Deriving Individual Obligations from Collective Obligations

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

A collective obligation is an obligation directed to a group of agents so that the group, as a whole, is obliged to achieve a given task. The problem investigated here is the impact of collective obligations on individual obligations, i.e. obligations directed to single agents of the group. In this case, we claim that the derivation of individual obligations from collective obligations depends on several parameters among which the ability of the agents (i.e. what they can do) and their own personal commitments (i.e. what they are determined to do). As for checking if these obligations are fulfilled or not, we need to know what are the actual actions performed by the agents.

Categories and Subject Descriptors

F.4.1 [**Mathematical Logic**]: Modal Logic; I.2.4 [**Knowledge Representation Formalisms and Methods**]: Modal Logic; I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems

General Terms

Theory

Keywords

modal logic, deontic logic, multiagent systems

1. INTRODUCTION

This poster studies the relation between collective obligations directed to a group of agents and the individual obligations directed to the single agents of the group. We study this relation in the case when the group of agents is not structured by any hierarchical structure and has no representative agent. Notice that this work has been partially studied in [2].

A collective obligation addressed to a group of agents is such that this group, as a whole, is obliged to achieve a given task. This comes to say that a given task is assigned as a goal to the group as a whole. For instance, (cf. [3]), when a mother says: "*Boys, you have to set the table*", she defines an obligation aimed at the group of her boys. The goal assigned to the boys is to set the table and the mother expects that the table will be set by some actions performed by her boys. Whether only one of her boys or all of them will bring it about that the table is set is not specified by the mother.

Understanding how the collective obligations are translated into individual obligations is the problem which is investigated here. We claim that the derivation of individual obligations from collective obligations depends on several parameters among which the ability of the agents and their own personal commitments. Latter on, by examining the actual actions of each agent of the group, one can check if these obligations are satisfied or violated.

2. COLLECTIVE OBLIGATIONS

This poster addresses the question of the translation of a collective obligation into individual obligations in the rather general case when the collective obligations are conditional ones. We use Boutilier's ideality operator $I(-|-)$ [1] to express such obligations. For instance, $I(s|d)$ means "if there is a dog, then it there should be a signal".

We consider a finite set of agents $\mathcal{A} = \{a_1, \ldots, a_n\}$. In the following, we suppose that Σ is a set of conditional preferences expressing the obligations directed to $\mathcal{A}$.

Starting from Boutilier's work, as we have shown in [2], in order to derive obligations we need to consider what the agents *can* do and *cannot* do. For representing ability, we partition for each agent a_i the atoms into two classes : C_{a_i} which represents the atoms the agent can change the truth value of and $\overline{C}_{a_i}$ which represents the atoms the agent can change the truth value of. From this partition, we can build the set of atoms controllable by $\mathcal{A}$: $C = \bigcup_{a_i \in \mathcal{A}} C_{a_i}$. We then extend this notion to propositions following Boutilier.

Given a knowledge base KB (a set of propositional formulas which represents the common beliefs of the agents about the world), we can define $UI(KB)$: $UI(KB)$ is the set of formulas which are true in the world and whose truth value cannot be changed by the group of agents.

From those definitions, we can derive the group's obligations :

Definition 1. **The group $\mathcal{A}$ has the obligation of φ toward the agent who directed the collective obligation** iff $\Sigma \models I(\varphi | UI(KB))$ with φ controllable by $\mathcal{A}$. This is noted $O_{\mathcal{A}}\phi$.

AAMAS'03, July 14–18, 2003, Melbourne, Australia.
ACM 1-58113-683-8/03/0007.

The group's obligation are derived from the conditional obligations Σ directed to the group given what is fixed (represented by $UI(KB)$) and given what the group controls. We can go further and direct those obligations to the sub-groups that can really fulfil them :

Definition 2. Let ϕ be a proposition. Let $\mathcal{A}_\phi$ be the union of the minimal subsets of $\mathcal{A}$ that control ϕ. We say that **the sub-group $\mathcal{A}_\phi$ has the obligation of ϕ, toward $\mathcal{A}$** iff $\Sigma \models I(\phi|UI(KB))$. It is denoted by $O^{\mathcal{A}}_{\mathcal{A}_\phi}\phi$.

Example 1. Let us consider a group $\mathcal{A}$ of three agents named Alice (denoted by A), John (denoted by J) and Tom (denoted by T). That group is addressed the following obligations: *"if the statistics are collected (s), then the financial proposal (fp) and the scientific proposal (sp) should be written"* and *"if the statistics are not collected, then the financial proposal should not be written, but the scientifically proposal should be"*. This scenario is translated into the following set of formulas: $\{I(fp \wedge sp|s), I(\neg fp \wedge sp|\neg s)\}$.

Let us suppose that $KB = \{s, \neg fp, \neg sp\}$. Let us also suppose that $C_A = C_T = \{fp\}$ and that $C_J = \{sp\}$. So $\mathcal{A}$ controls both fp and sp.

In this case, $UI(KB) = \{s\}$ and $\mathcal{A}$ has the obligation of $fp \wedge sp$, thus $\mathcal{A}$ has the obligation of fp and the obligation of sp. Moreover, as fp is controllable by both Alice and Tom, then $\{A, T\}$ has the obligation toward $\mathcal{A}$ to achieve fp. Finally, as John is the only agent which controls sp, $\{J\}$ has the obligation toward $\mathcal{A}$ to achieve sp.

