Power in Norm Negotiation

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

[1]Dipartimento di Informatica - Università di Torino - Italy. email: guido@di.unito.it
[2]University of Luxembourg - Luxembourg. e-mail: leendert@vandertorre.com

Abstract. In social mechanism design, norm negotiation creates individual or contractual obligations fulfilling goals of the agents. The social delegation cycle distinguishes among social goal negotiation, obligation and sanction negotiation and norm acceptance. Power may affect norm negotiation in various ways, and we therefore introduce a new formalization of the social delegation cycle based on power and dependence, without referring to the rule structure of norms, actions, decision variables, tasks, and so on.

1 Introduction

Normative multiagent systems [1–4] provide agents with abilities to automatically devise organizations and societies coordinating their behavior via obligations, norms and social laws. A distinguishing feature from group planning is that also sanctions and control systems for the individual or contractual obligations can be created. Since agents may have conflicting goals with respect to the norms that emerge, they can negotiate amongst each other which norm will be created.

The social delegation cycle [5] explains the negotiation of new social norms from cognitive agent goals in three steps. First individual agents or their representatives negotiate social goals, then a social goal is negotiated in a social norm, and finally the social norm is accepted by an agent [6] when it recognizes it as a norm, the norm contributes to the goals of the agent, and it is obeyed by the other agents. A model of norm negotiation explains also what it means, for example, to recognize or to obey a norm, and how new norms interact with existing ones.

Power may affect norm negotiation in various ways, and we therefore propose to analyze the norm negotiation 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 [7–10], as Castelfranchi observes both to enrich agent theory and to develop experimental, conceptual and theoretical new instruments for the social sciences [11].

To motivate our social-cognitive model, we contrast it with an abstract description of the social delegation cycle using game theoretic artificial social systems. The problem studied in artificial social systems is the design, emergence or more generally the creation of social laws. Shoham and Tennenholtz [12] introduce social laws in a setting without utilities, and they define *rational* social laws as social laws that improve a social game variable [13]. We follow Tennenholtz' presentation for stable social laws [14].

Moreover, we also contrast our social-cognitive model with existing highly detailed models of the social delegation cycle, like the ones we have proposed within normative

multiagent systems [5]. The challenge to define social mechanisms, as we see it, is to go beyond the classical game theoretic model by introducing social and cognitive concepts and a negotiation model, but doing so in a minimal way. In the model proposed in this paper we therefore keep goals and obligations abstract and we do not describe them by first-order (or propositional) logic or their rule structure, we do not introduce decisions, actions, tasks, and so on. Similar concerns are also mentioned by Wooldridge and Dunne in their qualitative game theory [15].

The layout of this paper is as follows. In Section 2 we discuss an abstract model of the social delegation cycle in Tennenholtz' game-theoretic artificial social systems. In Section 3 we define our social-cognitive conceptual model of multiagent systems in which we study and formalize the social delegation cycle, and in Section 4 we define the negotiation protocol. In Section 5 we formalize goal negotiation, in Section 6 we formalize norm negotiation, and in Section 7 we formalize the acceptance relation.

2 Social delegation cycle using artificial social systems

In Tennenholtz' game-theoretic artificial social systems, the goals or desires of agents are represented by their utilities. A game or multi-agent encounter is a set of agents with for each agent a set of strategies and a utility function defined on each possible combination of strategies. Tennenholtz only defines games for two agents to keep the presentation of artificial social systems as simple as possible, but he also observes [14, footnote 4] that the extension to the multi-agent case is straightforward.

Definition 1. *A* game *(or a* multi-agent encounter*) is a tuple $\langle N, S, T, U_1, U_2\rangle$, where $N = \{1, 2\}$ is a set of agents, S and T are the sets of strategies available to agents 1 and 2 respectively, and $U_1 : S \times T \to \mathbb{R}$ and $U_2 : S \times T \to \mathbb{R}$ are utility functions for agents 1 and 2, respectively.*

The social goal is represented by a minimal value for the social game variable. Tennenholtz [14] uses as game variable the maximin value. This represents safety level decisions, in the sense that the agent optimizes its worst outcome assuming the other agents may follow any of their possible behaviors.

