacquer it. Let us remark that with Boutilier's definition of CK goals, the only CK goals deducible in this case are $s \rightarrow l$ et $\neg s \rightarrow \neg l$. Our formalism better represents the decision process of the agent.

3. EXTENSION TO THE MULTI-AGENT CASE

In the following, we will consider a finite set of agents $\mathcal{A} = \{a_1, \ldots, a_n\}$.

Example 3. Let us consider the following scenario. The following preferences are assigned to a group of two agents $\{a_1, a_2\}$:

- if the door is sanded, then it should be lacquered and not covered with paper.
- if the door is not sanded, then it should be covered with paper and not lacquered.

The representation with QDT of this scenario is: $\mathcal{P} = \{I(l \wedge \neg p|s), I(p \wedge \neg l|\neg s)\}$.

If we take as a model the mono-agent formalism developed in the previous section, we have to know the group's influenceable propositions to deduce the CK goals of the group. We will therefore extend the controllability notion to a group of agents and then the CK goals of the group.

3.1 Controllability extension

For each agent of the group we have partitioned the literals of $PROP$ into two classes: the literals controllable by the agent and the literals uncontrollable by the agent. We extend this proposition to the multi-agent case by emitting two assumption. The first one is that each agent controls at least one literal:

Assumption 1. $\forall a_i \in \mathcal{A} \quad C_{a_i} \neq \phi$

The second one is about the controllability domains of the agents:

Assumption 2. The controllability domains of the agents are not necessary disjoint.

A literal can be controllable by two distinct agents. In this case, the two agents are "concurrent" to make this literal true for instance.

We now define the notion of controllability for a group of agents:

Definition 9. Let $lit(PROP)$ be the set of literals in the propositional language. the set of controllable literals by the group of agents is $C = \bigcup_{a_i \in \mathcal{A}}$ and the set of uncontrollable literals by the group of agents is $\overline{C} = lit(PROP) - C$.

The extension to propositions is the same as in definition 5.

Let us take an example: a group of agents $\{a_1, a_2\}$ is such that p is controllable by a_1 and r is controllable by a_2. Is $(p \vee q) \wedge (r \vee s)$ controllable by the group ?

In this case, as p is controllable by a_1 and r is controllable by a_2, $(p \vee q) \wedge (r \vee s)$ is controllable by $\{a_1, a_2\}$. The worlds that do not satisfy $(p \vee q) \wedge (r \vee s)$ are worlds which satisfy $\neg p$ or $\neg r$. As p and r are controllable by $\{a_1, a_2\}$, the group can keep or make p and r true.

$(p \wedge q) \vee (r \wedge s)$ is only influenceable: the worlds that satisfy $\neg q \wedge \neg s$ are not contexts for $(p \wedge q) \vee (r \wedge s)$. The contexts of $(p \wedge q) \vee (r \wedge s)$ are the worlds that satisfy $(q \vee s)$.

3.2 CK goals of the group

We can now define the notion of CK goal for a group of agents. We must precise the definitions of KB and $NC(KB)$ in the multi-agent case.

The definition of KB in the multi-agent case is not easy. For an agent, KB represents the beliefs it has about the actual world. Thus two agents may have contradictory beliefs. In this case, it seems to be difficult to determine a common KB for the group[4]. We will suppose that the agents in $\mathcal{A}$ share the same world representation.

Assumption 3. The agents of $\mathcal{A}$ share the same world representation. This representation is characterized by a finite and consistent set of propositional formulas noted KB.

The representation of the world for the group of agents is also KB.

The set of non-contextual propositions of KB is defined in the same way as in the mono-agent case. As in the mono-agent case, we assume that the group knows the truth value of all the literals for which $Cl(KB)$ is not a context. The definition of a CK goal for $\mathcal{A}$ is:

[4]Notice that we can use merging methods [13] to solve this problem.

Definition 10. Let $\mathcal{P}$ be a set of conditional preferences. φ is a CK goal for $\mathcal{A}$ iff $\mathcal{P} \models I(\varphi | NC(KB))$ and KB is a context for φ.

