

Welfare Engineering in Multiagent Systems

Ulle Endriss¹ and Nicolas Maudet²

¹ Department of Computing, Imperial College London
180 Queen's Gate, London SW7 2AZ (UK)

Email: ue@doc.ic.ac.uk

² Department of Computing, School of Informatics, City University
Northampton Square, London EC1V OHB (UK)

Email: maudet@soi.city.ac.uk

Abstract. A multiagent system may be regarded as an artificial society of autonomous software agents. Welfare economics provides formal models of how the distribution of resources amongst the members of a society affects the well-being of that society as a whole. In multiagent systems research, the concept of social welfare is usually given a utilitarian interpretation, i.e. whatever increases the average welfare of the agents inhabiting a society is taken to be beneficial for society as well. While this is indeed appropriate for a wide range of applications, we believe that it is worthwhile to also consider some of the other social welfare orderings that have been studied in the social sciences. In this paper, we put forward an engineering approach to welfare economics in multiagent systems by investigating the following question: Given a particular social welfare ordering appropriate for some application domain, how can we design practical criteria that will allow agents to decide locally whether or not a proposed deal would further social welfare with respect to that ordering? In particular, we review previous results on negotiating Pareto optimal allocations of resources as well as allocations that maximise egalitarian social welfare under this general perspective. We also provide new results on negotiating Lorenz optimal allocations, which may be regarded as a compromise between the utilitarian and the egalitarian approaches. Finally, we briefly discuss elitist agent societies, where social welfare is tied to the welfare of the most successful agent, as well as the notion of envy-freeness.

1 Introduction

Multiagent systems have been successfully applied in a variety of different areas, ranging from electronic commerce [12], over collaborative planning [5], to the fair sharing of resources provided by an earth observation satellite [8]. We may think of a multiagent system as a “society” of autonomous software agents. Given a “solution” to a problem generated by such a society, we may assess the quality of that solution using tools from formal economical sciences. A typical example would be the notion of Pareto optimality: A situation (or state of the system) is called Pareto optimal iff there is no other situation that would make

at least one of the agents in the society happier without making any of the others worse off. Besides Pareto optimality, many other notions of *social welfare* have been put forward in philosophy, sociology, and economics. In the context of multiagent systems, on the other hand, only Pareto optimality and the utilitarian programme (where only increases in average utility are considered to be socially beneficial) have found broad application.

In this paper, we shall argue that also notions such as egalitarian social welfare, Lorenz optimality, or envy-freeness may be usefully exploited when designing multiagent systems. In particular, we are going to discuss scenarios in which autonomous agents negotiate with each other in order to agree on the redistribution of a number of resources. We put forward an engineering approach to welfare economics in multiagent systems by showing how we can design criteria that will allow agents to decide locally whether or not a proposed deal would further social welfare according to the metric chosen by the system designer. Here, we do not use the word engineering in the sense of constructing a software (or even a physical) artefact, but rather to characterise the process of providing practical guidelines for designing artificial societies that exhibit particular properties we are interested in. This involves manipulating the decision making capacity of the agents inhabiting such a society appropriately. While classical welfare economics is concerned with the characterisation of the property of economic well-being with respect to the allocation of resources in a society, in this paper, we promote *welfare engineering* as the process of “engineering” appropriate behaviour profiles for individual agents in such a way that particular desirable properties can be guaranteed to emerge at the level of society.

The remainder of this paper is structured as follows. After motivating the need for different social welfare orderings in Section 2, we are going to define our basic negotiation framework in Section 3. In Section 4, we discuss previous results for two particular instances of this framework, namely for *utilitarian* systems (where a society of agents should at least be able to achieve a Pareto optimal allocation of resources) and for *egalitarian* systems (where society should aim at improving the individual welfare of its weakest member). We then move on to Section 5 and the case of artificial societies where *Lorenz optimal* allocations of resources are desirable (a compromise between the utilitarian and the egalitarian agenda). Before concluding, we briefly discuss *elitist* agent societies, where social welfare is tied to the welfare of the happiest agent, and *envy-free* allocations of resources in Section 6.