Definition 2. *Let S and T be the sets of strategies available to agent 1 and 2, respectively, and let U_i be the utility function of agent i. Define $U_1(s, T) = \min_{t\in T} U_1(s, t)$ for $s \in S$, and $U_2(S, t) = \min_{s\in S} U_2(s, t)$ for $t \in T$. The* maximin value for agent *1 (*respectively *2) is defined by $\max_{s\in S} U_1(s, T)$ (respectively $\max_{t\in T} U_2(S, t)$). A strategy of agent i leading to the corresponding maximin value is called a* maximin strategy for agent i.

The social norm is represented by a social law, characterized as a restriction of the strategies available to the agents. It is *useful* with respect to an efficiency parameter e if each agent can choose a strategy that guarantees it a payoff of at least e.

Definition 3. *Given a game $g = \langle N, S, T, U_1, U_2\rangle$ and an efficiency parameter e, we define a social law to be a restriction of S to $\overline{S} \subseteq S$, and of T to $\overline{T} \subseteq T$. The social law is* useful *if the following holds: there exists $s \in \overline{S}$ such that $U_1(s, \overline{T}) \geq e$, and there*

exists $t \in \overline{T}$ *such that* $U_2(\overline{S}, t) \geq e$*. A (useful) convention is a (useful) social law where* $|\overline{S}| = |\overline{T}| = 1$.

A social law is *quasi-stable* if an agent does not profit from violating the law, as long as the other agent conforms to the social law (i.e., selects strategies allowed by the law). Quasi-stable conventions correspond to Nash equilibria.

Definition 4. *Given a game* $g = \langle N, S, T, U_1, U_2 \rangle$*, and an efficiency parameter* e*, a* quasi-stable social law *is a useful social law (with respect to* e*) which restricts* S *to* $\overline{S}$ *and* T *to* $\overline{T}$*, and satisfies the following: there is no* $s' \in S - \overline{S}$ *which satisfies* $U_1(s', \overline{T}) > \max_{s \in \overline{S}} U_1(s, \overline{T})$*, and there is no* $t' \in T - \overline{T}$ *which satisfies* $U_2(\overline{S}, t') > \max_{t \in \overline{T}} U_2(\overline{S}, t)$.

The efficiency parameter can be seen as a social kind of *utility aspiration level*, as studied by Simon [16]. Such aspiration levels have been studied to deal with limited or resource-bounded reasoning, and have led to the development of goals and planning in artificial intelligence; we therefore use a goal based ontology in this paper. The three steps of the social delegation cycle in this classical game-theoretic setting can be represented as follows. Goal negotiation implies that the efficiency parameter is higher than the utility the agents expect without the norm, for example represented by the Nash equilibria of the game. Norm negotiation implies that the social law is useful (with respect to the efficiency parameter). The acceptance relation implies that the social law is quasi-stable.

We use the game-theoretical model to motivate our conceptual model of normative multiagent systems. Due to the uniform description of agents in the game-theoretic model, it is less clear how to distinguish among kinds of agents. For example, the unique utility aspiration level does not distinguish the powers of agents to negotiate a better deal for themselves than for the other agents. Moreover, the formalization of the social delegation cycle does neither give a clue how the efficiency parameter is negotiated, nor how the social law is negotiated. For example, the goals or desires of the agents as well as other mental attitudes may play a role in this negotiation. There is no sanction or control system in the model (adding a normative system to encode enforceable social laws to the artificial social system complicates the model [17]). Finally, an additional drawback is that the three ingredients of the model (agent goals, social goals, and social laws) are formalized in three completely different ways.

3 Power viewpoint on normative multiagent systems

In this paper we follow the definition of power as the ability of agents to achieve goals. Thus, an agent is more powerful than another agent if it can achieve more goals.

For example, in the so-called power view on multi-agent systems [18], 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, since goals can be conflicting in the sense that achieving some goals may make it impossible to achieve other goals, 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 [15], though they do not call the ability of agents to achieve goals their power. To model trade-offs among goals of agents, we introduce a priority relation among goals.