4. EFFECTIVE GOALS OF AN AGENT

We can now determine what are the tasks that each agent must achieve in order to achieve the group's goals. Let us resume the example given in the beginning of section 3: in the case where a_1 can lacquer the door or cover it with paper and a_2 can sand it, we can suppose that a_1's task depends on a_2's *commitments* . For instance, if a_2 commits itself to sand the door, a_1 has the task to lacquer it.

4.1 Agents' commitments

Given a literal controllable by an agent, the agent can express three positions on this literal:

- the agent can express that it will do an action that will keep or make this the literal true. We say that the agent commits itself to achieve the literal.
- the agent can express that it will not do an action that can make the literal true. We will say that the agent commits itself not to achieve the literal.
- finally, the agent can express nothing about the literal. We will say that the agent does not commit itself neither to achieve the literal nor not to achieve the literal.

To represent the commitments of each agent a_i, we will use three subsets of C_{a_i} : Com_{+,a_i}, Com_{-,a_i} and P_{a_i}. We define them in the following way:

- if l is a literal, if l is controllable by a_i and $l \in Com_{+,a_i}$, it means that "the agent a_i commits itself to achieve l";
- if l is a literal, if l is controllable by a_i and $l \in Com_{-,a_i}$, it means that "the agent a_i commits itself not to achieve l".
- $P_{a_i} = C_{a_i} - (Com_{+,a_i} \cup Com_{-,a_i})$ is the set of controllable literals by a_i and for which a_i does not commit itself to anything (i.e. a_i does not commit itself neither to achieve them nor not to achieve them).

We impose two constraints on those sets.

Constraint 1. $\forall a_i \in \mathcal{A} \quad Com_{+,a_i}$ is consistent.

Constraint 2. $\forall a_i \in \mathcal{A} \quad Com_{+,a_i} \cap Com_{-,a_i} = \phi$

Those two constraints express a kind of consistency for the agent's commitments. The first constraint express the fact that an agent does not commit itself to achieve both l and $\neg l$. The second constraint express the fact that an agent cannot commit itself both to achieve l and not to achieve l.

Definition 11. $Com_{+,\mathcal{A}}$ is the set of positive commitments of the agents:
$Com_{+,\mathcal{A}} = \bigcup_{a_i \in \mathcal{A}} Com_{+,a_i}$

$Com_{-,\mathcal{A}}$ is the set of "negative" commitments of the agents:
$Com_{-,\mathcal{A}} = \{l \in KB \ : \ \forall a_i \in \mathcal{A} \ \neg l$ controllable by $a_i \Rightarrow \neg l \in Com_{-,a_i}\}$

The meaning of $Com_{-,\mathcal{A}}$ is the following: if all the agents that control a literal l commit themselves not to achieve l and $\neg l \in KB$, we will consider that $\neg l$ will remain true. We suppose that there is no external intervention.

An hypothesis that we do on the agents' commitments is: *every CK goal of $\mathcal{A}$ is consistent with the union of $Com_{+,\mathcal{A}}$ and of $Com_{-,\mathcal{A}}$.*

Assumption 4. For every formula φ such that $\mathcal{P} \models I(\varphi|NC(KB))$ and KB is a context for φ, then :

$$Com_{-,\mathcal{A}} \cup Com_{+,\mathcal{A}} \cup \{\varphi\} \text{ is consistent.}$$

This restriction allows to eliminate some problematic cases like:

- the case where an agent which controls l commits itself to achieve l and another one which controls $\neg l$ commits itself to achieve $\neg l$ (i.e. $Com_{+,\mathcal{A}}$ not consistent).
- the case where a literal, which is not consistent with the group's CK goals, is true in KB and will remain true because the agents of the group which could make it false do not commit themselves to it.
- the case where the positive and negative commitments of the group are not consistent with some CK goal of the group.

If we want to distribute goals to a group of agents, we must first check the consistency of the agents' commitments with the group's CK goals. If the consistency is not verified, the agents must review their commitments. We do not solve the possible conflicts between the agents' commitments and the group's CK goals.

4.2 Effective goals

If the assumption 4 is verified, then the agents' commitments are consistent with the group's CK goals. The goals of each agent do not only depend on $NC(KB)$, but also on the commitments of the other agents. It seems to be intuitively correct to derive the effective goals of an agent $a_i \in \mathcal{A}$ from:

- the propositions of KB for which KB is not a context, i.e. $NC(KB)$.
- the set of positive commitments of the agents, i.e. $Com_{+,\mathcal{A}}$.
- the set of "negative" commitments of the agents, i.e. $Com_{-,\mathcal{A}}$.

