ELSEVIER

Artificial Intelligence 102 (1998) 1–20

Artificial Intelligence

# On stable social laws and qualitative equilibria [☆]

Moshe Tennenholtz [1]

*Faculty of Industrial Engineering and Management, Technion–Israel Institute of Technology, Haifa 32000, Israel*

Received 6 October 1996; received in revised form 1 January 1998

**Abstract**

This paper introduces and investigates the notion of qualitative equilibria, or *stable social laws*, in the context of qualitative decision making. Previous work in qualitative decision theory has used the *maximin* decision criterion for modelling qualitative decision making. When several decision-makers share a common environment, a corresponding notion of equilibrium can be defined. This notion can be associated with the concept of a *stable social law.* This paper initiates a basic study of stable social laws; in particular, it discusses the stability benefits one obtains from using social laws rather than simple conventions, the existence of stable social laws under various assumptions, the computation of stable social laws, and the representation of stable social laws in a graph-theoretic framework. © 1998 Elsevier Science B.V. All rights reserved.

*Keywords:* Social laws; Qualitative decision making

## 1. Introduction

General coordination mechanisms are essential tools for efficient reasoning in multi-agent AI systems. Coordination mechanisms are a major issue of study in the fields of mathematical economics and game theory as well. Much work in these fields concentrates on the notion of an *equilibrium.* An equilibrium is a joint behavior of agents, where it is irrational for each agent to deviate from that behavior. The notion of an equilibrium discussed in the game theory and mathematical economics literature refers to agents which are expected utility maximizers. However, much work in AI has been concerned with more qualitative forms of rational decision making. In particular, work in AI has been

[☆] A preliminary version of this paper appears in the Proceedings of the 5th International Conference on Knowledge Representation and Reasoning.

[1] Email: moshet@ie.technion.ac.il.

PII: S0004-3702(98)00033-2

concerned with agents which attempt to maximize their worst case payoff. Although, at first, this behavior may look questionable from a decision-theoretic perspective, it is known to capture the behavior of risk-averse agents [9,19,38], and it is appropriate in the context of qualitative decision theory [8,15,19,59]. Moreover, in [9] Brafman and Tennenholtz have shown general conditions under which an agent can be viewed *as if* it were a *maximin* agent (i.e., an agent which maximizes its worst case payoff). However, the corresponding notion of equilibrium has not yet been investigated. In this paper we introduce this notion and investigate its properties. The concept of qualitative equilibrium turns out to coincide with the notion of a *stable social law*, to be introduced later in this paper. For ease of exposition we introduce the notion of a stable social law in a self-contained fashion, as an extension to previous work on *artificial social systems*.

Some work on multi-agent systems assumes that agents are controlled by a single entity which dictates their behavior at each point in time, while some other work is concerned with decentralized systems where no global controller exists. A significant part of the theory developed for decentralized multi-agent systems [7,17] deals with conflict resolution in multi-agent encounters. The basic theme of work on this subject is that in decentralized systems agents will reach states of conflict and appropriate negotiation mechanisms would be needed in order to resolve these conflicts. The result of the negotiation process is a deal that the agents will follow. Work in AI has been mostly concerned with agents that conform to agreed-upon deals. Agents may not follow irrational negotiation protocols, but will conform to deals obtained by following rational negotiation protocols [20,37,62].[2] This differs from work in game-theory [25,48] where a joint strategy is considered unstable (and therefore unsatisfactory from a design perspective) if an agent has a rational incentive to deviate from it. The Artificial Social Systems approach (e.g., [45,57]) exposes a spectrum between a totally centralized approach and a totally decentralized approach to coordination. The basic idea of the Artificial Social Systems approach is to add a mechanism, called a social law, that will minimize the need for both centralized control and on-line resolution of conflicts. In a mobile robots setting, for example, such a social law may consist of various traffic constraints [56]. More generally, a social law is a set of *restrictions* on the agents' activities which allow them enough freedom on one hand, but at the same time constrain them so that they will not interfere with each other. In particular, a social law makes certain conflicts unreachable, and as a result improves the system efficiency. Notice that mechanisms for conflict resolution can serve as part of the social law; they will be used in situations where conflicts can't be prevented in advance.

The motivation for the theory of artificial social systems has been the design of artificial multi-agent systems, and as such it assumes that the agents will obey the law supplied by the designer. However, if each agent is designed by a different designer then some laws might be considered irrational. Therefore, at the current stage, the artificial social systems approach and approaches to conflict resolution are somewhat complementary; the resolution of conflicts in multi-agent encounters is part of a more general theory of social laws, but the theory of artificial social systems has neglected the stability of social laws in

[2] See [53] for a detailed discussion of this point.

multi-agent encounters. In this paper we wish to bridge part of the gap between the theory of artificial social systems and the theory of conflict resolution in multi-agent encounters, by considering *stable social laws* for multi-agent encounters. A social law for a multi-agent encounter is a restriction of the set of available actions (in the encounter) to a set of socially allowed actions. Stable social laws make deviation from them irrational. Notice that a convention is a particular type of a social law; a convention determines a particular joint action for the agents to follow (e.g., keep the right of the road), while a social law allows several such actions and prohibits others. As it turns out, this distinction is quite important and useful.

We will discuss social laws for multi-agent encounters using a game-theoretic framework which is tailored for assumptions made in the AI literature, and especially in recent work on qualitative decision making. In particular, in most of this paper we will assume that the agents are risk-averse agents, which use the *maximin decision criterion.*

More specifically, given a set of possible behaviors of the other agents, the aim of an agent is to optimize its worst case outcome assuming the other agents may follow any of these behaviors. This kind of behavior is appropriate where there is some ordinal relation on possible outcomes. In such situations all that matters to agents is the order of payoffs and not their exact value. The precise conditions under which such modelling is an appropriate one are discussed in [9]. Moreover, as we discuss in Section 4, this modeling perspective is quite natural in many multi-agent systems, where a payoff corresponds to the achievement of a particular user's specification. We will require that a social law suggested for a particular encounter will guarantee to the agents a certain payoff, and that it will be stable; there should be no incentive to deviate from it assuming the agents are risk-averse agents. Hence, a stable social law corresponds to a notion of qualitative equilibrium for risk-averse agents. In a later stage, we show how our discussion and results can be extended to other basic qualitative decision-making settings.

We start by introducing our framework. In particular, in Section 3 we define the notion of stable social laws. In Section 4 we discuss the intuition and formal adequacy of maximin in multi-agent systems. Having the basic framework, in Section 5 we show that the set of multi-agent encounters for which a stable convention exists is a strict subset of the set of multi-agent encounters for which there is an appropriate stable social law; however, we show that there exists situations where no stable social law exists. Then, in Section 6 we initiate a computational study of stable social laws; we formulate the corresponding computational problem and show that the general problem of coming up with a stable social law is intractable; the proof of our result sheds light on the structure of stable social laws; in addition, we point to an interesting restriction on our framework under which the synthesis of stable social laws is polynomial. We then return back to the question of the existence of stable social laws; in Section 7 we first show how this question can be formulated in standard graph-theoretic terms, and then expose a class of encounters where simple graph-theoretic conditions imply the existence of stable social laws. In Section 8 we discuss how our ideas can be applied in other qualitative decision making contexts, and in Section 9 we further discuss the meaning of our study and results and the connection of our work to the existing literature.