2 The *Veil of Ignorance* in Multiagent Systems

In the introduction to this paper we have claimed that multiagent systems research could benefit from considering notions of social welfare that go beyond the utilitarian agenda which aims solely at maximising the sum of the utility levels enjoyed by the individual agents in a system. The question what social welfare ordering is appropriate has been the subject of intense debate in philosophy and the social sciences for a long time. This debate has, in particular, addressed the

respective benefits and drawbacks of utilitarianism on the one hand and egalitarianism on the other [6, 10, 13]. While, under the utilitarian view, social welfare is identified with average utility (or, equivalently, the sum of all individuals' utilities), egalitarian social welfare is measured in terms of the individual welfare of a society's poorest member. Precise definitions of the respective social welfare orderings are given in Section 4.

Different notions of social welfare induce different kinds of social principles. For instance, in an egalitarian system, improving one's personal welfare at the expense of a poorer member of society would be considered inappropriate. A famous argument put forward in defence of egalitarianism is Rawls' *veil of ignorance* [10]. This argument is based on the following thought experiment. To decide what form of society could rightfully be called *just*, a rational person should ask herself the following question:

Without knowing what your position in society (class, race, sex, ...) will be, what kind of society would you choose to live in?

The idea is to decide on a suitable set of social principles that should apply to everyone in society by excluding any kind of bias amongst those who choose the principles. The argument goes that behind this *veil of ignorance* (of not knowing your own future role within the society whose principles you are asked to decide upon), any rational person would choose an egalitarian system, as it insures even the unluckiest members of society a certain minimal level of welfare.

One may or may not agree with this line of reasoning. What we are interested in here is the structure of the thought experiment itself. As far as *human* society is concerned, this is a highly abstract construction (some would argue, *too* abstract to yield any reliable social guidelines). However, for an *artificial* society it can be of very practical concern. Before agreeing to be represented by a software agent in such a society, one would naturally want to know under what principles this society operates. If the agent's objective is to negotiate on behalf of its owner, then the owner has to agree to accept whatever the outcome of a specific negotiation may be. That is, in the context of multiagent systems, we may reformulate the central question of the *veil of ignorance* as follows:

If you were to send a software agent into an artificial society to negotiate on your behalf, what would you consider acceptable principles for that society to operate by?

There is no single answer to this question; it depends on the purpose of the agent society under consideration. For instance, for the application described in [8], where agents need to agree on the access to an earth observation satellite which has been funded jointly by the owners of these agents, it is important that each one of them receives a "fair" share of the common resource. Here, a society governed by egalitarian principles may be the most appropriate. In an electronic commerce application running on the Internet where agents have no commitments to each other, on the other hand, egalitarian principles seem of little relevance. In such a case, it may be in the interest of the system designer to ensure at least Pareto optimal outcomes.

3 Resource Allocation by Negotiation

The general framework within which we are going to investigate welfare engineering is that of *resource allocation by negotiation*, where a number of agents negotiate the redistribution of a number of discrete (i.e. non-divisible) resources in order to benefit either themselves or the artificial society they inhabit. A negotiation scenario consists of a finite set of *agents* \mathcal{A} and a finite set of *resources* \mathcal{R} . Within such a scenario, a resource *allocation* A is a partitioning of \mathcal{R} amongst the agents in \mathcal{A} . For example, for an allocation A with $A(i) = \{r_3, r_7\}$ it would be the case that agent i owns resources r_3 and r_7 . Given a particular allocation of resources, agents may agree on a *deal* to exchange some of the resources they currently hold. In general, a single deal may involve any number of resources and any number of agents. It transforms an allocation of resources A into a new allocation A' , i.e. we can define a deal as a pair $\delta = (A, A')$ of allocations. We also define the set of agents involved in δ as $\mathcal{A}^\delta = \{i \in \mathcal{A} \mid A(i) \neq A'(i)\}$.

Every agent $i \in \mathcal{A}$ is equipped with a *utility function* $u_i : 2^{\mathcal{R}} \rightarrow \mathbb{R}$ to measure its individual welfare with respect to the set of resources it currently holds. We abbreviate $u_i(A) = u_i(A(i))$ for the utility value assigned by agent i to the set of resources it holds for allocation A . An agent may or may not find a particular deal *acceptable*. Here are some examples for possible acceptability criteria:

- A purely *selfish* agent may only accept deals $\delta = (A, A')$ that strictly improve its personal welfare: $u_i(A) < u_i(A')$.
- A *selfish but cooperative* agent may also be content with deals that do leave its own welfare constant: $u_i(A) \leq u_i(A')$.
- A *demanding* agent may require an increase of, say, 10 units for each and every deal it is asked to participate in: $u_i(A) + 10 \leq u_i(A')$.
- A *masochist* agent may insist on losing utility: $u_i(A) > u_i(A')$, etc.