Definition 5. *Let a multiagent system be a tuple $\langle A, G, goals, power, \geq \rangle$ where:*

- *the set of agents A and the set of goals G are two finite disjoint sets;*
- *$goals : A \rightarrow 2^G$ is a function that associates with each agent the goals the agent desires to achieve;*
- *$power : 2^A \rightarrow 2^{2^G}$ is a function that associates with each set of agents the sets of goals the set of agents can achieve;*
- *$\geq$: $A \rightarrow \subseteq 2^G \times 2^G$ is a function that associates with each agent a partial pre-ordering on the sets of his goals;*

To model the role of power in norm negotiation, we extend the basic power view in a couple of ways. To model obligations we introduce a set of norms, 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. In particular, we relate norms to goals in the following two ways.

- 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 group of agents does not have the power to achieve these goals.

Since we accept norms without sanctions, we do not assume that the sanction affects at least one goal of each agent of the group the obligation belongs to.

Definition 6. *Let a normative multi-agent system be a tuple $\langle MAS, N, O, V \rangle$ extending a multiagent system $MAS = \langle A, G, goals, power, \geq \rangle$ where:*

- *the set of norms N is a finite set disjoint from A and G;*
- *$O : N \times A \rightarrow 2^G$ is a function that associates with each norm and agent the goals the agent must achieve to fulfill the norm; We assume for all $n \in N$ and $a \in A$ that $O(n, a) \in power(\{a\})$;*
- *$V : N \times A \rightarrow 2^G$ is a function that associates with each norm and agent the goals that will not be achieved if the norm is violated by agent a; We assume for each $B \subseteq A$ and $H \in power(B)$ that $(\cup_{a \in A} V(n, a)) \cap H = \emptyset$.*

An alternative way to represent normative multiagent systems replaces the function *power* by a function representing dependencies between agents. For example, a function of minimal dependence can be defined as follows. Agent a depends on agent set $B \subseteq A$ regarding the goal g if $g \in goals(a)$, $g \notin power(\{a\})$, $g \in power(B)$, and there is no $C \subset B$ such that $g \in power(C)$. Note that dependence defined in this way is more abstract than power, in the sense that we have defined dependence in terms of power, but we cannot define power in terms of dependence.

4 Generic negotiation protocol

A negotiation protocol is described by a set of sequences of negotiation actions which either lead to success or failure. In this paper we only consider protocols in which the agents propose a so-called deal, and when an agent has made such a proposal, then the other agents can either accept or reject it (following an order ($\succ$) of the agents). Moreover, they can also end the negotiation process without any result.

Definition 7 (Negotiation Protocol). *A negotiation protocol is a tuple* $\langle \mathit{Ag}, \mathit{deals}, \mathit{actions}, \mathit{valid}, \mathit{finished}, \mathit{broken}, \succ \rangle$*, where:*

- *the agents Ag, deals and actions are three disjoint sets, such that actions* $=$ $\{\mathit{propose}(a,d), \mathit{accept}(a,d), \mathit{reject}(a,d) \mid a \in \mathit{Ag}, d \in \mathit{deals}\} \cup \{\mathit{breakit}(a) \mid a \in \mathit{Ag}\}$.
- *valid, finished, broken are sets of finite sequences of actions.*

We now instantiate this generic protocol for negotiations in normative multiagent systems. We assume that a sequence of actions (a history) is valid when each agent does an action respecting this order. Then, after each proposal, the other agents have to accept or reject this proposal, again respecting the order, until they all accept it or one of them rejects it. When it is an agent's turn to make a proposal, it can also end the negotiation by breaking it. The history is *finished* when all agent have accepted the last deal, and *broken* when the last agent has ended the negotiations.