## 2. The basic framework

In this section we introduce our basic framework, which is built upon a basic game-theoretic model.

### 2.1. The basic model

In general AI planning systems, agents are assumed to perform *conditional plans.*[3] A conditional plan is a (perhaps partial) function from the local state of an agent to action. Conditional plans can be treated as *protocols* in distributed systems terms, or as *strategies* in game-theoretic terms. In the sequel we will make use of a game-theoretic model; therefore, we will adopt the term *strategy*.

Multi-agent encounters can be represented as a game. In this paper we will consider two-person games, where two agents participate in an encounter.[4] We will be concerned with finite games where each agent has a finite number of strategies. A *joint strategy* for the agents consists of a pair of strategies, one for each agent. Each joint strategy is associated with a certain payoff for each of the agents, as determined by their *utility functions*. The above-mentioned terms are classical game-theoretic terms which capture general multi-agent encounters.

Formally, we have:

**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 $U_1 : S \times T \rightarrow \mathbb{R}$ and $U_2 : S \times T \rightarrow \mathbb{R}$ are utility functions for agents 1 and 2, respectively.

One interesting point refers to the knowledge of the agents about the structure of the game. In this work, we assume that agents are familiar with the sets of actions available to the different agents, but an agent might be aware only of its own payoff function. Our results are appropriate both for the case where the payoff functions are common-knowledge among the agents and for the case an agent knows only its individual payoff function.

How should agents behave in a multi-agent encounter prescribed by a given game? The system's designer may wish to guarantee to the agents a particular payoff. This guarantee is crucial for agents who are committed to obtain particular goals and to successfully perform particular tasks on behalf of their users in a multi-agent system. This brings us to the modeling of agents as maximin reasoners (see the discussion in Section 4). In Section 8 we show how our results can be extended to other basic settings of qualitative decision making.

**Definition 2.** Let $S$ and $T$ be the sets of strategies available to agent 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 $s \in S$,

[3] Plans with complete information and other forms of plans will be taken as restrictions on the general form of plans considered in this paper; this point will not affect the discussion or results presented in this paper.

[4] The concepts defined in this paper can be easily extended to the case of $k \geqslant 2$ agents, where $k$ is a constant.

and $U_2(S, s) = \min_{t \in S} U_2(t, s)$ for $s \in T$. 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*.

### 2.2. Conventions and social laws

As we have mentioned an agent may wish to guarantee that a particular specification or user's requirements are indeed obtained. The achievement of a particular set of requirements will be associated with a particular payoff. The system designer (e.g., the government in certain applications, the system administrator or other mediators on other applications) can specify a social law that will enable the agents to obtain "reasonable" payoffs. One way of capturing such reasonable behavior is by requiring that the payoff for each agent will be at least $t$.[5]

Given a game and a requirement that the agents will be able to obtain a payoff of at least $t$, the designer may supply the agents with an appropriate convention: a joint strategy for which the utility for both agents is greater than or equal to $t$. A convention is a special case of a social law. A social law in a multi-agent encounter is a restriction on the set of strategies available to the agents; a convention will simply restrict the behavior to a one particular joint strategy. The role of the designer is to select a law that allows each agent at least one strategy which guarantees a payoff of at least $t$.

**Definition 3.** Given a game $g = \langle N, S, T, U_1, U_2 \rangle$ and an efficiency parameter $t$, 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 hold: there exists $s \in \overline{S}$ such that $U_1(s, \overline{T}) \geqslant t$, and there exists $k \in \overline{T}$ such that $U_2(\overline{S}, k) \geqslant t$. A (useful) convention is a (useful) social law where $|\overline{S}| = |\overline{T}| = 1$.

In general, a useful social law is a restriction on each agent's activities which enables each agent to act individually and succeed reasonably well, as long as all the agents conform to the law (see the discussion and the general semantics in [46,57]). At this point the idea of using social laws for coordinating agents' activities in a multi-agent encounter may seem a bit strange; why should we care about social laws if every efficiency degree which can be obtained by a social law can already be obtained by an appropriate simple convention? However, as we will later see, social laws can serve as much more useful entities than simple conventions for agents participating in a multi-agent encounter.

## 3. Stable social laws

The concept of social laws which has been discussed in previous work defines a general methodology for the design of multi-agent systems. In this work we are concerned with social laws for multi-agent encounters. Although that's a most popular setting for the study of the resolution of conflicts, up to date the power of social laws has been illustrated in

[5] The case where different thresholds are associated with different agents can be treated similarly.

more complex settings [56,57]. As we shall see, social laws may serve as useful tools for multi-agent encounters as well.

**Definition 4.** Given a game $g = \langle N, S, T, U_1, U_2 \rangle$ 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 $s' \in S - \overline{S}$ 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 $U_2(\overline{S}, t') > \max_{t \in \overline{T}}\{U_2(\overline{S}, t)\}$.

Our definition is in the spirit of classical game theory; we require that deviation by one agent will be irrational given that the other agent sticks to the suggested behavior. Notice that irrationality here refers to the notion of maximin behavior. Indeed, if an agent's aim is to guarantee the fulfillment of his commitment to the user of obtaining payoff $k$, then it is irrational for him to deviate to a strategy that might risk this guaranteed performance.[6] Hence, a quasi-stable social law will make a deviation from the social law irrational as long as the other agent obeys the law. However, the above definition of stability may not be satisfactory in our context. In a multi-agent encounter an agent has specific goals to obtain, and there is no reason to assume an agent will execute a strategy which yields to it a payoff which is lower than the payoff guaranteed to it by another strategy, assuming the other agent obeys the law. Putting it in other terms, given that we talk about a specific encounter with specific goals, there is no reason to include in the set of allowed strategies a strategy which is (maximin) dominated by another strategy in that set. This requirement is consistent with models of stable social situations discussed in the game theory literature [31]. Therefore we have:

**Definition 5.** A quasi-stable social law is a *stable social law* if the payoff guaranteed to each of the agents is independent of the strategy (conforming to the law) it selects, as long as the other agent conforms to the social law (i.e., selects strategies allowed by the law). Namely, given a game $g = \langle N, S, T, U_1, U_2 \rangle$ and an efficiency parameter $q$, the quasi-stable social law that restricts the agents to $\overline{S}$ and $\overline{T}$ respectively is a *stable social law* if: for all $s_1, s_2 \in \overline{S}$, and $t_1, t_2 \in \overline{T}$, we have that $U_1(s_1, t_1) = U_1(s_2, t_2)$ and $U_2(s_1, t_1) = U_2(s_2, t_2)$.

Notice that a stable social law is an equilibrium concept which one may wish to associate with the *maximin* decision criterion. Hence, our study of stable social laws can be interpreted as a basic study of equilibria in the context of qualitative decision making. Our discussion and results can be extended to qualitative decision making contexts in which the agents adopt decision criteria which are different from *maximin*. We further discuss this point in Section 8.