The above are all examples where agents' decisions are based entirely on their own utility functions. This need not be the case:

- A *disciple* of agent *guru* may only accept deals $\delta = (A, A')$ that increase the welfare of the latter: $u_{guru}(A) < u_{guru}(A')$.
- A *team worker* may require the overall utility of a particular group of agents to increase:

$$\sum_{j \in Team} u_j(A) < \sum_{j \in Team} u_j(A')$$

Besides the acceptability criteria adopted by individual agents, the *negotiation protocol* in operation may also restrict the range of possible deals $\delta = (A, A')$:

- For instance, a particular protocol may not allow for more than two agents to be involved in any one deal: $|\mathcal{A}^\delta| \leq 2$.

A social welfare ordering formalises the notion of a society's "preferences" given the preferences of its members (the agents) [1, 9]. We are going to see several examples in the following sections. A particular deal may affect social welfare either positively or negatively. Our objective is to design criteria for the acceptability of deals that will guarantee positive or even optimal outcomes of negotiations.

We should stress here that we have made a number of simplifying assumptions in the definition of our negotiation framework. For instance, we do not take into account the possible costs incurred by trading agents when they redistribute bundles of resources (neither when measuring social welfare nor when modelling the utility functions of individual agents). Furthermore, our framework is static in the sense that agents' utility functions do not change over time. In a system that also allows for the modelling of agents' beliefs and goals in a dynamic fashion, this may not always be appropriate. An agent may, for instance, find out that a particular resource is in fact not required to achieve a particular goal, or it may simply decide to drop that goal for whatever reason. In a dynamic setting, such changes should be reflected by a revision of the agent's utility function. Still, while assuming constant utility functions for the entire life time of an agent may be unrealistic, it does indeed seem reasonable that utility functions do not change for the duration of a particular negotiation process. It is this level of abstraction that our negotiation framework is intended to model.

The most widely studied mechanisms for the reallocation of resources in multiagent systems are *auctions*. We should stress that our scenario of resource allocation by negotiation is *not* an auction. Auctions are mechanisms to help agents agree on a price at which an item (or a set of items) is to be sold [7]. In our work, on the other hand, we are not concerned with this aspect of negotiation, but only with the patterns of resource exchanges that agents actually carry out.

4 Results for Utilitarian and Egalitarian Systems

In this section, we summarise and discuss previous results for the cases of agent societies that are governed by either *utilitarian* or *egalitarian* principles [3, 4].

Definition 1 (Utilitarian social welfare). *The utilitarian social welfare $sw_u(A)$ of an allocation of resources A is defined as follows:*

$$sw_u(A) = \sum_{i \in \mathcal{A}} u_i(A)$$

In systems without explicit utility transfers (i.e. in systems where agents cannot pay each other in order to accept otherwise disadvantageous deals), it is not always possible to negotiate an allocation of resources that maximises utilitarian social welfare without individual agents having to accept a loss in utility. A simple example would be a system with two agents 1 and 2 and a single resource r with $u_1(\{r\}) < u_2(\{r\})$. If agent 1 initially owns the resource, then giving r to agent 2 would increase utilitarian social welfare, but agent 1 may not be prepared to do this. This is why, for utilitarian systems, it is more realistic to aim for allocations that are at least *Pareto optimal*:

Definition 2 (Pareto optimality). *An allocation of resources A is Pareto optimal iff there is no other allocation A' such that $sw_u(A) < sw_u(A')$ and $u_i(A) \leq u_i(A')$ for all $i \in \mathcal{A}$.*

In [3], agents that are selfish but cooperative have been identified as appropriate for utilitarian systems without explicit utility transfers. Such agents will be prepared to accept a deal whenever it is *cooperatively rational*:

Definition 3 (Cooperatively rational deals). *A deal $\delta = (A, A')$ is called cooperatively rational iff $u_i(A) \leq u_i(A')$ for all $i \in \mathcal{A}$ and there exists an agent $j \in \mathcal{A}$ such that $u_j(A) < u_j(A')$.*

The second part of this definition ensures that at least one agent (say, the one proposing the deal) will have a strictly positive payoff for every cooperatively rational deal. This condition is required to ensure the termination of a negotiation process. The following result is proved in [3]:

Theorem 1 (Pareto optimal outcomes). *Any sequence of cooperatively rational deals will eventually result in a Pareto optimal allocation of resources.*

The importance of this result lies in the fact that *any* sequence of deals will lead to a Pareto optimal allocation as long as agents only agree to deals that are cooperatively rational. This means that agents can arrange cooperatively rational deals locally, as they come up; they do not need to plan ahead for society to be able to eventually reach an optimal situation.

