Fair Distribution of Collective Obligations

Guido Boella[1] and **Leendert van der Torre**[2]

Abstract. In social mechanism design, obligation distribution creates individual or contractual obligations that imply a collective obligation. A distinguishing feature from group planning is that also the sanction of the collective obligation has to be distributed, for example by creating sanctions for the individual or contractual obligations. In this paper we address fairness in obligation distribution for more or less powerful agents, in the sense that some agents can perform more or less actions than others. Based on this power to perform actions, we characterize a trade-off in negotiation power. On the one hand, more powerful agents may have a disadvantage during the negotiation, as they may be one of the few or even the only agent who can see to some of the actions that have to be performed to fulfill the collective obligation. On the other hand, powerful agents may have an advantage in some negotiation protocols, as they have a larger variety of proposals to choose from. Moreover, powerful agents have an advantage because they can choose from a larger set of possible coalitions. We present an ontology and measures to find a fair trade-off between these two forces in social mechanism design.

1 Ontology

Power may affect the distribution of obligation in various ways, and we therefore propose to analyze the obligation distribution problem in terms of social concepts like power and dependence. Power has been identified as a central concept for modeling social phenomena in multi-agent systems by various authors [5, 6, 8, 12], as Castelfranchi observes both to enrich agent theory and to develop experimental, conceptual and theoretical new instruments for the social sciences [6]. In most of these proposals, power is interpreted as the ability of agents to achieve goals. For example, in the so-called power view on multi-agent systems [2], a multi-agent system consists of a set of agents (A), a set of goals (G), a function that associates with each agent the goals the agent desires to achieve ($goals$), and a function that associates with each agent the sets of goals it can achieve ($power$). To be precise, to represent conflicts the function $goals$ returns a set of set of goals for each set of agents. Such abstract structures have been studied as qualitative games by Wooldridge and Dunne [11].

To model the role of power in obligation distribution in normative multiagent systems, we associate with each norm the set of agents that has to fulfill it, and of each norm we represent how to fulfill it, and what happens when it is not fulfilled.

- First, we associate with each norm n a set of goals $O(n) \subseteq G$. Achieving these normative goals $O(n)$ means that the norm n has been fulfilled; not achieving these goals means that the norm is violated. We assume that every normative goal can be achieved by the group, i.e., that the group has the power to achieve it.
- Second, we associate with each norm a set of goals $V(n)$ which will not be achieved if the norm is violated (i.e., when its goals are not achieved), this is the sanction associated to the norm. We assume that the sanction affects at least one goal of each agent of the group the obligation belongs to, and that the group of agents does not have the power to achieve these goals.

Definition 1 *Let a normative multi-agent system be a tuple* $\langle A, G, goals, power, N, OD, O, V \rangle$ *where:*

- *the agents A, goals G and norms N are three finite disjoint sets;*
- $goals : A \to 2^G$ *is a function that associates with each agent the goals the agent desires;*
- $power : 2^A \to 2^{2^G}$ *is a function that associates with each set of agents the sets of goals the set of agents can achieve;*
- $OD : N \to 2^A$ *is a function that associates with every norm a set of agents that have to see to it that the norm is fulfilled.*
- $O : N \to 2^G$ *is a function that associates with each norm the goals which must be achieved to fulfill the norm; We assume for all $n \in N$ that there exists $H \in power(OD(n))$ with $O(n) \subseteq H$;*
- $V : N \to 2^G$ *is a function that associates with each norm the goals that will not be achieved if the norm is violated; We assume that for all $n \in N$ and $a \in OD(n)$ that $V(n) \cap goals(a) \neq \emptyset$, and for all $B \subseteq OD(n)$ and $H \in power(B)$ that $V(n) \cap H = \emptyset$.*

2 Power of the people

Some studies of obligation distribution consider logical relations among collective and individual obligations [9]. Cholvy and Garion [7] claim that if there is a set of agents subject to an obligation to perform some task, then the derivation of individual obligations from collective obligations depends on several parameters, among which the ability of the agents. If an agent is the only one able to perform a part of that task, then it is obliged to do that part and it is also obliged to do that towards the other members of the set.

For example, they consider three children who are obliged by their mother to prepare the table for dinner. The oldest child is the only one who is tall enough to get the glasses on the cupboard. The whole group is responsible for the violation of the collective obligation, but in case the violation is due to the fact that the oldest boy did not bring the glasses, only he can be taken responsible by the group because he was the only one able to take the glasses.

This disadvantage of more powerful agents is measured for each agent a and norm n as the number of normative goals it has the power to achieve, each weighted by the number of other agents within the group that can achieve the same goal. $m_1(a, n) = 0$ if $a \notin OD(n)$, and otherwise:

- $m_1(a, n) = \Sigma_{g \in power(a) \cap O(n)} \frac{1}{|\{b \in OD(n) | g \in power(b)\}|}$

[1] Dipartimento di Informatica - Università di Torino - Italy. E-mail: guido@di.unito.it

[2] Department of Computer Science, University of Luxembourg. E-mail: leendert@vandertorre.com

3 Power of the king

An analysis based on power and dependence, as is common in social theory, suggests that the ability possessed by only one agent makes the remaining agents depend on him, since they lack the power to do part of the task they are obliged to. In this sense the oldest boy is in the best position, rather than having an additional burden and being sanctioned both for not respecting the collective obligation and his own obligation.