Definition 8 (NMAS protocol). *Given a normative multiagent system* $\langle MAS, N, O, V \rangle$ *extending a multiagent system* $MAS = \langle A, G, \mathit{goals}, \mathit{power}, \geq \rangle$*, a negotiation protocol for NMAS is a tuple* $NP = \langle A, \mathit{deals}, \mathit{actions}, \mathit{valid}, \mathit{finished}, \mathit{broken}, \succ \rangle$*, where:*

- $\succ \subseteq A \times A$ *is a total order on* A*,*
- *a history* h *is a sequence of actions, and* $\mathit{valid}(h)$ *holds if:*
  - *the propose and breakit actions in the sequence respect* $\succ$*,*
  - *each propose is followed by a sequence of accept or reject actions respecting* $\succ$ *until either all agents have accepted the deal or one agent has rejected it,*
  - *there is no double occurrence of a proposal* $\mathit{propose}(a,d)$ *of the same deal by any agent* $a \in \mathit{Ag}$*, and*
  - *the sequence* h *ends iff either all agents have accepted the last proposal (*$\mathit{finished}(h)$*) or the last agent has broken the negotiation (*$\mathit{broken}(h)$*) instead of making a new proposal.*

In theory we can add additional penalties when agents break the negotiation. However, since it is in the interest of all agents to reach an agreement, we do not introduce such sanctions. In this respect norm negotiation differs from negotiation about obligation distribution [19], where it may be the interest of some agents to see to it that no agreement is reached. In such cases, sanctions must be added to the negotiation protocol to motivate the agents to reach an agreement.

Example 1. Assume three agents and the following history.
$action_1 : propose(a_1, d_1)$
$action_2 : accept(a_2, d_1)$
$action_3 : reject(a_3, d_1)$
$action_4 : propose(a_2, d_2)$
$action_5 : accept(a_3, d_2)$
$action_6 : accept(a_1, d_2)$

We have $valid(h)$, because the order of action respects $\succeq$, and we have $accepted(h)$, because the history ends with acceptance by all agents ($action_5$ and $action_6$) after a proposal ($action_4$).

The open issue of the generic negotiation protocol is the set of deals which can be proposed. They depend on the kind of negotiation. In social goal negotiation the deals represent a social goal, and in norm negotiation the deals contain the obligations of the agents and the associated control system based on sanctions.

5 Social goal negotiation

We characterize the allowed deals during goal negotiation as a set of goals which contains for each agent a goal it desires. Moreover, we add two restrictions. First, we only allow goals the agents have the power to achieve. Moreover, we have to consider the existing normative system, which may already contain norms that look after the goals of the agents. We therefore restrict ourselves to new goals. Additional constraints may be added, for example excluding goals an agent can see to itself. However, since such additional restrictions may be unrealistic in some applications (e.g., one may delegate some tasks to a secretary even when one has the power to see to these tasks oneself), we do not consider such additional constraints.

Definition 9 (Deals in goal negotiation). *In the goal negotiation protocol, a deal $d \in$ deals is a set of goals satisfying the following restrictions:*

1. $d \in$ *power*(A)
2. *for all $a \in A$ there exists a $g \in d$ such that*
   (a) $g \in goals(a)$
   (b) there does not exist a norm n in N such that $g \in \cup_{a \in A} O(n, a)$

The following example illustrates a case in which each agent may desire to be alive. In artificial social systems, it would be based on the utility to be alive.

Example 2. Let $MAS = \langle A, G, goals, power, \geq \rangle$ be a multi-agent system with goals $G = \{not\text{-}killed\text{-}by_{a,b} \mid a, b \in A\}$, and $not\text{-}killed\text{-}by_{a,b} \in$ *goals*(a) for all $a, b \in A$. Moreover, let $NMAS = \langle MAS, N, O, V \rangle$ with N the empty set. Then G is a social goal (i.e., an element of *deals*) iff $G \in$ *power*(A).

We could easily further refine our model by defining more abstract goals such as "we have a safe society" and by adding a goal hierarchy reflecting that if some goal is fulfilled, another goal is fulfilled too. For example, if the goal safe-society if fulfilled

then the goals $not\text{-}killed\text{-}by_{a,b}$ are all fulfilled too. However, to keep our model simple, and to focus on the social delegation cycle, we do not do so in this paper.

We now consider a variant of the running example from [5]. Three agents can work together in various ways. They can make a coalition to each perform a task, or they can distribute five tasks among them and obtain an even more desirable social goal.

