The Evolution of Artificial Social Systems

Guido Boella
Dipartimento di Informatica
Università di Torino
Italy
guido@di.unito.it

Leendert van der Torre
CWI Amsterdam
and Delft University of Technology
The Netherlands
torre@cwi.nl

1 Introduction

The basic idea of the artificial social systems approach of Shoham and Tennenholtz [1995; 1997] is to add a mechanism, called a social law, that will minimize the need for both centralized control and on-line resolution of conflicts. A social law is defined as a set of *restrictions* on the agents' activities which allow them enough freedom on the one hand, but at the same time constrain them so that they will not interfere with each other. Several variants have been introduced to reason about the design and emergence of social laws. However, existing models of artificial social systems cannot be used for the evolution of such systems, because these models do not contain an explicit representation of the social laws in force. In this paper we use enforceable social laws [Boella and van der Torre, 2005] to address the question how artificial social systems can be extended to reason about the evolution of artificial social systems.

2 Artificial social systems and social laws

Shoham and Tennenholtz [1995] introduce social laws in a setting without utilities. They define also *rational* social laws [Shoham and Tennenholtz, 1997] as social laws that improve a social game variable. 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. We extend artificial social systems with a control system, called a normative system, to model enforceable social laws. Following Boella and Lesmo [2002], the normative system is represented by a socially constructed agent called the normative agent or agent 0. In [Boella and van der Torre, 2005], the normative system is represented by the set of control strategies of agent 0, but not by a utility function.

Definition 1 *A normative* game *(or a* normative multi-agent encounter*) is a tuple $\langle N, R, S, T, U_1, U_2\rangle$, where $N = \{0, 1, 2\}$ is a set of agents, R, S and T are the sets of strategies available to agents 0, 1 and 2 respectively, and $U_1 : R \times S \times T \to \mathbb{R}$ and $U_2 : R \times S \times T \to \mathbb{R}$ are real-valued utility functions for agents 1 and 2, respectively.*

We use here as game variable the maximin value, following Tennenholtz [2000]. This represents safety level decisions, see Tennenholtz' paper for a motivation.

Definition 2 *Let R, S and T be the sets of strategies available to agent 0, 1 and 2, respectively, and let U_i be the utility function of agent i. Define $U_1(R, s, T) = \min_{r\in R, t\in T} U_1(r, s, t)$ for $s \in S$, and $U_2(R, S, t) = \min_{r\in S, s\in S} U_2(r, s, t)$ for $t \in T$. The* maximin value for agent *1 (*respectively *2) is defined by $\max_{s\in S} U_1(R, s, T)$ (respectively $\max_{t\in T} U_2(R, S, t)$). A strategy of agent i leading to the corresponding maximin value is called a* maximin strategy for agent i.

A social law is useful with respect to an efficiency parameter q if each agent can choose a strategy that guarantees it a payoff of at least q.

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

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).

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

When the set of strategies R of agent 0 is a singleton, then our definitions reduce to those of Tennenholtz [2000]. With the extension of agent 0 representing the control system we define enforceable social laws as quasi-stable social laws in normative games where the strategies of agent 0 may have been restricted [Boella and van der Torre, 2005].

Definition 5 *Given a normative game $g = \langle N, R, S, T, U_1, U_2\rangle$, and an efficiency parameter q, a social law (i.e., a restriction of S to $\overline{S} \subseteq S$, and of $\overline{T} \subseteq T$) is* enforceable *if there is a restriction of R to $\overline{R} \subseteq R$ such that $\overline{S}, \overline{T}$ is quasi-stable in the normative game $g = \langle N, \overline{R}, S, T, U_1, U_2\rangle$.*

Computational problems can be defined to find enforceable social laws (with respect to an efficiency parameter).

3 Representing social laws

We extend normative games with a utility function of agent 0, to represent the norms which are enforced. Since agent 0 is a socially constructed agent, in the sense of Searle [1995], its utility function can be updated. In particular, the enforcement of a social law by $\overline{R} \subseteq R$ is represented by giving $\overline{R}$ strategies a high utility, and $R \setminus \overline{R}$ strategies a low utility. Moreover, we go beyond the framework of enforceable social laws by varying the utility of agent 0 depending on the strategies played by the other agents, and by considering incremental updates of the utility function to represent the evolution of artificial social systems. Formally, we extend a normative game with a utility function $U_0 : R \times S \times T \Rightarrow I\!R$, we define $U_0(r, S, T) = \min_{s \in S, t \in T} U_0(r, s, t)$ for $r \in R$, and we define useful and quasi-stable social laws in the obvious way. Enforced social laws are defined as follows.