3. INDIVIDUAL OBLIGATIONS

In order to derive individual obligations, we think that we have to consider each agent's commitments. Given an atom it controls, an agent may have three positions, represented by three sets. $Com_{+,a_i} \subseteq C_{a_i}$ is the set of atoms a_i controls such that a_i commits itself to make them true. $Com_{-,a_i} \subseteq C_{a_i}$ is the set of atoms a_i controls such that a_i commits itself not to make them true. $P_{a_i} = C_{a_i} \setminus (Com_{+,a_i} \cup Com_{-,a_i})$ is the set of atoms a_i controls such that a_i does not commit to make them true nor commits not to make them true.

We impose some consistency constraints on those sets (not detailed here). For instance, an agent cannot commit itself to make an atom true and to make it not true.

We can now now characterise the obligations that are directed to some agents of the group, given the obligations of the group and given the agent's commitments. Individual obligations are defined by:

Definition 3. Let ϕ be a proposition such that $O_\mathcal{A}\phi$ holds. Let a_i be an agent of $\mathcal{A}$. If there is some minimal $\{l_1, \ldots, l_m\} \subseteq Com_{+,a_i}$ such that $\models l_1 \wedge \ldots \wedge l_m \rightarrow \phi$, we say that a_i **is obligated to satisfy** $l_1 \wedge \ldots \wedge l_m$ **toward** $\mathcal{A}_\phi$. This is denoted by $O^{\mathcal{A}_\phi}_{a_i}(l_1 \wedge \ldots \wedge l_m)$.

For checking if the different obligations introduced previously are violated or not, we must examine the results of the agents' actions. Let KB_n be the state of the world resulting from the actions of the agents and let ϕ be such that $O_\mathcal{A}\phi$.

- if $KB_n \models \phi$ then the collective obligation is not violated. We say that the collective obligation is fulfilled.
- if $KB_n \not\models \phi$ then $O_\mathcal{A}(\phi)$ is violated. The whole group $\mathcal{A}$ is taken as responsible for the violation, by the agent who directed the collective obligation. We consider $\mathcal{A}_\phi$. Thus, since $KB_n \not\models \phi$, $O^{\mathcal{A}}_{\mathcal{A}_\phi}(\phi)$ is violated too and $\mathcal{A}_\phi$ is taken as responsible, by $\mathcal{A}$, for this violation.
- let us consider all the agents a_i such that there is some φ such that $O^{\mathcal{A}_\phi}_{a_i}(\varphi)$. If $KB_n \not\models \varphi$, the obligation $O^{\mathcal{A}_\phi}_{a_i}(\varphi)$ is violated too and a_i can be taken as responsible by $\mathcal{A}_\phi$ of the violation of its commitment i.e. $O^{\mathcal{A}_\phi}_{a_i}\varphi$. Moreover, if $KB_n \not\models \phi$, a_i can be taken by $\mathcal{A}_\phi$ responsible for the violation of $O^{\mathcal{A}}_{\mathcal{A}_\phi}(\phi)$.

Example 2. (ex. 1 continuing) Let us suppose that Alice commits herself to write the financial proposal. In this case, $Com_{+,A} = \{fp\}$ and we can derive $O^{\{A,T\}}_{A}(fp)$ (because $O_\mathcal{A}(fp)$ holds). Alice is obligated to achieve fp toward $\{A, T\}$.

Assume that Alice writes the financial proposal, that John writes the scientific proposal and that Tom does nothing. In this case, $KB_{next} = \{s, sp, fp\}$ and all the obligations are fulfilled.

Assume now that Alice does not write the financial proposal, but that Tom writes the financial proposal. Assume also that John writes the scientific proposal. In this case, the collective obligation $O_\mathcal{A}(fp \wedge sp)$ is satisfied, $O^{\mathcal{A}}_{\{A,T\}}(fp)$ is satisfied too, but $O^{\{A,T\}}_{A}(fp)$ is violated. Even if the group fulfilled its obligations, the obligation of Alice toward $\{A, T\}$ to achieve fp is violated.

4. CONCLUSION

In this poster, we have presented a preliminary work about collective obligations, i.e. obligations directed to a group of agents. The first step was to determine the obligations of the group, given what is fixed in the world and given what this group as a whole, can do. Then we considered that, if the group is obliged to make A true, then it induces another obligation to the very sub-group who control A: that sub-group is obliged, toward the whole group, to make A true. These definitions of obligation are direct extensions, to the multi-agent case, of one definition provided by Boutilier in the single-agent case.

As for individual obligations, they are induced as soon as an agent commits itself to satisfy, by one of its action, an obligation of the group. Checking if these obligations are violated or not need to consider the state of the world obtained after the agents' actual actions.

5. REFERENCES

[1] C. Boutilier. Toward a logic for qualitative decision theory. In *Principles of Knowledge Representation and Reasoning (KR'94)*. J. Doyle, E. Sandewall and P. Torasso Editors, 1994.

[2] 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 (ΔEON'02)*, pages 55–71, London, may 2002.

[3] L. Royakkers and F. Dignum. No organization without obligations: how to formalize collective obligation ? In M. Klusch and L. Kerschberg, editors, *DEXA'00 (LNCS-1873)*, pages 191–207. Springer-Verlag, 2000.