Mechanism Design for Resource Bounded Agents

Noa E. Kfir-Dahav Dov Monderer Moshe Tennenholtz
Faculty of Industrial Engineering and Management
Technion – Israel Institute of Technology
Haifa 32000, Israel

January 19, 1999

Abstract

The theory of mechanism design deals with the design of protocols for non-cooperative multi-agent systems. A major task of this theory is the design of protocols that will maximize the social welfare of the agents. An ideal mechanism will optimize social welfare and will be strategy-proof, i.e. the dominant strategy of each agent will be to participate in the mechanism and to reveal his true goal and worth, as well as budget-balanced, i.e., the mechanism should not impose any payments from the center/organizer to the agents. Indeed, the Clarke's mechanism, which is central to information economics and to games with incomplete information satisfies these properties. However, we show that the Clarke's mechanism employs the use of procedures for optimizing social welfare, which are NP-hard. Hence, these procedures should be replaced by heuristics. We present a set of natural properties (axioms) of such heuristics that, when satisfied, enable to obtain the desired strategy-proofness and budget balance properties. Our result enables to extend the central protocol of the theory of mechanism design to the context of resource-bounded agents.

1 Introduction

Recent work in Artificial Intelligence has dealt extensively with economic and game-theoretic mechanisms (e.g. [12, 9, 14, 17, 11]). The theory of mechanism design deals with the design of protocols to be used in non-cooperative environments. In many cases, the objective of these protocols is the optimization of social welfare. The idea is that a center takes care of an interaction in which the agents have different worth functions over a set of possible states. The center has full control on the choice of the state, and its goal is to determine an efficient state, where a state is efficient (in the economics' sense) if there does not exist another state which is weakly preferred by all agents and strictly preferred by at least one agent. If the worth functions of the agents are known to the center, it can find an efficient state by optimizing an appropriate social welfare function. The most common such function is the sum of the worth functions of the agents. However, in many environments, the center is ignorant about the worth functions of the agents. In such cases the center wishes to design a protocol that extract the missing information from the agents. A mechanism is a protocol that allocates a state according to inputs given by the agents. An input may be a worth function or any other signal which is correlated with the worth function. Each agent is assumed to be a selfish entity (i.e., a human or an artificial agent programmed and designed by him/her) wishing to maximize its own utility. Such an agent may exercise its sovereignty by not sending any input or by sending a false input.

A mechanism in which a dominant strategy for each agent is to send its true input is called a strategy-proof mechanism. In a strategy-proof mechanism each agent is participating and telling the truth because it will be worth off if it does not do so, independent of the other agents' choice of actions. A strategy-proof mechanism that selects a state, which maximizes the sum of the worth functions of the agents is called strongly efficient. All economic mechanisms use monetary transfers in order to obtain a desired behavior. A mechanism is budget-balanced if the total transfer from the agents to the center is nonnegative, i.e., in a budget-balanced mechanism, the center does not subsidize the agents in order to obtain the desired result.

The goal of some of the theory of information economics and mechanism design [10], as well as of some of the theory of games with incomplete information [6], is to find a strongly efficient mechanism, which is strategy-proof and budget balanced. Although many negative results exist in the related literature (see [6]), a central positive result has been obtained. In a very general context, as the one we will present in this paper, strongly efficient, strategy-proof, and budget balanced mechanism exists. This mechanism is the famous Clarke's mechanism [3].

Work on multi-agent systems in AI [1, 4, 12, 2] share with work in information economics similar concerns, and therefore there is no doubt that the introduction of the Clarke's mechanism is of fundamental importance to AI too [5]. One major issue that work in AI and Computer Science need to address when applying the Clarke's mechanism is concerned with the fact that the participants (the agents and the center) are resource bounded. Discussions on resource bounds in the context of mechanism design are rare. Exceptions include work on resource bounds in the context of auctions [13] and in the context of coalition formation [15]. In [5], Ephrati and Rosenschein deal with the communication burden in the context of the Clarke's mechanism. However, as we show, there is a major obstacle in applying the Clarke's mechanism even in very simple settings. Formally, this obstacle is captured by the fact that the problem of optimizing social welfare (when the utilities and goals are known) is NP-hard. Given that the Clarke's mechanism (as well as all other related mechanisms [8, 7, 16]) is based on procedures for optimizing social welfare, this result leads to a fundamental problem.