In the rest of this paper we discuss stable social laws.[7] Given a multi-agent encounter, a stable social law will guarantee to the agents a particular payoff, similarly to the way a particular payoff can be guaranteed by a simple convention. As it turns out however, the

[6] See also the discussion in Section 4.

[7] Similar results can be obtained when we consider quasi-stable social laws.

difference between social laws and conventions stems from the fact that social laws may be more stable than conventions in multi-agent encounters. This will be the topic of Section 5.

## 4. A perspective on safety level decisions

The notion of equilibrium is of major importance in Economics. Although this concept is still a major controversial issue (see [2]), it is the central solution concept discussed in the Economics literature. The concept of social laws is quite natural for many domains, and has been studied and applied in other works [4,5,46,47,57]. We believe however that the concept of stable social laws, defined in this paper, should be further discussed before a study of some of its properties is presented. A major concern one may have is the applicability of the maximin (safety level) kind of reasoning to multi-agent domains. Although we believe the study presented in this paper is of considerable importance due to the fact it complements previous work carried out in game theory and due to the fact it bridges some of the gap between the theory of social laws and the theory of conflict resolution, the applicability of the related concepts should be also discussed.

Assume a system consists of several robots each of which is controlled by a different authority/programmer (as in Stanford's Gofer project [13]). Each such authority/programmer may be hired in order to obtain various tasks for different users/companies. A similar situation arises when several software houses compete in a market, where they perform tasks on behalf of their clients. More generally, an agent needs to satisfy a user's specification for a product, where the specification consists of a set of goals/tasks that should be obtained. This is the situation in current Automated Guided Vehicles systems, in current software projects, etc. In such systems a guarantee for a particular delivery is required. Assume there are $n$ possible goals one may consider, a contract will specify $k \leqslant n$ of them that an agent will (and must) obtain. The payment for the agent is a function of the guaranteed delivery. The above set of situations exist in many systems and can be best captured by maximin reasoning/analysis in multi-agent systems. In the related settings one may interpret a payoff as the number of goals that the agent obtains when the agents follow particular strategies. The maximin value is the number of goals that the agent can guarantee to obtain and that it will be willing to commit on achieving them. Expected value and Bayesian reasoning are rarely used in the related systems, although these systems might not be purely competitive. Social laws that restrict agents' behavior in a deterministic fashion are indeed used in the related settings.

The above kind of analysis is quite intuitive, but a reader may attack it based on the interpretation we give to payoffs. Consider an interpretation as the one mentioned above, or a similar one (e.g., one may refer to the negation of the number of possible fails of a processor, as the payoff in a certain hardware system, etc.), the payoff function might not capture utility in standard economic terms. This concern however can be addressed by paying attention to the foundations of decision theory. Utilities are entities which are ascribed to an agent based on its actions. The fundamental work of Savage [54] serves as the justification for the use of maximal expected utility. In his work Savage supplies a set of conditions on an agent's choice among actions, under which it can be viewed as if it had probabilities, utilities, and it had used expected utility maximization for action selection.

Therefore, from a formal perspective, there is no meaning to utilities without mentioning the decision criterion that is used. Indeed, Brafman and Tennenholtz [9,10] have shown sound and complete conditions under which an agent can be viewed as if it had a utility function under which it used the maximin decision criterion. Hence, one should be careful before dismissing maximin or other decision criteria based on the formal grounds supplied in previous work (see the discussion at [10,11]). On the other hand, as illustrated above, intuitive meanings for payoffs under which maximin reasoning is appropriate, exist for many interesting multi-agent domains. We will elaborate on the use of maximin and safety level decision making in the discussion session.

## 5. Social laws versus conventions

Having a definition of stable social laws, one may ask: what are these laws good for? If we wish to guarantee a certain payoff for the agents, why can not we look for stable conventions, i.e., select a joint strategy from which a deviation would be irrational, assuming such a strategy exists?

The answer is supplied by the following result:

**Theorem 6.** *There exists games for which there are no stable conventions, but where appropriate stable social laws do exist.*

**Proof.** The proof follows by considering the following game:

| | agent 2 | | | |
|---|---|---|---|---|
| agent 1 | $A$ | $B$ | $C$ | $D$ |
| $A$ | (1,1) | (1,1) | (2,0) | (0,2) |
| $B$ | (1,1) | (1,1) | (0,2) | (2,0) |
| $C$ | (2,0) | (0,2) | (0.5,0.5) | (0.75,0.25) |
| $D$ | (0,2) | (2,0) | (0.25,0.75) | (0.5,0.5) |

Assume the designer wishes to guarantee the payoff 1. The fact that no stable convention exists follows by case analysis. If the agents are required to perform $(A, A)$ (i.e., each agent is required to perform $A$) or are required to perform $(A, B)$ (i.e., agent 1 is required to perform $A$ and agent 2 is required to perform $B$), then agent 2 may improve upon its situation by performing $D$; If the agents are required to perform $(B, B)$ or are required to perform $(B, A)$ then agent 2 can improve its situation by performing $C$. All other potential conventions are not useful, since at least one of the agents obtains a payoff which is less than 1. On the other hand, if we restrict both agents to perform actions taken from $\{A, B\}$ then a payoff of 1 is guaranteed for both of the agents and no deviation is rational. If agent 2 performs $C$ (respectively $D$) then it may lose if agent 1 has chosen to perform $A$ (respectively $B$). If agent 1 performs $C$ (respectively $D$) then it may lose if agent 2 has chosen to perform $B$ (respectively $A$). □

The above theorem reveals a new contribution of the theory of social laws: restricting the activities of the agents to a set of allowed actions rather than to a particular action is useful even in simple multi-agent encounters. This is due to the fact that social laws may be more stable than simple conventions. The intuition behind the above result is as follows. Although different strategies may lead to similar payoffs, different strategies may block different deviations by the agents. Therefore, the fact that the agent's behavior is only partially defined may improve the system efficiency. Assume for example that there are two agents, each of which can invest its money using four options, $A$, $B$, $C$, or $D$. If they will invest only in options $A$ and $B$ then they will get reasonable payoffs. However, if they are told to invest in particular options, e.g., one is told to invest in $A$ and the other is told to invest in $B$, then one of them may take this opportunity in order to gain more on behalf of the other using option $C$ or $D$. But, if both $C$ and $D$ yield low payoffs when they are applied against $A$ or $B$ (although not against both of them), such deviation can be prevented by telling each agent to choose (nondeterministically) from among options $A$ and $B$ (i.e., by supplying the social law: "do not use $C$ and $D$", rather than pointing to particular investments). The reader may get additional understanding of this situation by considering the proof of Theorem 6.