On the downside, deals involving any number of resources as well as agents may be *necessary* to reach an optimal allocation provided agents will only agree to deals that are cooperatively rational [3]. Realising such a negotiation protocol seems highly challenging and complex. However, in some cases we can get more favourable results, where a simpler class of deals is sufficient to guarantee an optimal outcome. This is, in particular, the case for so-called *0-1 scenarios* where every agent assigns a utility value of either 1 or 0 to each single resource (thereby specifying whether it does or does not *need* that resource) and where the utility value assigned to a set of resources is simply the sum of the single utilities (i.e. utility functions are additive). In this case, so-called *one-resource-at-a-time* deals (i.e. deals only involving a single resource and two agents) are sufficient to guarantee optimal outcomes in utilitarian systems [3]:

Theorem 2 (Maximising utilitarian welfare in 0-1 scenarios). *In 0-1 scenarios, any sequence of cooperatively rational one-resource-at-a-time deals will eventually result in an allocation of resources with maximal utilitarian welfare.*

As an aside, we remark here that in cases where we are interested in maximising social welfare in a utilitarian agent society with general utility functions, a framework that includes a monetary component that allows (selfish) agents to compensate their trading partners for otherwise disadvantageous deals would be more appropriate. A discussion of such a negotiation framework *with* money (i.e. with explicit utility transfers), as well as proofs for sufficiency and necessity results similar to those reported here, may be found in [3].³

³ See also the work by Sandholm on the closely related subject of sufficient and necessary contract types for optimal allocations of tasks [11].

Table 1. Utility functions for Bob and Mary

$u_{bob}(\{\}) = 0$	$u_{mary}(\{\}) = 0$
$u_{bob}(\{glass\}) = 3$	$u_{mary}(\{glass\}) = 5$
$u_{bob}(\{wine\}) = 12$	$u_{mary}(\{wine\}) = 7$
$u_{bob}(\{glass, wine\}) = 15$	$u_{mary}(\{glass, wine\}) = 17$

We now turn our attention to *egalitarian agent societies* [4]. The first goal of an egalitarian society should be to increase the welfare of its weakest member [9, 10, 13]. In other words, we can measure the social welfare of such a society by measuring the welfare of the agent that is currently worst off:

Definition 4 (Egalitarian social welfare). *The egalitarian social welfare $sw_e(A)$ of an allocation of resources A is defined as follows:*

$$sw_e(A) = \min\{u_i(A) \mid i \in \mathcal{A}\}$$

When searching the economics literature for a class of deals that would benefit society in an egalitarian system we soon encounter *Pigou-Dalton transfers* [9]. In the context of our framework, a Pigou-Dalton transfer (between agents i and j) can be defined as follows:

Definition 5 (Pigou-Dalton transfers). *A deal $\delta = (A, A')$ is called a Pigou-Dalton transfer iff it satisfies the following criteria:*

- Only two agents i and j are involved in the deal: $\mathcal{A}^\delta = \{i, j\}$.
- The deal is mean-preserving: $u_i(A) + u_j(A) = u_i(A') + u_j(A')$.
- The deal reduces inequality: $|u_i(A') - u_j(A')| < |u_i(A) - u_j(A)|$.

The second condition could be relaxed to $u_i(A) + u_j(A) \leq u_i(A') + u_j(A')$, to also allow for inequality-reducing deals that increase overall utility. Pigou-Dalton transfers capture certain egalitarian principles; but are they sufficient as acceptability criteria to guarantee optimal outcomes of negotiations for society?

Consider a scenario with two agents, Bob and Mary, and two resources, a bottle of wine and an empty glass. The utility functions for the two agents are given in Table 1. Bob attributes a high utility value to the wine and a low value to the glass. Furthermore, the value of both resources together is simply the sum of the individual utilities for Bob (no synergy effects). Mary ascribes a medium value to either resource and a very high value to the full set. Now suppose the initial allocation of resources is A with $A(bob) = \{glass\}$ and $A(mary) = \{wine\}$. The “inequality index” for this allocation is $|u_{bob}(A) - u_{mary}(A)| = 4$. We can easily check that inequality is in fact minimal for allocation A . However, allocation A' with $A'(bob) = \{wine\}$ and $A'(mary) = \{glass\}$ would result in higher egalitarian social welfare (namely 5 instead of 3). Hence, Pigou-Dalton transfers alone are not sufficient to guarantee optimal outcomes of negotiations in egalitarian agent societies. We need a more general acceptability criterion. To this end, we have put forward the class of *equitable* deals in [4]:

Definition 6 (Equitable deals). *A deal $\delta = (A, A')$ is called equitable iff we have $\min\{u_i(A) \mid i \in \mathcal{A}^\delta\} < \min\{u_i(A') \mid i \in \mathcal{A}^\delta\}$.*

As shown in [4], this is a sufficient acceptability criterion for deals to guarantee optimal negotiation results in egalitarian systems:

Theorem 3 (Maximising egalitarian welfare). *Any sequence of equitable deals will eventually result in an allocation with maximal egalitarian welfare.*

Again, the connections between the local acceptability criterion and the global welfare ordering are not that surprising. The importance of the theorem lies in the fact that it allows agents to converge towards a global optimum by agreeing on exchanges of resources *locally*, without having to consider the welfare of agents not involved into a particular deal. In the literature on multiagent systems, the *autonomy* of an agent (one of the central features distinguishing multiagent systems from other distributed systems) is sometimes equated with pure selfishness. Under such an interpretation of the agent paradigm, our notion of equitability would, of course, make little sense. We believe, however, that it is useful to distinguish different degrees of autonomy. An agent may well be autonomous in its decision in general, but still be required to follow certain rules imposed by society (and agreed to by the agent on entering that society).

From a purely practical point of view, our results for egalitarian agent societies may be of a lesser interest than those for utilitarian systems, because in the former case it has not been possible to define a deal acceptability criterion that only depends on a *single* agent. Of course, this coincides with our intuitions about egalitarian societies: maximising social welfare is only possible by means of cooperation and the sharing of information on agents' preferences.

5 Negotiating Lorenz Optimal Allocations of Resources

We are now going to introduce a welfare ordering that combines utilitarian and egalitarian notions of social welfare. The basic idea is to endorse deals that result in an improvement with respect to the utilitarian welfare function without causing a loss in egalitarian welfare, and vice versa. An appropriate welfare ordering for this kind of agent society is given by the notion of Lorenz domination [9].

For a society with n agents, let $\{u_1, \dots, u_n\}$ be the set of utility functions for that society. Then every allocation A determines a utility vector $\langle u_1(A), \dots, u_n(A) \rangle$ of length n . If we rearrange the elements of that vector in increasing order we obtain the *ordered utility vector* for allocation A , which we are going to denote by $\vec{u}(A)$. The number $\vec{u}_i(A)$ is the i th element in that ordered utility vector (for $1 \leq i \leq n$). That is, $\vec{u}_1(A)$ for instance, is the utility value assigned to allocation A by the currently weakest agent.

Definition 7 (Lorenz domination). *Let A and A' be allocations of resources for a society with n agents. Then A is Lorenz dominated by A' iff we have*

$$\sum_{i=1}^k \vec{u}_i(A) \leq \sum_{i=1}^k \vec{u}_i(A')$$

for all k with $1 \leq k \leq n$ and that inequality is strict in at least one case.

Table 2. A situation that is not Lorenz optimal

Agent 1	Agent 2	Agent 3
$A(1) = \{\}$	$A(2) = \{\}$	$A(3) = \{r_1, r_2\}$
$u_1(\{\}) = 0$	$u_2(\{\}) = 0$	$u_3(\{\}) = 0$
$u_1(\{r_1\}) = 6$	$u_2(\{r_1\}) = 1$	$u_3(\{r_1\}) = 1$
$u_1(\{r_2\}) = 1$	$u_2(\{r_2\}) = 6$	$u_3(\{r_2\}) = 1$
$u_1(\{r_1, r_2\}) = 7$	$u_2(\{r_1, r_2\}) = 7$	$u_3(\{r_1, r_2\}) = 10$

For any k with $1 \leq k \leq n$, the sum referred to in the above definition is the sum of the utility values assigned to the respective allocation of resources by the k weakest agents. For $k = 1$, this sum is equivalent to the egalitarian social welfare for that allocation. For $k = n$, it is equivalent to the utilitarian social welfare.

An allocation of resources is called *Lorenz optimal* iff it is not Lorenz dominated by any other allocation. When moving from one allocation of resources to another such that the latter Lorenz dominates the former we also speak of a *Lorenz improvement*.