Definition 6 *Given a normative game $g = \langle N, R, S, T, U_1, U_2 \rangle$, and an efficiency parameter q, a social law (i.e., a restriction of S to $\overline{S} \subseteq S$, and of $\overline{T} \subseteq T$) is* enforced *if there is a unique restriction of R to $\overline{R} \subseteq R$ such that $\overline{R}, \overline{S}, \overline{T}$ is quasi-stable.*

3.1 Identification of enforced social laws

The game in Table 1 illustrates that the computational problem to find quasi-stable laws corresponds in extended normative games to the identification of enforced social laws. The table should be read as follows. Strategies are represented by literals, i.e., atomic propositions or their negations. Each table represents the sub-game given a strategy of agent 0, represented by $\neg n$ and n, respectively. Agent 1 is playing columns and agent 2 is playing rows. The values in the tables represent the utilities of agent 0 (in italics), 1 and 2.

$\neg n$	p	$\neg p$	n	p	$\neg p$
q	*3*,3,3	*0*,4,1	q	*3*,3,3	*1*,2,1
$\neg q$	*0*,1,4	*1*,2,2	$\neg q$	*1*,1,2	*0*,2,2

Table 1: What is the enforced social law?

Agent 0 (the normative system) can play strategy $\neg n$ or n, agent 1 can play strategy p or $\neg p$, agent 2 can play strategy q or $\neg q$. When the normative system plays $\neg n$, the sub-game of agent 1 and 2 is a classical prisoner's dilemma. Intuitively, the strategy $\neg n$ corresponds to the state before the social law is introduced, and n corresponds to the introduction of a control system that sanctions an agent for deviating from p, q. For example, the utility of agent 1 in $\neg p, q, n$ (2) is lower than its utility in $\neg p, q, \neg n$ (4) due to this sanction.

When the normative system plays n, the agents are always worse off compared to the normative agent playing $\neg n$, all else being equal. Nevertheless, due to the dynamics of the game, the overall outcome is better for both agents. For example, in the sub-game defined by strategy $\neg n$, the only Nash equilibrium is $2, 2$. Now suppose we set the efficiency parameter to 3, which means that all agents will be better off. If the normative system plays n, then the sub-game has a Nash equilibrium which is the (Pareto optimal) $3, 3$. This explains why the agents accept the possibility to be sanctioned.

3.2 Iterated design of enforced social laws

The social law design problem is, given a normative game, to define a new utility function for the normative system. The principle that we like to maintain as much as possible from the existing social laws can be represented by the use of the principle of minimal change. Table 2 represents the evolution of an artificial social system by an incremental increase of the utility of agent 0 to the efficiency parameter of the new social law.

$\neg n_1, \neg n_2$	p_1	p_2	$\neg p_1, \neg p_2$
q_1	*0*,3,3	*0*,4,1	*0*,6,0
q_2	*0*,1,4	*0*,2,2	*0*,0,0
$\neg q_1, \neg q_2$	*0*,0,6	*0*,0,0	*0*,0,0
n_1	p_1	p_2	$\neg p_1, \neg p_2$
q_1	*1*,3,3	*1*,4,1	*1*,0,0
q_2	*1*,1,4	*1*,2,2	*1*,0,0
$\neg q_1, \neg q_2$	*1*,0,0	*1*,0,0	*1*,0,0
n_2	p_1	p_2	$\neg p_1, \neg p_2$
q_1	*3*,3,3	*3*,1,1	*3*,0,0
q_2	*3*,1,1	*3*,0,0	*3*,0,0
$\neg q_1, \neg q_2$	*3*,0,0	*3*,0,0	*3*,0,0

Table 2: Iterated design

The first table represents that the normative system does not impose a control system, the second table represents that there is a sanction for playing $\neg p_1, \neg p_2$ or $\neg q_1, \neg q_2$, and the third table represents that there is an additional sanction for playing something else than p_1 and q_1. The first social law is $\overline{S} = \{p_1, p_2\}, \overline{T} = \{q_1, q_2\}$ based on control system $\overline{R} = \{n_1, n_2\}$, and the second social law is $\overline{S} = \{p_1\}, \overline{T} = \{q_1\}$ based on control system $\overline{R} = \{n_2\}$.

References

[Boella and Lesmo, 2002] G. Boella and L. Lesmo. A game theoretic approach to norms. *Cognitive Science Quarterly*, pages 492–512, 2002.

[Boella and van der Torre, 2005] G. Boella and L. van der Torre. Enforceable social laws. In *Procs. of AAMAS'05*. ACM Press, 2005.

[Searle, 1995] J.R. Searle. *The Construction of Social Reality*. The Free Press, New York, 1995.

[Shoham and Tennenholtz, 1995] Y. Shoham and M. Tennenholtz. On social laws for artificial agent societies: off-line design. *Artificial Intelligence*, 73 (1-2):231 – 252, 1995.

[Shoham and Tennenholtz, 1997] Y. Shoham and M. Tennenholtz. On the emergence of social conventions: modeling, analysis, and simulations. *Artificial Intelligence*, 94 (1-2):139 – 166, 1997.

[Tennenholtz, 2000] M. Tennenholtz. On stable social laws and qualitative equilibria. *Artificial Intelligence*, 102 (1):1–20, 2000.