We have shown that social laws are more stable than conventions. We can also show:

**Theorem 7.** *There exist games for which no stable social laws exist, for any selection of the efficiency parameter.*

**Proof.** The proof follows by considering the following game:

| | agent 2 | |
|---|---|---|
| agent 1 | $A$ | $B$ |
| $A$ | (2.5,1) | (1.5,3) |
| $B$ | (2,2) | (4,0.5) |

A case analysis shows that no stable social law exists in the above-mentioned game. If both agents are required to perform $A$ then agent 2 can improve upon its situation by performing $B$. If both agents are required to perform $B$ then agent 2 can improve upon its situation by performing $A$. If agent 1 is required to perform $A$ and agent 2 is required to perform $B$ then agent 1 will improve its situation by adopting $B$. If agent 1 is required to perform $B$ and agent 2 is required to perform $A$ then agent 1 will improve its situation by performing $A$. Given that all of the payoffs in the matrix are different, it is clear that no law where an agent is required to perform either $A$ or $B$ (i.e., where both actions are allowed) will be stable. Notice that allowing both $A$ and $B$ is not stable, since one of the allowed actions will become more desirable than the other.[8]

Combining the above we get that no stable social law exists. □

[8] The question of whether desirable behaviors can be obtained by social laws that are not stable according to our definition is beyond the scope of this paper. The type of stable sets we consider somewhat resemble the situation discussed by Shapley when considering the notion of block dominance [55].

Hence, stable social laws are powerful but do not always exist. Given this observation, it may be of interest to supply a procedure for computing when a social law exists. Naturally, in cases where a stable social law exists it may be of interest to compute such a law. In addition, it may be of interest to characterize conditions for the existence of stable social laws. These are the topics of the following sections.

## 6. Computing stable social laws

In this section we take a look at the computation of stable social laws. In order to do so, we first need to decide on the representation of our input. We will use the standard game-theoretic representation in which a multi-agent encounter is represented by a game matrix.

The problem of computing a Stable Social Law (SSLP) is defined as follows:

**Definition 8** (*The Stable Social Law Problem* [*SSLP*]). Given a multi-agent encounter $g$, and an efficiency parameter $t$, find a Stable Social Law which guarantees to the agents a payoff which is greater than or equal to $t$ if such a law exists, and otherwise announce that no such law exists.

Notice that if we restrict ourselves to simple conventions, the computational problem is easy; however, as we have observed, conventions are not as useful as social laws. As the following theorem shows this does not come without a cost. We are able to show:

**Theorem 9.** *The SSLP is NP-complete.*

**Proof.** The proof that the problem is in NP is straightforward. The proof that the problem is NP-hard is by reduction from 3-SAT [27]. Let $\varphi$ be a 3-CNF formula with $m$ clauses. Our reduction will generate a 2-person game $g$, where the strategies for both agents are identical. The set of strategies for each agent is: $c_i^1, c_i^2, \ldots, c_i^7, d_i$ $(1 \leqslant i \leqslant m)$, where each $c_i^k$ is associated with a different truth assignment to clause $i$ (there are seven such assignments), and $d_i$ is an additional distinguished strategy which is associated with clause $i$. We take the efficiency parameter $t$ to be equal to 0, and let $t'$ be a positive real number. We will show that a stable social law for $g$ exists if and only if $\varphi$ is satisfiable.

We take $g$ to be a symmetric game, and specify the utility function of agent 1:

(1) $U_1(d_i, d_j) = -t'$ for all $i$, $j$.
(2) $U_1(c_i^k, c_j^l) = 0$ iff $c_i^k$ and $c_j^l$ correspond to consistent assignments.
(3) $U_1(c_i^k, c_j^l) = -t'$ iff $c_i^k$ and $c_j^l$ correspond to inconsistent assignments.
(4) $U_1(d_i, c_j^k) = t'$ for all $i \neq j$ and every $k$.
(5) $U_1(c_j^k, d_i) = -t'$ for all $i \neq j$ and every $k$.
(6) $U_1(d_i, c_i^k) = -t'$ for every $i$ and every $k$.
(7) $U_1(c_i^k, d_i) = t'$ for every $i$ and every $k$.

Now, consider a truth assignment $T$ which satisfies $\varphi$. We can define a social law which leaves each agent only with the strategies which their corresponding assignments are as determined by $T$ (and with no strategy of the form $d_i$). It is easy to see that we get a stable social law; the social law guarantees a payoff of 0 since the agents are left only with "consistent strategies", and deviations are irrational since there is a representative strategy of the form $c_i^k$ for each clause.

If there exists a stable social law then it can not leave the agents with strategies of the form $d_i$. This is due to the fact that when such actions are performed a payoff that is lower than 0 might be obtained. In addition, such a law must leave each agent with exactly one strategy for each clause. At least one strategy for each clause is required since otherwise a deviation to some $d_j$ will become rational. If more than one strategy is allowed for each clause then the agents may execute "inconsistent strategies" (which lead to negative payoffs). The allowed strategies need to be consistent (with respect to their corresponding assignments), since otherwise a payoff lower than 0 is possible. Hence, by combining the allowed strategies (i.e., their corresponding truth assignments) we get a satisfying assignment. □

The importance of the above theorem is twofold: first, it supplies an initial result in the computational study of stable social laws. Second, the proof of this result sheds additional light on the structure of stable social laws. As one can see, we need to restrict the behavior of agents, but still leave them with enough freedom for blocking deviations by the other agents. This observation is complementary to the observations made in [46] about the role of social laws. In [46] the authors refer to the *golden-mean* problem in *artificial social systems*, where the designer needs to restrict the behavior of agents in order that they won't interfere with one another but leave them with enough freedom for obtaining socially acceptable goals. The setting described in this paper refers to multi-agent encounters and not to general artificial social systems, but augment the discussion with the concept of stability; As we have explained, the introduction of stability explores another aspect of the golden-mean problem. Further understanding of this structure is obtained in the following section, where we supply a graph-theoretic representation of stable social laws.

Notice that in many situations the system's input has a much more succinct representation than the one discussed in the previous theorem. This points to the importance of the above hardness result. Indeed, it is not straightforward to show that the design of social laws in the framework of strategic-form representations is intractable [46]. Nevertheless, the above kind of representation has been found useful for modeling many multi-agent interactions [25]. Hence, given the previous theorem, it would be of interest to identify general cases where the problem of coming up with a stable social law (if such a law exists) is tractable. One case which is of interest is when the parties involved are of unequal power. One way of capturing this fact is by assuming that one party has many more strategies available to it than the other party does. Formally, we say that an agent is *logarithmically bounded* if the number of strategies available to it is $O(\log(n))$ where $n$ is the number of strategies available to the other agent. In this case we can show:

**Theorem 10.** *The SSLP when one of the agents is logarithmically bounded is polynomial.*[9]

**Proof.** Without loss of generality let agent 1 be the logarithmically bounded agent. We can efficiently enumerate the set of possible restrictions on its strategies since there are only polynomially many such possibilities. For each such restriction $r$, let us denote the set of nonprohibited strategies by $S_1(r)$. Given $S_1(r)$ we can gather the set of strategies of the other agent (i.e., agent 2) which guarantee a payoff greater than or equal to $t$ (where $t$ is the efficiency parameter) for agent 2 and exclude from them the ones that are dominated by other strategies of that agent (2). Let us denote this set of strategies by $S_2(r)$. Now, if there are strategies in $S_1(r)$ that are better than other strategies in $S_1(r)$ or if there exists a strategy in $S_1(r)$ which does not guarantee a payoff of $t$ (given the previously generated set of strategies for agent 2) then we should move and try a new restriction $r'$ on the strategies of agent 1. If that's not the case then we need to check whether there is a strategy for one of the agents which is not included in $S_1(r)$ and $S_2(r)$ respectively, and may yield a better payoff for the respective agent than what is guaranteed under $S_1(r)$ and $S_2(r)$. If there is such a deviation then we should try another $r'$ (if exists) and otherwise we should stop (an appropriate law has been found).