The above problem can not be simply neglected. We need to address the problem of obtaining desired behavior without guaranteeing efficient outcome for each possible input. Indeed, the best one can hope for is the use of clever heuristics or approximation algorithms in order to try and approach the optimal social welfare. However, this leaves us with a central problem, which one must address in order to deal with resource-bounded mechanism design: given heuristics for optimizing social welfare, can we still obtain the desired strategy-proofness and budget balance properties? In this paper we answer this question positively, and in a constructive manner.

The approach we take is as follows. We present several properties, termed axioms, that one may wish any heuristic for optimizing social welfare to have. Then, we use the original Clarke's mechanism where whenever we need to compute optimal social welfare we use the result of an appropriate heuristic instead. As we show, when the above properties are satisfied, the resulting mechanism is strategy-proof and budget balanced. Hence, we obtain a generalization of the Clarke's mechanism, which extends its use to the context of resource-bounded agents.

Although our axioms are natural properties one may wish any heuristic for optimizing social welfare to have, one might argue that it may not be easy to construct an algorithm that satisfies our axioms, We show that in various variants of our model any heuristic h that does not satisfy our axioms can be (constructively) replaced by a better heuristic h' that does satisfy them.

The paper is structured as follows. The basic model of group decision-making under incomplete information is presented in Section 2. In Section 3 we discuss the Clarke's mechanism, and we show that it makes use of NP-hard procedures for optimizing social welfare. In Section 4 we present a set of natural properties (axioms) that one may wish any (heuristic, approximation) algorithm for optimizing social welfare to have. In Section 5 we introduce the heuristic Clarke mechanism where procedures for optimizing social welfare are replaced by a corresponding heuristic. We show that any heuristic Clarke mechanism which makes use of a heuristic that satisfies the axioms is strategy proof and budget balanced. In Section 6 we discuss the issue of existence of heuristics that satisfy the axioms, and present the idea (whose detailed algorithmic aspects are presented in the full paper) of transforming any given heuristic to a desirable heuristic, i.e., a one that satisfies the axioms, without decreasing the efficiency level.

2 The Model - Definitions

Consider a set of agents, $N = \{1, \ldots, n\}$, who are functioning in a given shared environment. In addition, there is a center who is taken to be under the designer's control.[1] Assume that the environment can be described by a state which is a vector of k components. Let $K = \{1, \ldots, k\}$, and let S_j be the set of possible values of component j.

[1]As in the theory of mechanism design in economics, we look at the system from the perspective of the system's designer, where the center is the representative of the designer in the system.

Definition 1 *A* state *(of the world) is a tuple $s = (s_1, \ldots, s_k)$, such that for each j, $s_j \in S_j$. The set of all states is $S = S_1 \times \ldots \times S_k$.*

Definition 2 *For each $a \in S_j$ let $p_{j,a}$ be a primitive proposition meaning "the value of component j is a". LA will be the closure of $\{p_{j,a} : j \in K, a \in S_j\}$ under $\wedge, \vee$, and $\neg$. A state $s = (s_1, \ldots, s_k) \in S$ satisfies a primitive proposition $p_{j,a}$ if and only if $s_j = a$. The satisfaction of a formula $\varphi \in LA$ by a state $s \in S$ (denoted by $s \models \varphi$) is defined as in standard propositional calculus. The set of all satisfiable formulas is denoted by L.*

Each agent is characterized by a goal, which is a satisfiable formula and by a positive monetary worth it assigns to achieving its goal:

Definition 3 *A* goal *of agent i is a satisfiable formula $\varphi^i \in L$. Let $v^i_{goal} > 0$ be the* worth *agent i assigns to achieving its goal. The pair (φ^i, v^i_{goal}) is called the* type *of i. Let $T = L \times \Re_{++}$ be the set of all possible types, where $\Re_{++}$ denotes the set of all positive real numbers. Agent i's* worth function *(sometimes called valuation function), associated with its type $\tau^i = (\varphi^i, v^i_{goal})$, is denoted by $v^i = v^{\tau^i}$, where $v^i : S \to \Re_+$ is such that $v^i(s) = v^i_{goal}$ if $s \models \varphi^i$ and $v^i(s) = 0$ otherwise. The set of all possible worth functions is denoted by V.*