We are now going to try to establish connections between the global welfare measure induced by the notion of Lorenz domination on the one hand, and various local criteria on the acceptability of a proposed deal that individual agents may choose to apply on the other. For instance, it is an immediate consequence of Definitions 3 and 7 that, whenever $\delta = (A, A')$ is a cooperatively rational deal, then A must be Lorenz dominated by A' . As may easily be verified, any deal that amounts to a Pigou-Dalton transfer will also result in a Lorenz improvement. On the other hand, it is not difficult to construct examples that show that this is not the case for the class of equitable deals anymore (that is, while some equitable deals will indeed result in a Lorenz improvement, others will not).

Our next goal is to find a class of deals that captures the notion of Lorenz improvements in as far as, for any two allocations A and A' such that A is Lorenz dominated by A' , there exists a sequence of deals (or possibly even a single deal) belonging to that class leading from A to A' . Given that both cooperatively rational deals and Pigou-Dalton transfers always result in a Lorenz improvement, the union of these two classes of deals may seem like a promising candidate. In fact, according to a result reported by Moulin [9, Lemma 2.3], it is the case that any Lorenz improvement can be implemented by means of a sequence of Pareto improvements and Pigou-Dalton transfers.⁴ It is important to stress that this seemingly general result does *not* apply to our negotiation framework. To see this, consider the example shown in Table 2. The ordered utility vector for allocation A , which assigns both resources to agent 3, is $\vec{u}(A) = \langle 0, 0, 10 \rangle$, i.e. utilitarian social welfare is currently 10. Allocation A is Pareto optimal, because any other allocation would be strictly worse for agent 3. Hence, there can be no cooperatively rational deal that would be applicable in this situation. We also observe that any deal involving only two agents would at best result in a new allocation with a utilitarian social welfare of 7 (this would be a deal consisting

⁴ Note that every Pareto improvement corresponds to a cooperatively rational deal [3].

either of passing both resources on to one of the other agents, or of passing the “preferred” resource to either agent 1 or agent 2, respectively). Hence, no deal involving only two agents (and in particular no Pigou-Dalton transfer) could possibly result in a Lorenz improvement. However, there *is* an allocation that Lorenz dominates A , namely the allocation assigning to each one of the first two agents their respectively preferred resource. This allocation A' with $A'(1) = \{r_1\}$, $A'(2) = \{r_2\}$ and $A'(3) = \{\}$ has got the ordered utility vector $\langle 0, 6, 6 \rangle$. The reason why Moulin’s result is not applicable to our domain is that we cannot use Pigou-Dalton transfers to implement arbitrary utility transfers here. Any such transfer has to correspond to a move in our (discrete) negotiation space.

While this negative result emphasises, again, the high complexity of our negotiation framework, we can get better results for scenarios with restricted utility functions. Recall our definition of 0-1 scenarios where utility functions can only be used to indicate whether an agent does or does not need a particular resource: In such a scenario, $u_i(\{r\})$ is required to be either 0 or 1 for every agent $i \in \mathcal{A}$ and every (single) resource $r \in \mathcal{R}$. Furthermore, utility functions are required to be additive, i.e. we have $u_i(R) = \sum_{r \in R} u_i(\{r\})$ for every set of resources $R \subseteq \mathcal{R}$. As we shall see next, for 0-1 scenarios, the aforementioned result of Moulin *does* apply. In fact, we can even sharpen it a little by showing that only Pigou-Dalton transfers and cooperatively rational deals involving just a single resource and two agents are required to guarantee negotiation outcomes that are Lorenz optimal. We first give a formal definition of this class of deals:

Definition 8 (Simple Pareto-Pigou-Dalton deals). *A deal $\delta = (A, A')$ is called a simple Pareto-Pigou-Dalton deal iff it only involves a single resource and it is either cooperatively rational or a Pigou-Dalton transfer.*

We are now going to show that this class of deals is sufficient to guarantee Lorenz optimal outcomes of negotiations in 0-1 scenarios:

Theorem 4 (Lorenz optimal outcomes in 0-1 scenarios). *In 0-1 scenarios, any sequence of simple Pareto-Pigou-Dalton deals will eventually result in a Lorenz optimal allocation of resources.*

Proof. As pointed out earlier, any deal that is either cooperatively rational or a Pigou-Dalton transfer will result in a Lorenz improvement (not only in the case of 0-1 scenarios). Hence, given that there is only a finite number of different allocations, after a finite number of deals the system will have reached an allocation A where no more simple Pareto-Pigou-Dalton deals are possible (that is, negotiation must terminate).