We will denote the set of such formulas by $D(KB)$. This set will be used in the *conditional* part of $I(-|-)$ to deduce the effective goals of each agent.

Definition 12. We define:

$$D(KB) = NC(KB) \cup Com_{+,\mathcal{A}} \cup Com_{-,\mathcal{A}}$$

Property 1. $D(KB)$ is consistent.

Proof. $D(KB) = NC(KB) \cup Com_{+,A} \cup Com_{-,A}$.
From hypothesis 4, $Com_{+,A} \cup Com_{-,A}$ is consistent.
$NC(KB)$ is the set of formulas φ of $Cl(KB)$ such that KB is not a context for $\neg\varphi$. $NC(KB)$ is consistent because KB is consistent by definition.

Let us suppose that there is a literal l such that $l \in NC(KB)$ and $\neg l \in Com_{+,A} \cup Com_{-,A}$ (as $Com_{+,A} \cup Com_{-,A}$ is a set of literals, we can easily generalize this proof to the formulas of $NC(KB)$). In this case, KB is not a context for $\neg l$. But, $\neg l \in Com_{+,A} \cup Com_{-,A}$, thus $\neg l$ is controllable. Then by definition, KB is a context for $\neg l$ (because l is a literal). Thus $NC(KB) \cup Com_{+,A} \cup Com_{-,A}$ is consistent. □

Note that this definition is arbitrary: we can also derive the effective goals of each agent only from $NC(KB)$ for instance.

Now, we can define the notion of *effective goal* for an agent:

Definition 13. Let $\mathcal{P}$ be a set of preferences addressed to the group $\mathcal{A}$. φ is an effective goal for a_i, denoted by $EGoal_{a_i}(\varphi)$, iff $\mathcal{P} \models I(\varphi|D(KB))$ and KB is a context for φ for a_i.

As we use the $I(-|-)$ operator, we are sure that an agent cannot have contradictory goals. Like in the mono-agent case, we can define an effective atomic goals set by trivially extending definition 8.

It is also interesting to define the notion of *unfulfillment* of a CK goal φ.

Definition 14. Let φ be a CK goal $\mathcal{A}$. φ is not fulfilled (noted $Nonful(\varphi)$) iff:

$$\bigcup_{a_i \in \mathcal{A}} \{\varphi' \ : \ EGoal_{a_i}(\varphi')\} \nvdash \varphi$$

Let us present some properties of the formalism.

Property 2. Let l be a literal of $PROP$. If l is a CK goal of $\mathcal{A}$, then $\exists a_i \in \mathcal{A}$ such that $EGoal_{a_i}(l)$ holds.

Proof. Let l be a CK goal of the group $\mathcal{A}$. By definition, $\Sigma \models I(l|NC(KB))$ and KB is a context for l (0). Particularly, l is controllable by $\mathcal{A}$.

Let us suppose that l is not an effective goal for any agent of $\mathcal{A}$. Thus, $\forall a_i \in \mathcal{A}\ \Sigma \not\models I(l|D(KB))$ or l is not controllable by a_i (because l is a literal, every KB is a context for it if l is controllable).

As l is a literal, there is at least one agent a_i of $\mathcal{A}$ which controls l. So $\Sigma \not\models I(l|D(KB))$. Then:

$$\begin{array}{lll} & \exists M_0, M_0 \models \Sigma \text{ and } M \not\models I(l|D(KB)) & \\ \Longrightarrow & M_0 \not\models \overleftrightarrow{\Diamond}\,(D(KB) \wedge \Box(D(KB) \rightarrow l)) & \\ \Longrightarrow & \forall w \in W \quad M_0, w \not\models D(KB) \text{ or } \exists w' \leq w & \\ & M_0, w' \not\models D(KB) \rightarrow l & \\ \Longrightarrow & \forall w \in W \quad M_0, w \models D(KB) \quad \Rightarrow \quad \exists w' \leq w & \\ & M_0, w' \models D(KB) \wedge \neg l & (1) \end{array}$$

(0) implies $\exists w_0 \in W \quad M_0, w_0 \models UI(KB)$ and $\forall w \leq w_0 \quad M_0, w \models UI(KB) \rightarrow l$.