The above procedure exhausts in a systematic manner all possible stable social laws since each possible restriction on the behavior of agent 1 is checked, and for each such restriction the most general restriction on the second agent's behavior which still may be possible is generated. This enumeration procedure is polynomial since agent 1 has only $O(\log(n))$ strategies. Checking stability of a given set of restrictions is polynomial; all that we need to do in to compute for every action the worst case payoff which might be obtained when the other agent obeys the law; then we can compare these worst case payoffs to the payoff guaranteed by the law, in order to detect whether a rational deviation exists. □

## 7. Graph-theoretic representations of stable social laws

We can learn about the structure of stable social laws by studying the reduction used in Theorem 9. More generally, the study of a new equilibrium concept and of its use can greatly benefit from representation theorems which show what does this concept mean in terms of known concepts. In addition, in the context of this particular work, such representation theorems can supply conditions for the existence of stable social laws. The reduction used in the proof of Theorem 9 shows that a *special case* of the problem of finding a stable social law is isomorphic to a well-known problem. This has been useful for proving the above-mentioned result. However, it would be of interest to characterize the general Stable Social Laws concept by means of well-known terminology. In particular, in this section we make use of graph-theoretic terms in order to characterize the stable social law concept.

[9] The results presented in the previous section and Theorem 9 can be easily extended to the case of $k \geqslant 2$ agents. The technique used in the proof of Theorem 10 can be used also for the case $k \geqslant 3$ if there is only one party which is more powerful than the others.

We will make use of the following standard terms:

**Definition 11.** Let $G = (V, E)$ be a *graph*, where $V$ is a set of *nodes*, and $E \subseteq V^2$ is a set of *edges*. $G$ is *undirected* if, for all $v_1, v_2 \in V$, $(v_1, v_2) \in E$ iff $(v_2, v_1) \in E$, and is *directed* otherwise. A set $V' \subseteq V$ is an *independent set* if there are no $v', v'' \in V'$ which satisfy $(v', v'') \in E$. A set $V' \subseteq V$ is a *clique* if $(v', v'') \in E$ for all $v', v'' \in V'$. A node $v \in V$ is *nonisolated* relative to $V' \subseteq V$ if there is a vertex $v' \in V'$ such that $(v, v') \in E$. A set $V' \subseteq V$ is a *dominating set* if for each node $v' \in V - V'$ there is a node $v'' \in V'$ such that $(v', v'') \in E$. A node $v \in V$ is a *sink* if there is no $v'$ such that $(v, v') \in E$. The graph $G$ is *k-colorable* if we can color the nodes of the graph with $k$ colors in a way that $(v, v') \in E$ implies that $v$ and $v'$ have different colors.

We would now like to make a connection between the above-mentioned graph-theoretic terms and our notion of a stable social law. In the sequel we will be concerned with games where the sets of strategies, $S$, available to the agents are identical. We will also assume the game is symmetric in the sense that $U_1(s,t) = U_2(t,s)$ (i.e., the outcome of the agents is independent of their names). We will be interested in social laws that are fair, in the sense that if a strategy is prohibited for one agent then it is prohibited for all agents. For ease of exposition we will be concerned with social laws guaranteeing the value $t$ and no more than $t$.

**Definition 12.** Given a game $g$ and an efficiency parameter $t$, let $G_1 = (V, E_1)$, $G_2 = (V, E_2)$, $G_3 = (V, E_3)$ be directed graphs where $V$ is associated with the set of strategies $S$, and $E_i$ is defined as follows: $(s,q) \in E_1$ iff $U_1(s,q) \geqslant t$; $(s,q) \in E_2$ iff $U_1(s,q) = t$; $(s,q) \in E_3$ iff $U_1(s,q) \leqslant t$.

Given the above-mentioned graphs which are built based on the game $g$ and the efficiency parameter $t$, we can show the connection between stable social laws and standard graph-theoretic concepts:

**Theorem 13.** *Given a game $g$ and an efficiency parameter $t$, the corresponding graphs $G_1, G_2, G_3$ satisfy the following: a stable social law for $g$ exists iff there is a subset $V'$ of the nodes of the related graphs, such that $V'$ is a clique in $G_1$, a dominating set in $G_3$, and all nodes in $V'$ are nonisolated, relative to $V'$, in $G_2$.*

**Proof.** Assume that $V'$ satisfying the above properties exists; one can easily check that by prohibiting all strategies in $V - V'$ we get a stable social law. The efficiency is guaranteed by the requirement from $G_1$, and the fact that no deviation is rational is guaranteed by the requirement from $G_3$. The fact that no allowed action can be ignored is guaranteed by the requirement from $G_2$.

If there exists a stable social law then a payoff greater than or equal to $t$ should be guaranteed regardless of the (allowed) actions selected; this implies that the nodes associated with the allowed actions constitute a clique in $G_1$. Similarly, since no deviation is rational these nodes should correspond to a dominating set in $G_3$. In addition, since there is no reason to consider behaviors which are inferior to others in a stable social law we get that no node which corresponds to an allowed strategy would be isolated in $G_2$. □

The above theorem supplies an additional graph-theoretic understanding of the notion of stable social laws. A further look at such representations enables us to prove additional general existence theorems for stable social laws. One interesting general type of multi-agent encounters refers to games which are a combination of pure coordination and zero-sum games. The importance of such type of games is obvious; they allow agents either to agree and obtain "reasonable payoff" or to "fight" for "high payoff" taking the risk of obtaining "low payoff". These basic games are formally defined as follows:

**Definition 14.** Assuming without loss of generality that the efficiency parameter $t$ equals 0, a symmetric game $g$ is a *mixed coordination-competition game*, if the utility functions satisfy:

(1) $U_1(s,s) = 0$ for every $s \in S$.
(2) $U_1(s,q) > 0$ iff $U_1(q,s) < 0$ for every $s, q \in S$.