It is obvious that V equals the set of all functions $w = \alpha 1_A$, where $\alpha > 0$, $\emptyset \neq A \subseteq S$, and 1_A is the characteristic function of A. For $v^i \in V$, $v^i(s)$ is interpreted as the amount agent i is willing to pay if the state of the world is changed to s. That is, we assume that the agents' worth functions do not depend on the initial state. It is obvious that in order to describe v^i all we need to know is τ^i. However, the mapping $(\varphi^i, v^i_{goal}) \to v^i$ is not one to one. Given an initial state of the world, the center can change it to any other state. Such a change may incur a state-dependent cost, which is normalized to zero in this paper. This normalization is innocent and all our results can be adapted to the non-normalized case. We do not discuss in this paper the interesting case in which the state can be changed only by a joint series of operations performed by all (or some) of the agents (including the center).

The center's goal is to find a state that maximizes the social welfare according to the designer's criterion. It achieves this goal by maximizing an appropriate social welfare function of the agents' types, where any function $q : T^n \to R$ is considered as a legitimate social welfare function. The theory of mechanism design in economics assumes that the center does not know the true types of the agents and therefore it uses a mechanism to extract this information:

Definition 4 *A* mechanism *(for the set of agents $N = \{1, \ldots, n\}$) is a pair (M, f), where M is a set of messages which includes the null message, and f is an $n+1$ tuple of functions $(d, t_1, \ldots, t_n)$ with $d : M^n \times S \to S$ and $t_i : M^n \times S \to \Re$, $i \in N$. The functions t_i, $i \in N$ are assumed to satisfy the following property: If $m_i = null$, $t_i(m) = 0$.*

The set M is the set of messages each agent can choose from. Choosing the *null* message means sending no message, i.e. not participating in the decision process. If the agents send the vector of messages $m = (m_1 \ldots, m_n)$ and the initial state is s, then $d(m, s)$ will be the final state, and every agent i will transfer $t_i(m, s)$ to the center. An agent who chose any message except the *null* message will be called an *active agent*. An agent who chose the *null* message will be called an *inactive agent*. An active agent commits itself to the mechanism's rules. That is, he may have to transfer money to the center. If the agents in N send messages $m = (m_1, \ldots, m_n)$ then the utility of agent i will be $u^i(m, s) = v^i(d(m, s)) - t_i(m, s)$,
where v^i is the valuation function of i.

In this paper we will mainly focus on anonymous mechanisms:

Definition 5 *A mechanism (M, f) is* anonymous *if a permutation over the agents' messages will lead to an identical permutation of the transfers and to the same final state. That is, for every one-to-one function $\pi : N \to N$, $d(m^\pi, s) = d(m, s)$ and $t_{\pi(i)}(m^\pi, s) = t_i(m, s)$, for every $i \in N$, $m \in M^n$ and $s \in S$, where for $m \in M^n$, m^π is the vector of messages in M^n for which $m^\pi_{\pi(i)} = m_i$ for every $i \in N$.*

A special kind of mechanisms are the direct mechanisms. In a direct mechanism each agent is asked to declare its type:

Definition 6 *A* direct mechanism *is a mechanism (M, f) in which $M = T \cup \{null\}$.*

Notice that in a direct mechanism, an agent can try to manipulate the final state and transfers by declaring a false type.

An agent chooses a message based on its strategy. A strategy of an agent is a function which determines, based on the agent's type and the initial state, which message to send:

Definition 7 *A* strategy *for agent i in the mechanism (M, f) is a function $b^i : T \times S \to M$. The set of all strategies of i is denoted by B^i.*

Note that $B^i = B^j$ for all $i, j \in N$. We denote each of the sets B^i, $i \in N$, by B.