So, as $UI(KB) \subseteq D(KB)$:

$$\exists w_0 \in W \ M_0, w_0 \models UI(KB) \text{ and } \forall w \leq w_0 \ M_0, w \models D(KB) \rightarrow l$$

Let w_1 be a world such that $w_1 \leq w_0$. From (1), if $w_1 \models D(KB)$ then there is some $w' \in W$ such that $w' \leq w_1$ and $w' \models D(KB) \wedge \neg l$. But $w' \leq w_0$ (by transitivity of $\leq$) so $w' \models D(KB) \rightarrow l$ which is impossible.

Thus:

$$\begin{aligned} & \forall w \leq w_0 \quad w \models \neg D(KB) \\ \Longrightarrow \quad & \forall w \leq w_0 \quad w \models \neg(UI(KB) \wedge (Com_{+,\mathcal{A}} \wedge Com_{-,\mathcal{A}})) \\ \Longrightarrow \quad & \forall w \leq w_0 \quad w \models UI(KB) \rightarrow \neg(Com_{+,\mathcal{A}} \wedge Com_{-,\mathcal{A}}) \end{aligned}$$

But $M_0 \models I(l|UI(KB)$, so $\exists w_2 \in W \quad M_0, w_2 \models UI(KB)$ and $\forall w \leq w_2 \quad M_0, w \models UI(KB) \rightarrow l$.

So $M_0, min_{\leq}(w_0, w_2) \models l \wedge \neg(Com_{+,A} \wedge Com_{-,A}))$. This contradicts hypothesis 4. □

This property means that if l is a literal and a CK goal of the group, then there is at least one agent which will have for effective goal to achieve l. An immediate corollary of this property is the following [5]:

Property 3. Let $l_1, \ldots, l_n$ be n literals of $PROP$ such that $l_1 \wedge \ldots \wedge l_n$ is a CK goal of $\mathcal{A}$. Then $\forall i \in \{1, \ldots, n\}$ $\exists a_{l_i} \in \mathcal{A}$ such that $EGoal_{a_{l_i}}(l_i)$ holds.

For all the CK goals which are literals or conjunctions of literals, the process previously defined assign effective goals such that those CK goals are achieved.

5. EXAMPLE

Let us resume the previous example. The preferences imposed to the group $\{a_1, a_2\}$ are:

- if the door is sanded, then it should be lacquered and not covered with paper.
- if the door is not sanded, then it should be covered with paper and not lacquered.

The representation of this scenario within QDT is the following: $\mathcal{P} = \{I(l \wedge \neg p|s), I(p \wedge \neg l|\neg s)\}$. For each model of $\mathcal{P}$:

- $I(l \wedge \neg p|s)$ means that there is a world which satisfies s and such that all preferred worlds satisfy $s \rightarrow l \wedge \neg p$.
- $I(p \wedge \neg l|\neg s)$ means that there is a world which satisfies $\neg s$ and such that all preferred worlds satisfy $\neg s \rightarrow \neg l \wedge p$.

We suppose that there is no normality rules, so $Cl(KB) = KB$. Let us present some scenarios:

1. Let us suppose that $KB = \{s, \neg l, \neg p\}$ i.e. the agents know that the door is sanded but not lacquered nor covered with paper. Let suppose also that $\neg s$ is uncontrollable by the agents (i.e. the agents have no "means" to unsand the door), that $C_{a_1} = \{l\}$ and that $C_{a_2} = \{p, \neg p\}$ (i.e. a_1 can lacquer the door, a_2 can cover it with paper or remove the paper if necessary).

   In this case, $NC(KB) = \{s\}$, because KB is a context for l and for p. $l \wedge \neg p$ is a CK goal of the group[6]. If the agents do not commit themselves to anything, $D(KB) = \{s\}$, and then $EGoal_{a_1}(l)$ and $EGoal_{a_2}(\neg p)$ hold. a_1 has for atomic goal set $\{l\}$ (i.e. its only goal is to lacquer the door) and a_2 has $\{\neg p\}$ for atomic goals set (i.e. its only goal is not to cover the door with paper). This seems intuitively correct.