Example 3. Let $MAS = \langle \{a_1, a_2, a_3\}, \{g_1, g_2, g_3, g_4, g_5\}, \mathit{goals}, \mathit{power}, \geq \rangle$ be a multiagent system, where:

power: $\mathit{power}(a_1) = \{\{g_1\}, \{g_2\}, \{g_3\}\}$, $\mathit{power}(a_2) = \{\{g_2\}, \{g_3\}, \{g_4\}\}$, $\mathit{power}(a_3) = \{\{g_3\}, \{g_4\}, \{g_5\}\}$, if $G_1 \in \mathit{power}(A)$ and $G_2 \in \mathit{power}(B)$ then $G_1 \cup G_2 \in \mathit{power}(A \cup B)$. Agent a_1 has the power to achieve goal g_1, g_2, g_3, agent a_2 has the power to achieve goal g_2, g_3, g_4, and agent a_3 can achieve goal g_3, g_4, and g_5. There are no conflicts among goals.

goals: $\mathit{goals}(a_1) = \{g_4, g_5\}$, $\mathit{goals}(a_2) = \{g_1, g_5\}$, $\mathit{goals}(a_3) = \{g_1, g_2\}$. Each agent desires the tasks it cannot perform itself.

Moreover, let $NMAS = \langle MAS, N, O, V \rangle$ be a normative multiagent system with $N = \{n\}$, $O(n, a_1) = \{g_1\}$. Since there has to be some benefit for agent a_2 and a_3, the goals g_5 and g_2 have to be part of the social goal. Therefore, social goals (i.e., possible deals) are $\{g_2, g_5\}$ and $\{g_2, g_4, g_5\}$.

Finally, consider the negotiation. Assuming agent a_1 is first in the order $\succ$, he will propose $\{g_2, g_4, g_5\}$. The other agents may accept this, or reject it and agent a_2 will them propose $\{g_2, g_5\}$. The latter would be accepted by all agents, as they know that according to the protocol no other proposals can be made.

The example illustrates that the negotiation does not determine the outcome, in the sense that there are multiple outcomes possible. Additional constraints may be added to the negotiation strategy to further delimit the set of possible outcomes.

6 Social norm negotiation

We formalize the allowed deals during norm negotiation as obligations for each agent to see to some goals, such that all goals of the social goal are included. Again, to determine whether the obligations imply the social goal, we have to take the existing normative system into account. We assume that the normative system only creates obligations that can be fulfilled together with the already existing obligations.

Definition 10 (Fulfillable nmas). *A normative multiagent system $\langle MAS, N, O, V \rangle$ extending a multiagent system $MAS = \langle A, G, \mathit{goals}, \mathit{power}, \geq \rangle$ can be fulfilled if there exists a $G' \in \mathit{power}(A)$ such that all obligations are fulfilled $\cup_{n \in N, a \in A} O(n, a) \subseteq G'$.*

Creating a norm entails adding obligations and violations for the norm.

Definition 11 (Add norm). *Let $NMAS$ be a normative multiagent system $\langle MAS, N, O, V \rangle$ extending a multiagent system $MAS = \langle A, G, \mathit{goals}, \mathit{power}, \geq \rangle$. Adding a norm $n \notin N$ with a pair of functions $\langle d_1, d_2 \rangle$ for obligation $d_1 : A \to 2^G$ and for sanction $d_2 : A \to 2^G$ leads to the new normative multiagent system $\langle MAS, N \cup \{n\}, O \cup d_1(n), V \cup d_2(n) \rangle$.*

Moreover, if every agent fulfills it obligation, then the social goal is achieved.

Definition 12 (Deals in goal negotiation). *In the norm negotiation protocol, a deal $d \in$ deals for social goal S is a pair of functions $\langle d_1, d_2 \rangle$ for obligation $d_1 : A \to 2^G$ and for sanction $d_2 : A \to 2^G$ satisfying the following conditions:*

1. *Adding $\langle d_1, d_2 \rangle$ to $NMAS$ for a fresh variable n (i.e., not occurring in N) leads again to a normative multiagent system $NMAS'$;*
2. *$NMAS'$ achieves the social goal, $\cup_{a \in A} d_1(a) = S$.*
3. *If $NMAS$ is fulfillable, then $NMAS'$ is too.*

The following example models the norm not to kill someone else. Thus, this example of the social delegation cycle is an instance of the Kantian categorical imperative: you should not do to others what you don't want them to do to you. Note that we can also represent $alive_a$ as an abbreviation of the conjunction of $not\text{-}killed\text{-}by_{a,b}$ for all agents b when we extend the language with definitions, but this does not change the principle of the social mechanism.