3 The Clarke's mechanism and its complexity

Making a social choice using a mechanism is a well studied subject in game theory and information economics [6, 10]. As explained in the Introduction, several desired properties of such mechanisms are discussed in the literature and are satisfied by the Clarke's mechanism. We now define these properties as well as the Clarke's mechanism.

Definition 8 *Let (M, f) be a mechanism. A strategy $b^i \in B$ is a* dominant strategy *for agent i if for every $\tau_i \in T$, for every $s \in S$, and for every $b'^i \in B$, $u^i(b^i(\tau^i, s), m_{-i}, s) \geq u^i(b'^i(\tau^i, s), m_{-i}, s)$ for every $m_{-i} \in M^{n-1}$.*

Definition 9 *A mechanism (M, f) is* strongly efficient *if every agent i has at least one dominant strategy, and if for every n-tuple of dominant strategies $b = (b_1, \ldots, b_n) \in B^n$, for every $\tau \in T^n$, and for every $s' \in S$, $s^* = d((b^i(\tau^i, s'))_{i \in N}, s')$ maximizes $\sum_{i \in N} v^i(s)$ over $s \in S$, where $v^i = v^{\tau^i}$.*

Definition 10 *A mechanism (M, f) is* budget balanced *if for every $m \in M^n$, and every $s \in S$,*
$\sum_{i \in N} t_i(m, s) \geq 0$.

Definition 11 *A direct mechanism* (M, f), *is* strategy-proof *if participation and truth-telling is a dominant strategy for each agent. That is,* b^i *is a dominant strategy for every agent* i, *where* $b^i(\tau^i, s) = \tau^i$ *for every* $\tau^i \in T$ *and every* $s \in S$.

Before we describe the Clarke mechanism, we need a few notations. The vector $m \in M^n$ for which $m_i = null$ for all $i \in N$ is denoted by m^0. For every $m \in M^n$ we denote by $N(m)$ the set of all active agents defined by m. That is, $i \in N(m)$ if and only if $m_i \neq null$.

Definition 12 *A* Clarke's mechanism *is a direct mechanism* (M, f), *in which the following three properties hold for every* $s \in S$:

1. $d(m^0, s) = s$, *and for* $m \neq m^0$, $d(m, s)$ *is a state that maximizes* $\Sigma_{j \in N(m)} w^j(s')$ *over* $s' \in S$, *where* w^j *is the worth function associated with* m_j.

2. For every agent i *and for every* $m \in M^n$ *with* $m_i \neq null$ *for which there exists* $j \neq i$ *with* $m_j \neq null$, $t_i(m, s) = max_{s \in S} \sum_{j \neq i, j \in N(m)} w^j(s) - \sum_{j \neq i, j \in N(m)} w^j(s^*)$, *where* w^i *is the worth function associated with* m_i, *and* $s^* = d(m, s)$.

3. For every agent i *and for every* $m \in M^n$ *for which* $m_i \neq null$ *and* $m_j = null$ *for all* $j \neq i$, $t_i(m, s) = 0$.

Theorem 1 *A Clarke's mechanism is an anonymous, strongly efficient, budget balanced, strategy-proof mechanism.*

The proof of the above theorem can be derived from [3].

Notice that in general one may face a problem of obtaining the agents' worth functions from their types. In the general setting that we discuss in this paper this mapping is non-problematic. In particular, the computation of maximal social welfare can take the types of the agents as its input, and the corresponding algorithms need not necessarily transform the agents' types into complete (naive) representation of their utility functions.

If we want to use a Clarke's mechanism we would need to compute $s' \in argmax_{s \in S} \sum_{i \in N} v^i(s)$, and for each $i \in N$ we would need to compute $s' \in argmax_{s \in S} \sum_{j \neq i} v^j(s)$. Hence, we are interested in the following problem:

Definition 13 The optimal social welfare problem *[OSWP]: Given a set* N' *of* n' *agents, and a vector of types*
$(\tau^i)_{i\in N'} = ((\varphi^i, v^i_{goal}))_{i\in N'}$,
find $s' \in argmax_{s\in S} \sum_{i\in N'} v^i(s)$*, where* $v^i = v^{\tau^i}$.