2. Let us suppose that $KB = \{\neg s, \neg l, \neg p\}$, that $C_{a_1} = \{l, \neg l\}$ and that $C_{a_2} = \{s, p, \neg p\}$. In this case, $NC(KB) = \phi$ and $(l \wedge \neg p) \vee (\neg l \wedge p)$ is a CK goal of the group. If $D(KB) = \phi$ (i.e. the agents do not commit themselves to anything), no effective goal can be derived, because a_2 controls s and could make s true. So, $Nonful((l \wedge \neg p) \vee (\neg l \wedge p))$ holds.

   But if a_2 commits itself not to achieve s (i.e. it commits itself not to sand the door), then $Com_-(\{a_1, a_2\} = \{\neg s\}$ and $EGoal_{a_2}(p)$ and $EGoal_{a_1}(\neg l)$ can be deduced: a_2 has for effective goal to cover the door with paper and a_1 has for effective goal to keep the door unlacquered.

3. Let us suppose that $KB = \{\neg s, \neg l, \neg p\}$, that $C_{a_1} = \{l\}$ and that $C_{a_2} = \{s, p\}$. In this case, $NC(KB) = \phi$ and $l \vee p$ is a CK goal of the group. Let us suppose also that a_2 commits itself to achieve s, then $Com_+(a_2) = \{p\}$. In this case, $D(KB) = \{s\}$, and $EGoal_{a_1}(l)$ and $EGoal_{a_2}(s)$ hold. a_1 should lacquer the door, a_2 should sand it. a_2 does not have for effective goal not to cover the door with paper, because it does not control $\neg p$.

4. Let us suppose that $KB = \{\neg s, \neg l, \neg p\}$, that $C_{a_1} = \{l\}$ and that $C_{a_2} = \{s, p\}$. a_2 commits itself to achieve s, so $Com_+(a_2) = \{s\}$ and a_1 commits itself not to achieve l, so $Com_-(a_1) = \{l\}$. In this case, $NC(KB) = \phi, Com_+(\{a_1, a_2\}) = \{s\}$ and $Com_-(\{a_1, a_2\}) = \{\neg l\}$. $Com_+(\{a_1, a_2\}) \cup Com_-(\{a_1, a_2\}) \cup \{s \rightarrow l\}$ is not consistent. But $s \rightarrow l$ is a CK goal of $\{a_1, a_2\}$, thus hypothesis 4 is not verified. a_1 and a_2 must review their commitments.

6. CONCLUSION

This work addresses the problem of deriving individual goals from goals assigned to a group of agents and an agency model for each agent. We have started from Boutilier's logical interpretation of decision theory. His formalism allows to deduce the goals of a single agent knowing its preferences, its abilities and its beliefs.

We have first focused on extending Boutilier's model of controllability, because we thought it was too restrictive. Partitioning the atoms of the language into two classes (the controllable atoms and the uncontrollable atoms) leads to counterintuitive conclusions: for instance, if an agent can sand the door, it can also unsand it. Our model is based on a partition of the literal of the language and avoids such conclusions.

We have then extended the controllability and CK goals notions to a group of agents and defined the commitments of a single agent using sets of literals. We were then able to determine the effective goals of each agent of the group.

This work could be extended in many directions.

The agency model can be refined: we can for instance suppose that the agents may have incomplete beliefs about the uninfluenceable propositions or that they may not share

[5] Because if $l_1 \wedge \ldots \wedge l_m$ is a CK goal of $\mathcal{A}$, $\forall i \in \{1, \ldots, l\}$ l_i is a CK goal of $\mathcal{A}$.

[6] In fact, it is the only one that is interesting. We can also deduce for instance that $(l \wedge \neg p) \vee p$ is a CK goal of the group.

the same beliefs about the world. Indeed, the "common knowledge" assumption is very strong. If we relax it, there may be conflicts between the agents beliefs. We suggest then to use some merging methods to solve such conflicts (cf. [13, 6, 7]). Those methods are used to build a common belief set from several belief sets which can be contradictory. Moreover, Let us remark that the agents beliefs are in fact knowledge, because we suppose implicitly that the beliefs of the group of agents are true in the real world. We could also suppose that the agents have "real" beliefs (so they can be false in the actual world) and analyze the impact of this assumption on our work.