The dependence on more powerful agents can be made explicit when the negotiation process is made explicit. For example, the oldest boy has more power in the negotiation for the distribution of the task, and, by exercising this power, he may end up doing less than the other boys. An agreement is the result of a negotiation which has to take into account the dependence relations among the agents. E.g., the kids are first obliged to negotiate the distribution of the obligation, and then, only if they find a successful distribution of obligations, the oldest kid becomes obliged to see to the glasses. Note that in this negotiation model the distribution is not centralized, which is another reason why it is more complex that a planning problem.

In the negotiation protocol proposed in [4], each agent in turn makes a proposal of a complete distribution, which can be accepted or rejected by the other agents. Agents may not repeat their own proposal. If no proposal is accepted, then the norm will be violated. Roughly, when an agent can make many proposals, he has the advantage that he can propose deals which are beneficial to himself in all rounds. This models the intuition that an agent with more possible proposals has more power in the negotiation. This holds regardless of the penalty for breaking the negotiation.

For example, assume for a norm n that an agent $a \in OD(n)$ is allowed to propose a distribution $d(n) : OD(n) \rightarrow 2^{power(OD(n))}$ that associates with every agent some of the goals he has the power to achieve, such that the distribution implies all the goals of the norm $O(n) \subseteq \cup_{a \in OD(n)} d(n)(a)$ (note that here we assume that if a goal can be achieved by a set of agents, then it can be achieved by an individual agent). With $D(n)$ we refer to the set of all such $d(n)$. Moreover, assume that an agent cannot make two proposals in which itself is dealt the same goals. The measure of this example is the number of elements of $D(n)$ in which $d(n)(a)$ is distinct.

- $m_2(a, n) = |\{d(n)(a) \mid d(n) \in D(n)\}|$

If the distributions that may be proposed have to satisfy other criteria, for example related to the distribution of the sanction [4], then the definition of d can be adapted accordingly.

4 Coalition power

Agents depend on more powerful agents not only to accomplish collective obligations, but also to fulfill their other desires. Thus, if we not only consider the collective obligation that has to be distributed but also the other desires of the agents, then powerful agents have an additional advantage in negotiation. For example, they may threaten to do something the other agents dislike, who therefore will accept an inferior solution.

A framework to study such dependencies is coalition formation. In contrast to the negotiation model in the previous section, coalition formation considers not only a single collective obligation to be distributed, but many of them. Such a setting is given in our definition of normative multi-agent system in Definition 1. A way to calculate the possible coalitions from a power view on normative multi-agent systems is given in [1, 3]. In that paper, a coalition is only considered in a power structure if each agent contributes and receives something in a coalition (called do-ut des).

- $m_3(a) = |coalitions(a)|$

Of the three measures proposed in this paper, this third measure is the most abstract. It is tempting to make it more precise by quantifying over the goals of the norms the agent is involved in. However, the impact of the dependencies in coalition formation on obligation distribution has not been studied in any detail thus far, and therefore we have chosen this more general measure. Consequently, it can only be taken as an indication of the impact on obligation distribution.

5 Concluding remarks

In some social mechanisms the agents with more power will end up doing more, but in other ones they will end up doing less. In the extreme case, a very powerful agent might do nothing at all - or the distribution is not made at all. The measures discussed in this paper can be used to enforce fairness in the distribution of agents. Another solution is to build a normative system that ensures a fair distribution. In this approach, there are norms that, for example, indicate the number of deals an agent can propose.

Most multi-agent system models concerned with complex cognitive tasks like obligation creation, negotiation and distribution [4, 7, 9, 10] are based on relatively detailed cognitive models incorporating, for example, beliefs, obligations, desires, intentions, goals, preferences, and so on. Some of these elements can be used to make the measures more precise. For example, another advantage may have more reliable knowledge about the goals and abilities of other agents. If we extend the model with such beliefs of agents about the power and goals of other agents, then this advantage can be measured too.

REFERENCES

[1] G. Boella, L. Sauro, and L. van der Torre. An abstraction from power to coalition structures. In *Proceedings of ECAI 2004*, pages 965–966, 2004.

[2] G. Boella, L. Sauro, and L. van der Torre. Social viewpoints on multiagent systems. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS'04)*, pages 1358–1359, 2004.

[3] G. Boella, L. Sauro, and L. van der Torre. Strengthening admissible coalitions. In *Proceedings of ECAI 2006*, 2006.

[4] G. Boella and L. van der Torre. Negotiating the distribution of obligations with sanctions among autonomous agents. In *Procs. of ECAI 2004*, pages 13–17, 2004.

[5] S. Brainov and T. Sandholm. Power, dependence and stability in multiagent plans. In *Procs. of AAAI99*, pages 11–16, 1999.

[6] C. Castelfranchi. Micro-macro constitution of power. *ProtoSociology*, 18-19, 2003.

[7] L. Cholvy and C. Garion. Collective obligations, commitments and individual obligations: a preliminary study. In *Procs. of* Δ*EON02*, London, 2002.

[8] R. Conte and J. Sichman. Multi-agent dependence by dependence graphs. In *Procs of AAMAS02*, pages 483–490, 2002.

[9] F. Dignum and L. Royakkers. Collective obligation and commitment. In *Procs. of 5th Int. Conference on Law in the Information Society*, Florence, 1998.

[10] J. Ferber, O. Gutknecht, and F. Michel. From agents to organizations: an organizational view of multiagent systems. In *Procs. of AOSE03*, LNCS 2935, page 214230. Springer, 2003.

[11] M. Wooldridge and P.E. Dunne. On the computational complexity of qualitative coalitional games. *Artif. Intell.*, 158(1):27–73, 2004.

[12] Lopez y Lopez. *Social Power and Norms: Impact on agent behaviour*. PhD thesis, June 2003.