Example 4. Assume $\{not\text{-}killed\text{-}by_{a,b} \mid a \in A\} \subseteq power(b)$ for all agents b. A possible deal for the social goal that there are no murders is that $d_1(b) = \{not\text{-}killed\text{-}by_{a,b} \mid a \in A\}$ for all agents $b \in A$.

The second example illustrates the negotiation protocol.

Example 5. Consider the social goal $\{g_2, g_4, g_5\}$. A possible solution here is that each agent sees to one of the goals. For the social goal $\{g_2, g_5\}$, there will always be one of the agents who does not have to see to any goal.

Sanctions can be added in the obvious way. In the norm negotiation as defined thus far, the need for sanctions has not been formalized yet. For this need, we have to consider the acceptance of norms.

7 Norm acceptance

An agent accepts a norm when it believes that the other agents will fulfill their obligations, and the obligation implies the goals the cycle started with. For the former we use the quasi-stability of the norm (e.g., if the norm is a convention, then we require that the norm is a Nash equilibrium). Each agent b fulfills the norm *given that all other agents fulfill the norm*. Again we have to take the existing normative system into account, so we add the condition that all other norms are fulfilled. In general, it may mean that an agent does something which it does not like to do, but it fears the sanction more than this dislike. We use the trade-off among goals $\geq$.

Definition 13 (Stability). *A choice c of agent $b \in A$ in $NMAS$ with new norm n is $c \in power(b)$ such that $\cup_{m \in N \setminus \{n\}} O(m, b) \subseteq d$. The choices of the other agents are $oc = \cup_{a \in A \setminus \{b\}, m \in N} O(n, a) \cup V(n, a)$. The effect of choice c is $c \cup oc \cup V(n, a)$ if $O(n) \subseteq c$, $c \cup oc$ otherwise. $NMAS$ is stable if $\forall b \in A$, there is a choice c such that $O(n, b) \subseteq c$, and there is no choice $c' \geq (b) c$ with $O(n, b) \not\subseteq c'$.*

Finally, we have to test whether the new situation is better than the old one for all agents. Here we assume that the outcome in both the original multiagent system as in the new multiagent system is a Nash equilibrium, and we demand that each Nash outcome in the new system is better than each Nash outcome in the original normative multiagent system. The formalization of these concepts is along the same lines as the definition of acceptance in Definition 13.

Example 6. The norm in the 'don't kill me' example is quasi-stable, since there is no reason for an agent to divert from the norm. Moreover, it is effective since it sees to his goals of the agents. If we assume that agents minimize the set of goals they see to if it does not affect the priority of the agents, or there are some agents who would like to kill other agents, then a sanction has to be added to make sure that no-one is killed. The priority of the goal not to be sanctioned should be higher than the priority to kill.

8 Concluding remarks

In this paper we introduce a norm negotiation model based on power and dependence structures. It is based on a distinction between social goal negotiation and the negotiation of the obligations with their control system. Roughly, the social goals are the benefits of the new norm for the agents, and the obligations are the costs of the new norm for the agents in the sense that agents risk being sanctioned. Moreover, in particular when representatives of the agents negotiate the social goals and norms, the agents still have to accept the negotiated norms. The norm is accepted when the norm is quasi-stable in the sense that agents will act according to the norm, and effective in the sense that fulfilment of the norm leads to achievement of the agents' desires – i.e., when the benefits outweigh the costs.

Our new model is based on a minimal extension of Tennenholtz' game theoretic model of the social delegation cycle. We add a negotiation protocol, sanction and control, and besides acceptance also effectiveness. It is minimal in the sense that, compared to our earlier model [5] in normative multiagent systems, we do not represent the rule structure of norms, we do not use decision variables, and so on. Also, as discussed in this paper, we do not add goal hierarchy, definitions, etc. The model therefore focusses on various uses of power: the power as ability to achieve goals and in negotiation.