An interesting point about mixed coordination-competition games is that they can be represented by a single graph, $\bar{G}$, which is defined as follows: the nodes of $\bar{G} = (\bar{V}, \bar{E})$ correspond to the different strategies, and the set of edges $\bar{E}$ is defined as follows: $(s,t) \in \bar{E}$ iff $U_1(s,t) < 0$. Given this graph structure, we can prove the existence of stable social laws for an interesting class of encounters:

**Theorem 15.** *Given a mixed coordination-competition game $g$, if the corresponding graph $\bar{G}$ has a sink or is 2-colorable then an appropriate stable social law exists.*

**Proof.** If there is a sink in the graph then we can choose the corresponding strategy as a convention (i.e., both agents will be required to play only the corresponding strategy). Otherwise, if the graph is 2-colorable then we can color the graph by *red* and *blue* and prohibit all (and only) red strategies (for both agents). Clearly, two blue strategies will yield the desired payoff since the graph is 2-colorable. No deviation to red strategy is rational since the graph has no sinks and neighbors of a red strategy should be blue (i.e., a deviation may result in a negative payoff). □

## 8. Other qualitative equilibria

The previous sections have been concerned with qualitative equilibria, where the decision criterion is the *maximin* decision criterion. As we have mentioned before, similar observations and results do hold for other decision criteria as well. In this section we take a look at two other basic decision criteria, the *minimax regret* decision criterion, and the *competitive ratio* decision criterion, and show how our previous study can be adapted to the context of these decision criteria as well.

**Definition 16.** Let $S_i$ be a set of strategies available to agent $i$, and let $u_i$ be the utility function of agent $i$. Given $s \in S_1$, and $q \in S_2$, define

$$u_1(s, q, S_1) = \max_{t \in S_1} u_1(t, q) - u_1(s, q).$$

Given $q \in S_1$ and $s \in S_2$ define

$$u_2(q, s, S_2) = \max_{t \in S_2} u_2(q, t) - u_2(q, s).$$

The *minimax regret value for agent* 1 (*respectively* 2) is defined by $\min_{s \in S_1} \max_{q \in S_2} u_1(s, q, S_1)$ (respectively $\min_{t \in S_2} \max_{q \in S_1} u_2(q, t, S_2)$). A strategy of agent $i$ leading to the corresponding minimax regret value is called a *minimax regret strategy for agent* $i$.

The minimax regret decision criterion is a basic decision criterion [38,41] which captures the following idea. Given a set of available strategies for agent $i$, if the other agent, $j$, would have known the actual strategy to be performed by agent $i$ then it could choose a corresponding optimal strategy. The amount of regret of agent $j$ when performing a strategy $s_j$ when agent $i$ performs a strategy $s_i$ is the lost obtained by performing $s_j$ instead of performing the optimal strategy against $s_i$ available to agent $j$. For each strategy of agent $j$ one can compute the maximal regret this agent may have while performing this strategy; the strategy which minimizes the maximal regret is the minimax regret strategy. The intuition behind this decision rule is that the agent would not like to lose much relative to the case where it would have known the other agent's action.

A related decision rule, the *competitive ratio* decision rule, which is popular in the theoretical computer science literature [49], and which has been recently discussed in the AI literature [44], is similar to the minimax regret decision rule; in this decision rule we consider the ratio between the payoff obtained by a particular strategy to the payoff obtained by the corresponding optimal one, instead of considering the difference between these payoffs.

The definition of a useful social law remains as in the previous section, but the definition of quasi-stable social laws and of stable social laws will be based on minimax regret or competitive ratio respectively. Hence, an agent may deviate from the prescribed social law if it has a strategy which leads to a minimax regret (or to a competitive ratio) value which is lower than the one obtained by conforming to the law. The rationale of the related settings is that the designer may wish to guarantee a particular payoff for the agents, but an agent may not be risk-averse and might use decision criteria such as the minimax regret or the competitive ratio while selecting its action.

As it turns out, social laws are more stable than simple conventions also when the agents adopt the minimax regret or the competitive ratio decision rules. Consider the game matrix of the proof of Theorem 6. If the agents are required to perform the action $A$, then the minimax regret value for an agent who considers deviating is obtained by the action $C$. Similar results are obtained whenever each agent is required to stick to a particular action. However, if the agents are allowed to perform "only $A$ or $B$" then the maximal regret values of both $C$ and $D$ (2) are lower than the maximal regret values of $A$ and $B$ (1), and we get a stable social law. A similar result can be obtained for the case of the competitive ratio decision criterion.

The above discussion shows that the basic results obtained in the context of qualitative equilibria for risk-averse agents, can be obtained also in other contexts of qualitative decision-making. As it turns out, the computational results obtained in the case of maximin can also be extended to the case of minimax regret and competitive ratio. The key idea which enables to extend Theorem 9 and its proof to the case of *minimax regret* and *competitive ratio* is the following one. For ease of exposition, we present it for the case of *minimax regret*; the case of *competitive ratio* is treated similarly. The payoffs in the proof of Theorem 9 can be either 0 (i.e., satisfactory), $t$ (i.e., high), and $-t$ (i.e., low). In order to have a similar proof in the context of minimax regret we will change the negative payoffs from $-t$ to $-2t$ (i.e., very low). Given this modification, the minimax regret of strategies which are not part of the law mentioned in the proof of Theorem 9 will be at least $2t$, while by conforming to the law the agents will have a regret of at most $t$; this modification is needed in order to guarantee the stability of the laws prescribed in our reduction. The other details of the reduction and proof of Theorem 9 remain as in the case of *maximin*. A modification of the proof of Theorem 10 to the context of minimax regret is quite straightforward as well. In this case we still enumerate the possible restrictions on the strategies of agent 1, but when collecting the strategies of agent 2 we have to be careful to gather only the strategies which minimize the maximal regret of agent 2 given the corresponding restriction on the strategies of agent 1. Having this observation, Theorem 10 can be proved for the case of minimax regret as well.

## 9. Discussion

In this work we have introduced a theory of stable social laws, or qualitative equilibria, for qualitative decision makers. Our work bridges some of the gap between work on Artificial Social Systems and work on conflict resolution in game theory and AI. Social laws have been shown to be a basic and useful tool for the coordination of multi-agent systems [4,5,12,42,45,47,57]. However, the stability of social laws in a system of rational agents has been neglected so far. This work extends previous work on social laws for artificial agent societies by considering stable social laws for multi-agent encounters.

Two major lines of research related to our work are work in the field of game theory and work in the field of Distributed Artificial Intelligence [DAI]. Related work on rational deals and negotiations in DAI (e.g., [37,51,62]) can be viewed as contributions to game theory. A very interesting property of this work is that it considers deals among rational agents who will not deviate from agreed-upon deals. [10] In difference to this assumption, our work is concerned with agents who will deviate from agreed-upon deals if they have a rational incentive to do so. The safety level kind of reasoning/analysis is not new of course to computer science. This is also true for its (somewhat implicit) use in multi-agent systems. In fact, work in software engineering has used a social laws paradigm for the imposition of protocols in distributed systems [42,43]. The idea in this work is that a desired behavior of an agent should be guaranteed regardless of the behavior of other agents (obeying the corresponding law). Naturally, the different specifications which may be satisfied (i.e., the

[10] This is not to say that other assumptions are not treated by the DAI literature; see, for example [53].

sets of goals/tasks which may be obtained) by each agent might be of different quality, which can be captured by corresponding payoffs. In this case the performance guarantee, that is a crucial factor in computerized systems, can be captured by maximin-like analysis.