Given that OSWP is a basic ingredient in the Clarke's mechanism, as well as in any other mechanism for optimizing social welfare with complete and with incomplete information, we may wish to understand the potential use of it for the case of resource-bounded agents. The proof of the following theorem is relatively standard. It appears in the full version.

Theorem 2 *OSWP is NP-hard.*

4 Heuristics for the Optimization of Social Welfare

In the previous section we have shown that the problem of computing a state which maximizes the sum of worth over S is computationally hard. This suggests a major problem and challenge to the application and adaptation of the Clarke's mechanism to computational contexts. The approach we suggest to adopt is the following one. Since OSWP is hard, we need to make use of some heuristics or approximations for its solution. The question is whether mechanisms that are constructed with heuristics for OSWP will have the desired properties of the Clarke's mechanism (excluding efficiency of course). Hence, the idea is that we can not use computations of optimal social welfare, but only heuristics which attempt to approach that value. We would like to look for natural properties of heuristics that if hold, then the other desired properties of the Clarke's mechanism are still satisfied! In this section we provide such properties– Axioms 1–3. In the next section we prove that a mechanism which is defined with a heuristic that satisfies Axiom 1–3 is strategy-proof and budget balanced.

Notice that although we use the term heuristic, common in the AI literature, we wish to emphasize that our treatment and our results can be applied to any non-optimal algorithm h which stops after polynomial time, that one may use to solve the OSWP. We take the inputs of the algorithm h to be $M = T \cup \{null\}$. Hence, as is typically assumed in the theory of computation, we assume that h is uniform, in the sense that it can be used for inputs of different length.

Definition 14 *A* heuristic *is a function* $h : M^n \setminus \{m^0\} \to S$. *The* social worth function *associated with a heuristic* h *is the function* $h_{sw} : M^n \setminus \{m^0\} \to R$ *defined by:* $h_{sw}(m) = \sum_{i \in N(m)} w^i(h(m))$, *where* $w^i = w^{m_i}$ *for every* $i \in N(m)$.

We now describe three natural properties that a heuristic h can have. The issue of existence of heuristics that satisfy these properties is discussed in Section 6. We wish to emphasize that the following axioms refer to properties one may wish an algorithm for optimizing social welfare under complete information (i.e. for solving the OSWP) to have. In particular, $N(m)$ will refer to the non-null elements in the input tuple m. We use the term "the message sent by agent i" only for convenience; technically, the message sent by agent i is the i-th input of the heuristic h. The axioms below refer to natural algorithmic properties one may wish a heuristic for optimizing social welfare under complete information to have, and the way h treats its inputs has a-priori nothing to do with the fact we have in the background a setting for decision-making with incomplete information.

Axiom 1:
For every $m \in M^n$, $m \neq m^0$, there exists $j \in N$ with $m_j \neq null$, such that $h(m) \models \psi^j$, where $m_j = (\psi^j, v^j_{goal})$.

If h satisfies Axiom 1, then h satisfies at least one of the goals given in its input. In particular, a heuristic h for solving the OSWP should be able to obtain more than the 0 outcome (e.g. by selecting a state that satisfies at least one goal; recall we assume that all goals are individually satisfiable).

Axiom 2:
For every $m, m' \in M^n \setminus \{m^0\}$, for which there exists a unique input $i \in N(m)$ such that $m'_i \neq m_i$,

$$h_{sw}(m) \geq \sum_{i \in N(m)} w^i(h(m')),$$

where w^i is the worth function associated with m_i (and not with m'_i).

If h satisfies Axiom 2, then a unilateral change in one of the input elements cannot increase the approximate social welfare defined by h_{sw} (for the original input). Hence, if the heuristic h for solving OSWP will run instead of on the input m, on an input that differs from m on one element, this will not improve its outcome with respect to m (although the social welfare obtained by h for the modified input may be higher than the social welfare obtained by it for the original input).

Axiom 3:
For every input i and for every $m \in M^n$ with $m_i \neq null$ for which there exists $j \neq i$ with $m_j \neq null$,

$$\sum_{j \neq i, j \in N(m)} w^j(h(null, m_{-i})) \geq \sum_{j \neq i, j \in N(m)} w^j(h(m)).$$

If h satisfies Axiom 3, then the output it obtains for running on an input of l input elements (i.e., assigning the value "null" to all other input elements), can not be improved (as far as this input is concerned) by running it with $l+1$ input elements.