It would be also interesting to compare the commitment notion developed here with the notion of "controllable and fixed" variables introduced in [5]. In this paper, we show how Boutilier's formalism can be adapted to reason with deontic notions, particularly Contrary-to-Duties. The agency model has been extended to deal correctly with this problem by using two types of variables: the controllable and fixed variables and the controllable and unfixed variables (following Carmo and Jones' terminology in [3]). The controllable and fixed variables are variables that the agent controls, but such that it does not decide to change the truth value of the variable. This notion is close to commitment.

Let us notice also that the distribution process presented here is not selective. Two agents that control both the same literal have the same effective goal about this literal. For instance, an agent a_1 may have the effective goal to lacquer to door and to sand it, and an agent a_2 may have the effective goal to sand the door. In such a case, it may be interesting to allocate only to a_2 the task to sand the door. We have worked on distribution strategies in order to avoid such derivations, but due to lack of space, we cannot present this in this paper. Those results are presented in [10].

Finally, a first study on collective responsibility and individual obligation [8] could be integrated in this framework, in order to deal with normative problems linked to the goals distribution process.

7. REFERENCES

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

[2] 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.

[3] J. Carmo and A. Jones. Deontic logic and contrary-to-duties. In *Handbook of Philosophical Logic*, volume 8: Extensions to Classical Systems 2. Kluwer Publishing Company, 2001.

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

[5] 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.

[6] L. Cholvy and C. Garion. A logic to reason on contradictory beliefs with a majority approach. In *Proceedings of IJCAI'01 Workshop on Inconsistency in Data and Knowledge*, pages 22–27, 2001.

[7] L. Cholvy and C. Garion. Answering queries addressed to merged databases: a query evaluator which implements a majority approach. In M-S Hacid, Z.W. Raś, D.A. Zighed, and Y. Kodratoff, editors, *Foundations of Intelligent Systems - Proceedings of the 13^{th} International Symposium on Methodologies for Intelligent Systems, ISMIS 2002*, volume 2366 of *Lecture Notes in Artificial Intelligence*, pages 131–139. Springer, June 2002.

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

[9] P.R. Cohen and H.J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.

[10] 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.

[11] M. Georgeff, B. Pell, M.E. Pollack, M. Tambe, and M. Wooldridge. The Belief-Desire-Intention model of agency. In J. Muller, M. Singh, and A. Rao, editors, *Intelligent Agents V*, volume 1365 of *Lecture Notes in Artificial Intelligence*. Springer Publishers, 1999.

[12] B.J. Grosz, L. Hunsberger, and S. Kraus. Planning and acting together. *AI Magazine*, 20(4):23–34, 1999.

[13] S. Konieczny and R. Pino-Pérez. On the logic of merging. In *Proceedings of the Sixth International Conferences on Principles of Knowledge Representation and Reasoning (KR'98)*, pages 488–498, Trento, Italy, June 1998. Morgan Kaufmann.

[14] S. Kraus. Negociation and cooperation in multi-agent environments. *Artificial Intelligence Journal - Special Issue on Economic Principles of Multi-Agent Systems*, 94(1-2):79–98, 1997.

[15] C. Lafage and J. Lang. Logical representation of preferences for group decision making. In A.G. Cohn, F. Giunchiglia, and B. Salman, editors, *Proceedings of the Seventh International Conference KR 2000*, pages 457–468, Beckenridge, Colorado (USA), 2000.

[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] O. Shehory and S. Kraus. Methods for task allocation via agent coalition formation. *Artificial Intelligence*, 101(1-2):165–200, 1998.

[18] M.P. Singh, A.S Rao, and M.P. Georgeff. Formal methods in DAI : Logic-based representation and reasoning. In G. Weiss, editor, *Multiagent Systems : A Modern Approach to Distributed Artificial Intelligence*, chapter 8, pages 331–376. MIT Press, 1999.

[19] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behaviour*. Princeton University Press, 1947.

[20] M.J. Wooldridge and N.R. Jennings. Agent theories, architectures, and languages : a survey. In M.J. Wooldridge and N.R. Jennings, editors, *Proceedings of the ECAI-94 Workshop on Agent Theories, Architectures, and Languages*, volume 890 of *LNAI*, pages 1–25. Springer-Verlag, 1994.