The present paper may be seen as a preliminary study of the expressive power of power and dependence views on multiagent systems. As has been argued by Castelfranchi and Conte for some time, and has been supported by various researchers since then, the power and dependence viewpoint have advantages over classical game theory. However, it remains an open question whether such power structures (or qualitative games in Wooldridge and Dunne's theory) cannot be mapped on classical games by mapping goals on outcomes, power on strategies, and so on. In future research we intend to study under which conditions or assumptions such mappings can be made.

There are several other issues for further research. First, the motivation of our model is to design social mechanisms. Second, we would like to perform formal analysis, like the complexity results obtained for qualitative games [15] or in game-theoretic artificial social systems [12–14]. Third, we like to study more general notions of norm

creation including permission creation and creation of constitutive norms or counts-as conditionals. Fourth, we are interested in the role of coalition formation, contract negotiation, and obligation distribution in the the new norm negotiation model. Finally, we would like to extend the model with the distinction between uncontrollable (or external) and controllable (or police) agents as studied by Brafman and Tennenholtz [20].

References

1. Conte, R., Falcone, R., Sartor, G.: Agents and norms: How to fill the gap? Artificial Intelligence and Law **7(1)** (1999) 1–15
2. Boella, G., van der Torre, L., Verhagen, H.: Introduction to normative multiagent systems. Computational and Mathematical Organizational Theory, Special issue on Normative Multiagent Systems **12(2-3)** (2006) 71–79
3. Boella, G., van der Torre, L.: A game theoretic approach to contracts in multiagent systems. IEEE Transactions on Systems, Man and Cybernetics - Part C **36(1)** (2006) 68–79
4. Boella, G., van der Torre, L.: Security policies for sharing knowledge in virtual communities. IEEE Transactions on Systems, Man and Cybernetics - Part A **36(3)** (2006) 439–450
5. Boella, G., van der Torre, L.: Norm negotiation in multiagent systems. International Journal of cooperative Information Systems (IJCIS) **16(1)** (2007)
6. Conte, R., Castelfranchi, C., Dignum, F.: Autonomous norm-acceptance. In: Intelligent Agents V (ATAL'98). LNAI 1555, Springer (1999) 99–112
7. Brainov, S., Sandholm, T.: Power, dependence and stability in multi-agent plans. In: Procs. of the 21st National Conference on Artificial Intelligence (AAAI'99). (1999) 11–16
8. Castelfranchi, C.: Modeling social action for AI agents. Artificial Intelligence **103(1-2)** (1998) 157–182
9. Conte, R., Sichman, J.: Multi-agent dependence by dependence graphs. In: Procs. of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'02). (2002) 483–490
10. Lopez y Lopez, F.: Social Power and Norms: Impact on agent behaviour. PhD thesis (2003)
11. Castelfranchi, C.: Micro-macro constitution of power. ProtoSociology **18-19** (2003)
12. Shoham, Y., Tennenholtz, M.: On social laws for artificial agent societies: off-line design. Artificial Intelligence **73 (1-2)** (1995) 231 – 252
13. Shoham, Y., Tennenholtz, M.: On the emergence of social conventions: Modeling, analysis and simulations. Artificial Intelligence **94(1–2)** (1997) 139–166
14. Tennenholtz, M.: On stable social laws and qualitative equilibria. Artificial Intelligence **102(1)** (1998) 1–20
15. Wooldridge, M., Dunne, P.: On the computational complexity of qualitative coalitional games. Artificial Intelligence **158(1)** (2004) 27–73
16. Simon, H.: A behavioral model of rational choice. The Quarterly Journal of Economics **69** (1955) 99–118
17. Boella, G., van der Torre, L.: Enforceable social laws. In: Procs. of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'05). (2005) 682–689
18. Boella, G., Sauro, L., van der Torre, L.: Social viewpoints on multiagent systems. In: Procs. of the 3rd International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS'04). (2004) 1358–1359
19. Boella, G., van der Torre, L.: Fair distribution of collective obligations. In: Procs. of the 17th European Conference on Artificial Intelligence (ECAI'06). (2006) 721–722
20. Brafman, R., Tennenholtz, M.: On partially controlled multi-agent systems. Journal of Artificial Intelligence Research (JAIR) **4** (1996) 477–507