5 The Heuristic Clarke's mechanism

As we mentioned, the Clarke's mechanism makes use of procedures for optimizing social welfare under complete information, which are NP-hard problems. The heuristic Clarke's mechanism is obtained by replacing these procedures by corresponding heuristics or approximations. Formally. we have:

Definition 15 *Let h be a heuristic. The* heuristic Clarke mechanism associated with h *is the direct mechanism (M, f) in which the following three properties hold for every $s \in S$:*

1. $d(m^0, s) = s$ and $d(m, s) = h(m)$ for every $m \neq m^0$.

2. For every agent i and for every $m \in M^n$ with $m_i \neq null$ for which there exists $j \neq i$ with $m_j \neq null$, $t_i(m, s) = \sum_{j \neq i, j \in N(m)} w^j(h(null, m_{-i})) - \sum_{j \neq i, j \in N(m)} w^j(h(m))$.

3. For every agent i and for every $m \in M^n$ for which $m_i \neq null$ and $m_j = null$ for all $j \neq i$, $t_i(m, s) = 0$.

At this point we have modified the Clarke's mechanism in a way where NP-hard procedures it uses for solving OSWP problems are replaced by heuristics. On the other hand, in the previous section we have presented a set of algorithmic axioms one may wish such heuristics to have (regardless of their use in the Clarke's mechanism, or in any other mechanism for group decision making with incomplete information). The following theorem shows that the basic mechanism discussed in the information economics literature, can be extended to the context of resource-bounded agents:

Theorem 3 *A heuristic Clarke's mechanism, which is associated with a heuristic h that satisfies Axioms 1–3 is a strategy-proof, budget balanced mechanism.*

Proof: In order to prove that the mechanism is strategy-proof we need to show that for every $s \in S$, $m_i = (\varphi^i, v^i_{goal}) \in T$, $m'_i \in M$, and $m_{-i} \in M^{n-1}$,

$$(*) \quad v^i(d(m,s)) - t_i(m,s) \geq v^i(d(m',s)) - t_i(m',s),$$

where $m = (m_i, m_{-i})$, $m' = (m'_i, m_{-i})$, and $v^i = v^{m_i}$. Indeed, if $m' = m^0$, then by the definition of mechanism, $t_i(m', s) = 0$, and by (3) in Definition 15, $t_i(m,s) = 0$. Moreover, by (1) in Definition 15, $d(m', s) = s$, and therefore the inequality (*) is equivalent to

$$v^i(h(m)) \geq v^i(s).$$

By Axiom 1, $h(m)$ satisfies φ^i, and therefore $v^i(h(m)) = v^i_{goal} \geq v^i(s)$.

If, on the other hand, $m' \neq m^0$, then (*) follows from Axiom 2.

It remains to show that the mechanism is budget balanced. Let $s \in S$ and let $m \in M^n$. If $m = m^0$, $t_i(m,s) = 0$ for every $i \in N$. Hence

$$\sum_{i \in N} t_i(m,s) \geq 0.$$

If $m \neq m^0$, then for every $i \notin N(m)$, $t_i(m,s) = 0$ by the definition of a mechanism. Hence, it sufficient to show that $t_i(m,s) \geq 0$ for every $i \in N(m)$. Let then $i \in N(m)$. If $m_j = null$ for every $j \neq i$, then $t_i(m,s) = 0$ by (3) in definition 15. If there exists $j \neq i$ with $m_j \neq null$, then the result follows from (2) in Definition 15 and Axiom 3. ▮