Now, for the sake of contradiction, let us assume this terminal allocation A is not optimal, i.e. there exists another allocation A' that Lorenz dominates A . Amongst other things, this implies $sw_u(A) \leq sw_u(A')$, i.e. we can distinguish two cases: either (i) there has been a strict increase in utilitarian welfare, or (ii) it has remained constant. In 0-1 scenarios, the former is only possible if there are (at least) one resource $r \in \mathcal{R}$ and two agents $i, j \in \mathcal{A}$ such that $u_i(\{r\}) = 0$ and $u_j(\{r\}) = 1$ as well as $r \in A(i)$ and $r \in A'(j)$, i.e. r has been moved from

agent i (who does not need it) to agent j (who does need it). But then the deal of moving only r from i to j would be cooperatively rational and hence also a simple Pareto-Pigou-Dalton deal. This contradicts our assumption of A being a terminal allocation.

Now let us assume that utilitarian social welfare remained constant, i.e. $sw_u(A) = sw_u(A')$. Let k be the smallest index such that $\vec{u}_k(A) < \vec{u}_k(A')$. (This is the first k for which the inequality in Definition 7 is strict.) Observe that we cannot have $k = |\mathcal{A}|$, as this would contradict $sw_u(A) = sw_u(A')$. We shall call the agents contributing the first k entries in the ordered utility vector $\vec{u}(A)$ the *poor* agents and the remaining ones the *rich* agents. Then, in a 0-1 scenario, there must be a resource $r \in \mathcal{R}$ that is owned by a rich agent i in allocation A and by a poor agent j in allocation A' and that is needed by both these agents, i.e. $u_i(\{r\}) = 1$ and $u_j(\{r\}) = 1$. But then moving this resource from i to j would constitute a Pigou-Dalton transfer (and hence also a simple Pareto-Pigou-Dalton deal) in allocation A , which again contradicts our earlier assumption of A being terminal. \square

In summary, we have shown that (i) any allocation of resources from which no simple Pareto-Pigou-Dalton deals are possible must be a Lorenz optimal allocation and (ii) that such an allocation will always be reached by implementing a finite number of simple Pareto-Pigou-Dalton deals. As with our earlier sufficiency results, agents do not need to worry about which deals to implement, as long as they are simple Pareto-Pigou-Dalton deals. The convergence to a global optimum is guaranteed by the theorem.

6 Further Examples: Elitism and Envy-freeness

In this section, we are going to briefly discuss two further notions of social welfare: *elitism* and *envy-freeness*. The former may be motivated by the fact that, for certain applications, a distributed multiagent system may merely serve as a means for helping a single agent in that system to achieve its goal. However, it may not always be known in advance which agent is most likely to achieve its goal and should therefore be supported by its peers. The welfare of such a society would be evaluated on the basis of the happiest agent (as opposed to the unhappiest agent, as in the case of egalitarian welfare):

Definition 9 (Elitist social welfare). *The elitist social welfare $sw_{el}(A)$ of an allocation of resources A is defined as follows:*

$$sw_{el}(A) = \max\{u_i(A) \mid i \in \mathcal{A}\}$$

In an *elitist agent society*, agents would cooperate in order to support their champion (the currently happiest agent). While such an approach to social welfare may seem somewhat unethical as far as human society is concerned, we believe that it could indeed be very appropriate for certain societies of artificial agents. A typical scenario could be where a system designer launches different agents with the same goal, with the aim that *at least one* agent achieves that goal—no

matter what happens to the others. As with the egalitarian agent societies, this does not contradict the idea of agents being *autonomous* entities. Agents may be physically distributed and make their own autonomous decisions on a variety of issues whilst also adhering to certain social principles, in this case elitist ones.

From a technical point of view, designing a criterion that will allow agents inhabiting an elitist agent society to decide locally whether or not to accept a particular deal is very similar to the egalitarian case [4]. In analogy to the case of equitable deals defined earlier, a suitable deal would have to increase the maximal individual welfare amongst the agents involved in any one deal.