Much work in game theory has been concerned with devising rational conventions for a group of rational agents; a rational agent may deviate from a prescribed joint strategy if this deviation will improve its own situation. More specifically, much work in game theory [25,38,48] has been devoted to the study of equilibrium in games; an equilibrium will have the property that there is no rational incentive for an agent to deviate from the equilibrium as long as other agents stick to it. The notion of an equilibrium has been adopted to the AI and DAI literature in various settings (e.g., [61]), as part of a general and important attempt to introduce social and organizational metaphors into the AI context [14,16,18,21,24,28,35,36,39,58]. A central notion in this regard is the notion of a rational agent adopted from the decision/game theory literature. Most work in game theory has associated the notion of a rational agent with the notion of expected utility maximization. This is not however the usual way a rational agent is viewed in the AI literature, such as in work on conditional planning [22,29,50,52,60]. Moreover, much of the field of knowledge representation and reasoning is concerned with qualitative notions of decision making, which are different from expected utility maximization; this is true for work in belief revision [26], nonmonotonic reasoning [30], qualitative physics [23], as well as for work on qualitative decision theory [8,11,15,19,38,59]. Work on safety level reasoning/analysis gets increasing attention in mainstream game theory in the recent years. Several works, including [3,6,32,40], and more recently [1,34], are concerned with long-run solution concepts, in which agents attempt to get closer to a safety level of a game. In the related work an agent is not able to observe the utility function of the other agents (although he may be familiar with the related strategies). Although one way to address the related issue is by using the theory of types developed by Harsanyi [33], the use of safety level analysis in such settings is a natural approach. Notice that our work may fit nicely also with the assumption that each agent knows only its utility function.

One may wish to explore the connections between the notion of stable social laws and the notion of mixed strategies (and equilibria in mixed strategies) discussed in game theory. However, a direct comparison between these concepts seems quite problematic. This is due to the fact that the interpretations of the utility function for maximin agents and for expected utility maximizers are different (see the discussion is Section 4, as well as [10]). Notice that in our context the actual selection among the strategies available to an agent is modelled as a nondeterministic choice rather than a probabilistic one. The nondeterminism refers to parameters which might be unknown to the designer and to the other agents and which may change from time to time. One may however wish to find some functional connections between mixed strategies and stable social laws, e.g., a stable social law may be the support of a mixed strategy equilibrium. Unfortunately, we were not able to prove or disprove such a claim. On the other hand, the idea of considering sets of strategies rather than a mixed strategy in equilibrium is consistent with the theory of social situations [31]. The restriction to a set of strategies (i.e., prohibiting some of the strategies) rather than pointing to a probabilistic mixture is a basic idea in the work of Minsky on social rules in the context of software engineering [42,43]. Needless to say that the enforcement of a social law is much more practical than the enforcement of a mixed strategy, due to the fact that coin flippings can not be usually observed by an outside observer.

Using a game-theoretic terminology, in this work we developed an equilibrium theory for qualitative decision makers, and in particular for risk-averse agents [9,19,38]. Our theory and results can therefore be interpreted both as an extension to the theory of Social Laws presented in the AI literature, as well as a contribution to the foundations of discrete/qualitative decision/game theory. We hope it can lead to further cross-fertilization between these fields.

## Acknowledgements

I wish to thank the participants of the inter-disciplinary game theory and multi-agent systems seminar at the Technion, especially Dov Monderer and Ron Holzman, for useful discussions on the concepts discussed in this paper and their interpretation. Special thanks also to Nathan Linial and Amir Pnueli for their comments.

## References

[1] P. Auer, N. Cesa-Bianchi, Y. Freund, R.E. Schapire, Gambling in a rigged casino: the adversial multi-armed bandit problem, in: Proceedings 36th Annual Symposium on Foundations of Computer Science, 1995, pp. 322–331.

[2] R. Aumann, Perspectives on bounded rationality, in: Proceedings 4th Conference on Theoretical Aspects of Reasoning about Knowledge, Pacific Grove, CA, 1991, pp. 108–117.

[3] A. Banos, On pseudo games, The Annals of Mathematical Statistics 39 (1968) 1932–1945.

[4] O. Ben-Yitzhak, M. Tennenholtz, On the synthesis of social laws for mobile robots: a study in artificial social systems (Part I), Computers and Artificial Intelligence 14 (1997).

[5] O. Ben-Yitzhak, M. Tennenholtz, On the synthesis of social laws for mobile robots: a study in artificial social systems (Part II), Computers and Artificial Intelligence, to appear.

[6] D. Blackwell, An analog of the minimax theorem for vector payoffs, Pacific J. Mathematic 6 (1956) 1–8.

[7] A.H. Bond, L. Gasser, Readings in Distributed Artificial Intelligence, Ablex Publishing Corporation, Norwood, NJ, 1988.

[8] C. Boutilier, Toward a logic for qualitative decision theory, in: Proceedings 4th International Conference on Principles of Knowledge Representation and Reasoning, Bonn, Germany, 1994, pp. 75–86.

[9] R. Brafman, M. Tennenholtz, On the foundations of qualitative decision theory, in: Proceedings AAAI-96, Portland, OR, 1996.

[10] R. Brafman, M. Tennenholtz, Axiom systems for qualitative decision criteria, in: Proceedings AAAI-97, Providence, RI, 1997.

[11] R. Brafman, M. Tennenholtz, Modeling agents as qualitative decision-makers, Artificial Intelligence 94 (1997), 217–268.

[12] W. Briggs, D. Cook, Flexible social laws, in: Proceedings 14th International Joint Conference on Artificial Intelligence (IJCAI-95), Montreal, Que., 1995, pp. 688–693.

[13] P. Caloud, W. Choi, J.-C Latombe, C. Le Pape, M. Yim, Indoor automation with many mobile robots, in: Proceedings IEEE International Workshop on Intelligent Robots and Systems, Tsuchiura, Japan, 1990.

[14] P.R. Cohen, H.J. Levesque, Teamwork, Nous 25 (4) (1991).

[15] A. Darwiche, M. Goldszmidt, On the relation between kappa calculus and probabilistic reasoning, in: Proceedings 10th Conference on Uncertainty in Artificial Intelligence (UAI-94), Seattle, WA, 1994, pp. 145–153.

[16] R. Davis, R.G. Smith, Negotiation as a metaphor for distributed problem solving, Artificial Intelligence 20 (1) (1983) 63–109.

[17] Y. Demazeau, J.P. Muller, Decentralized AI, North-Holland/Elsevier, 1990.

[18] J. Doyle, A society of mind: multiple perspectives, reasoned assumptions, and virtual copies, in: Proceedings 8th International Joint Conference on Artificial Intelligence (IJCAI-83), Karlsruhe, Germany, 1983, pp. 309–314.

[19] D. Dubois, H. Prade, Possibility theory as a basis for qualitative decision theory, in: Proceedings 14th International Joint Conference on Artificial Intelligence (IJCAI-95), Montreal, Que., 1995, pp. 1924–1930.