6 Discussion: Satisfying Axioms 1-3

The previous sections introduced an extension of the basic mechanism in information economics to the context of resource-bounded agents. We have observed that the major problem in applying the Clarke's mechanism to the context of resource-bounded agents is the use of procedures for optimizing social welfare, under **complete information** about the agents' worth functions. Hence, these procedures should be replaced by some heuristics or approximations. In order to extend the theory of mechanism design to the context of resource-bounded agents, we used an axiomatic approach. In this approach, we specified natural properties that an algorithm which attempts to optimize social welfare (i.e., the sum of the agents' worth functions), under complete information on the agents' worth functions, may have. We then showed that given such an algorithm, we can present a new mechanism (i.e. the heuristic Clarke mechanism) that will have the desired properties in the context of group decision-making with incomplete information. The heuristic Clarke mechanism does not necessarily select an efficient state. However, our axioms guarantee that it selects an *1− efficient* state, in the following sense: If we corrupt one of the inputs (worth functions) of this algorithm and run the algorithm on the corrupted input, then this will produce a state that is less desirable (in the weak sense) than the original one (i.e. the one produced on the right input) as far as the original input is concerned.

One may argue that it may not be easy to construct an algorithm that satisfies our axioms, or to verify that a given algorithm satisfies them. In the full paper we discuss several modifications of our basic model in which we tackle this issue. In particular in these models the following property holds:

Property 1 Given an algorithm h that does not satisfy our axioms, one can construct another algorithm h', which is at least as efficient as h, satisfies the (modified) axioms, and runs in a time which is polynomial in the running time of h.

This shows that in the modified models we can assume without loss of generality that every heuristic satisfies our axioms. As an example of a model that satisfies Property 1, consider the case in which some of the agents in the system are under the designer's control, while the rest of the agents are not under its control. Assume in addition that the number of agents which are not under the center's control, as well as the number of literals and possible valuations in each agent's goal are bounded by a constant k. The center can rely on the agents that are under its control to supply it with correct information about their goals and utilities, but can not rely on the other agents in this regard. We can now look at a restricted form of our axioms, where we refer to the potential corruption of one the uncontrolled agents' inputs.

Further discussion of the modified models are beyond the scope of this short version of our paper.

References

[1] A. H. Bond and L. Gasser. *Readings in Distributed Artificial Intelligence.* Ablex Publishing Corporation, 1988.

[2] C. Boutilier, Y. Shoham, and M.P. Wellman. Special issue on economic principles of multi-agent systems. *Artificial Intelligence*, 94, 1997.

[3] E. Clarke. Multipart pricing of public goods. *Public Choice*, 18:19–33, 1971.

[4] E. Durfee. What your computer really needs to know, you learned in kindergarten. In *10th National Conference on Artificial Intelligence*, pages 858–864, 1992.

[5] E. Ephrati and J. Rosenschein. Deriving consensus in multi-agent systems. *Artificial Intelligence*, 87, 1996.

[6] D. Fudenberg and J. Tirole. *Game Theory.* MIT Press, 1991.

[7] J.R. Green and J. Laffont. *Incentives in Public Decision Making.* North Holland, 1979.

[8] T. Groves. Incentives in teams. *Econometrica*, 41:617–631, 1973.

[9] S. Kraus. Negotiation and cooperation in multi-agent environments. *Artificial Intelligence*, 94, 1997.

[10] D. Kreps. *A Course in Microeconomic Theory.* Princeton University Press, 1990.

[11] D. Monderer and M. Tennenholtz. Optimal Auctions Revisited. In *Proceedings of AAAI-98*, 1998.

[12] J. S. Rosenschein and G. Zlotkin. *Rules of Encounter.* MIT Press, 1994.

[13] T. Sandholm. Limitations of the vickrey auction in computational multi-agent systems. In *2nd International Conference on Multi-Agent Systems*, 1996.

[14] T.W. Sandholm and V.R. Lesser. Equilibrium Analysis of the Possibilities of Unenforced Exchange in Multiagent Syustems. In *Proc. 14th International Joint Conference on Artificial Intelligence*, pages 694–701, 1995.

[15] T.W. Sandholm and V.R. Lesser. Coalitions Among Rationally Bounded Agents. *Artificial Intelligence*, 94(1):99–137, 1997.

[16] H.R. Varian. Economic mechanism design for computerized agents. Working paper, Berkeley University, 1995.

[17] Michael P. Wellman. A market-oriented programming environment and its application to distributed multicommodity flow problems. *Journal of Artificial Intelligence Research*, 1:1–23, 1993.