Our final example for an interesting approach to measuring social welfare in an agent society is the issue of envy-freeness [2]. For a particular allocation of resources, an agent may be “envious” of another agent if it would prefer that agent’s set of resources over its own. Ideally, an allocation should be envy-free:

Definition 10 (Envy-freeness). *An allocation of resources A is called envy-free iff we have $u_i(A(i)) \geq u_i(A(j))$ for all agents $i, j \in \mathcal{A}$.*

We should stress that envy-freeness is defined on the sole basis of an agent’s private preferences; that is, there is no need to take other agents’ utility functions into account. On the other hand, whether an agent is envious or not does not only depend on the resources it holds, but also on the resources it *could* hold and whether any of the other agents currently hold a preferred bundle. As we shall see, this somewhat paradoxical situation makes envy-freeness far less amenable to our methodology than any of the other notions of social welfare we have discussed in this paper.

Envy-freeness is desirable (though not always achievable) in societies of self-interested agents in cases where agents have to collaborate with each other over a longer period of time. In such a case, should an agent believe that it has been ripped off, it would have an incentive to leave the coalition which may be disadvantageous for other agents or the society as a whole. In other words, envy-freeness plays an important role with respect to the stability of a group. Unfortunately, envy-free allocations do not always exist. A simple example would be a system with two agents and just a single resource, which is valued by both of them. Then, whichever agent holds that single resource, will be envied by the other agent. To be able to measure different degrees of enviousness, we could, for example, count the number of agents that are envious for a given allocation. However, it is not possible to define a *local* acceptability criterion in terms of the utility functions of the agents involved in a deal (and only those) that indicates whether the deal in question would reduce envy according to such a metric.

7 Conclusion

We have argued that a wide spectrum of social welfare orderings (rather than just those induced by the well known utilitarian social welfare function and the concept of Pareto optimality) can be of interest to agent-based applications.

In an artificial society where agents negotiate over the allocation of resources, different social principles induce different local criteria on the acceptability of a proposed deal. Both in previous work [3, 4] and in the present paper, we have exemplified the idea of *welfare engineering* by designing such local criteria for different social welfare orderings, which in turn are motivated by different types of applications. In particular, we have shown that, for the relatively simple 0-1 scenarios, Lorenz optimal allocations can be achieved using one-to-one negotiation by implementing deals that are either inequality-reducing or that increase the welfare of both agents involved. We have also discussed the case of elitist agent societies and we have pointed out some of the difficulties associated with designing agents that would be able to negotiate allocations of resources where the degree of envy between the agents in a society is minimal.

Acknowledgements. We would like to thank the ESAW referees for their helpful comments. This work has been supported by the European Union as part of the SOCS project (IST-2001-32530).

References

1. K. J. Arrow. *Social Choice and Individual Values*. John Wiley and Sons, 1963.
2. S. J. Brams and A. D. Taylor. *Fair Division: From Cake-cutting to Dispute Resolution*. Cambridge University Press, 1996.
3. U. Endriss, N. Maudet, F. Sadri, and F. Toni. On Optimal Outcomes of Negotiations over Resources. In *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 177–184. ACM Press, 2003.
4. U. Endriss, N. Maudet, F. Sadri, and F. Toni. Resource Allocation in Egalitarian Agent Societies. In *Secondes Journées Francophones sur les Modèles Formels d'Interaction*, pages 101–110. Cépaduès-Éditions, 2003.
5. B. J. Grosz and S. Kraus. Collaborative Plans for Complex Group Action. *Artificial Intelligence*, 86(2):269–357, 1996.
6. J. C. Harsanyi. Can the Maximin Principle Serve as a Basis for Morality? *American Political Science Review*, 69:594–609, 1975.
7. G. E. Kersten, S. J. Noronha, and J. Teich. Are All E-Commerce Negotiations Auctions? In *Proceedings of the 4th International Conference on the Design of Cooperative Systems*, 2000.
8. M. Lemaître, G. Verfaillie, and N. Bataille. Exploiting a Common Property Resource under a Fairness Constraint: A Case Study. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, pages 206–211. Morgan Kaufmann Publishers, 1999.
9. H. Moulin. *Axioms of Cooperative Decision Making*. Cambridge University Press, 1988.
10. J. Rawls. *A Theory of Justice*. Oxford University Press, 1971.
11. T. W. Sandholm. Contract Types for Satisficing Task Allocation: I Theoretical Results. In *AAAI Spring Symposium: Satisficing Models*, 1998.
12. T. W. Sandholm. Distributed Rational Decision Making. In G. Weiss, editor, *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, pages 201–258. MIT Press, 1999.
13. A. K. Sen. *Collective Choice and Social Welfare*. Holden Day, 1970.