[20] E.H. Durfee, J. Lee, P.J. Gmytrasiewicz, Overeager reciprocal rationality and mixed strategy equilibiria, in: Proceedings AAAI-93, Washington, DC, 1993, pp. 225–230.

[21] E.H. Durfee, V.R. Lesser, D.D. Corkill, Coherent cooperation among communicating problem solvers, IEEE Trans. Comput. 36 (1987) 1275–1291.

[22] O. Etzioni, S. Hanks, D. Weld, D. Draper, N. Lesh, M. Williamson, An approach to planning with incomplete information, in: Proceedings 3rd Conference on Principles of Knowledge Representation and Reasoning (KR-92), Cambridge, MA, 1992, pp. 115–125.

[23] K.D. Forbus, Qualitative process theory, Artificial Intelligence 24 (1984) 85–168; also in: D. Bobrow (Ed.), Qualitative Reasoning about Physical Systems, The MIT Press, 1985.

[24] M.S. Fox, An organizational view of distributed systems, IEEE Trans. Systems Man Cybernet. 11 (1981) 70–80.

[25] D. Fudenberg, J. Tirole, Game Theory, MIT Press, 1991.

[26] P. Gardenfors, Belief Revision, Cambridge University Press, 1992.

[27] M. Garey, D. Johnson, Computers and Intractability—A Guide to the Theory of NP-Completeness, W.H. Freeman and Company, 1979.

[28] L. Gasser, Social knowledge and social action: Heterogeneity in practice, in: Proceedings 13th International Joint Conference on Artificial Intelligence (IJCAI-93), Chambery, France, 1993, pp. 751–757.

[29] M.R. Genesereth, I.R. Nourbakhsh, Time saving tips for problem solving with incomplete information, in: Proceedings AAAI-93, Washington, DC, 1993.

[30] M.L. Ginsberg (Ed.), Readings in Nonmonotonic Reasoning, Morgan Kaufmann, 1987.

[31] J. Greenberg, The Theory of Social Situations: An Alternative Game-Theoretic Approach, Cambridge University Press, 1990.

[32] J. Hannan, Approximation to bayes risk in repeated play, in: M. Dresher, A.W. Tucker, P. Wolfe (Eds.), Contributions to the Theory of Games, Vol. III, Annals of Mathematics Studies 39, Princeton University Press, 1957, pp. 97–139.

[33] J.C. Harsanyi, Games with incomplete information played by bayesian players, parts i, ii, iii, Management Science 14 (1967) 159–182.

[34] S. Hart, A. Mas-Colell, A simple adaptive procedure leading to correlated equilibrium, Discussion paper 126, Center for Rationality and Interactive Decision Theory, Hebrew University, 1997.

[35] T. Ishida, M. Yokoo, L. Gasser, An organizational approach to adaptive production systems, in: Proceedings AAAI-90, Boston, MA, pp. 52–58.

[36] N.R. Jennings, Controlling cooperative problem solving in industrial multi-agent systems using joint intentions, Artificial Intelligence 75 (1995) 195–240.

[37] S. Kraus, J. Wilkenfeld, The function of time in cooperative negotiations, in: Proceedings AAAI-91, Anaheim, CA, pp. 179–184.

[38] R.D. Luce, H. Raiffa, Games and Decisions—Introduction and Critical Survey, John Wiley, 1957.

[39] T.W. Malone, Modeling coordination in organizations and markets, Management Science 33(10) (1987) 1317–1332.

[40] N. Megiddo, On repeated games with incomplete information played by nonbayesian players, Internat. J. Game Theory 9 (1980) 157–167.

[41] J. Milnor, Games against nature, in: R.M. Thrall, C.H. Coombs, R.L. Davis (Eds.), Decision Processes, John Wiley, 1954.

[42] N.H. Minsky, The imposition of protocols over open distributed systems, IEEE Trans. Software Engineering 17 (2) (1991) 183–195.

[43] N.H. Minsky, Law-governed systems, Software Engineering J. (September 1991) 285–302.

[44] D. Monderer, M. Tennenholtz, Dynamic nonbayesian decision-making, J. Artif. Intell. Res. 7 (1997) 231–248.

[45] Y. Moses, M. Tennenholtz, Artificial social systems, Part I: basic principles, Technical Report CS90-12, Weizmann Institute (1990).

[46] Y. Moses, M. Tennenholtz, Artificial social systems, Computers and Artificial Intelligence 14 (6) (1995) 533–562.

[47] S. Onn, M. Tennenholtz, Determination of social laws for agent mobilization, Artificial Intelligence 95 (1997) 155–167.

[48] G. Owen, Game Theory, 2nd ed., Academic Press, 1982.

[49] C.H. Papadimitriou, M. Yannakakis, Shortest paths without a map, in: Proceedings 16th International Colloquium on Automata, Languages and Programming, 1989, pp. 610–620.

[50] M.A. Peot, D.E. Smith, Conditional nonlinear planning, in: Proceedings 1st International Conference on AI Planning Systems, 1992, pp. 189–197.

[51] J.S. Rosenschein, M.R. Genesereth, Deals among rational agents, in: Proceedings 9th International Joint Conference on Artificial Intelligence (IJCAI-85), Los Angeles, CA, 1985, pp. 91–99.

[52] S. Safra, M. Tennenholtz, On planning while learning, J. Artif. Intell. Res. 2 (1994) 111–129.

[53] T.W. Sandholm, V.R. Lesser, Equilibrium analysis of the possibilities of unenforced exchange in multiagent systems, in: Proceedings 14th International Joint Conference on Artificial Intelligence (IJCAI-95), Montreal, Que., 1995, pp. 694–701.

[54] L.J. Savage, The Foundations of Statistics, John Wiley, New York, 1954. Revised and enlarged edition, Dover, New York, 1972.

[55] L. Shapley, Game theory, Lecture Notes, Mathematics Dept., UCLA, 1990.

[56] Y. Shoham, M. Tennenholtz, On traffic laws for mobile robots, in: Proceedings 1st Conference on AI Planning Systems (AIPS-92), 1992.

[57] Y. Shoham, M. Tennenholtz, Social laws for artificial agent societies: off-line design, Artificial Intelligence 73 (1995) 231–252.

[58] H.A. Simon, The Sciences of the Artificial, The MIT Press, 1981.

[59] S.W. Tan, J. Pearl, Specification and evaluation of preferences under uncertainty, in: Proceedings 4th International Conference on Principles of Knowledge Representation and Reasoning (KR-94), Bonn, Germany, 1994, pp. 530–539.

[60] D.H.D. Warren, Generating conditional plans and programs, in: Proceedings Summer Conference on AI and Simulation of Behavior, Edinburgh, 1976.

[61] M.P. Wellman, A market-oriented programming environment and its application to distributed multicommodity flow problems, J. Artif. Intell. Res. 1 (1993) 1–23.

[62] G. Zlotkin, J.S. Rosenschein, A domain theory for task oriented negotiation, in: Proceedings 13th International Joint Conference on Artificial Intelligence (IJCAI-93), Chambery, France, 1993, pp. 416–422.