Negotiating Socially Optimal Allocations of Resources: An Overview

Ulle Endriss,[1] Nicolas Maudet,[2] Fariba Sadri[1] and Francesca Toni[1]

[1]Department of Computing, Imperial College London (UK)
Email: {ue,fs,ft}@doc.ic.ac.uk

[2]LAMSADE, Université Paris-Dauphine (France)
Email: maudet@lamsade.dauphine.fr

22 March 2005 (Revised: 15 July 2005)

Abstract

A multiagent system may be thought of as an artificial society of autonomous software agents and we can apply concepts borrowed from welfare economics and social choice theory to assess the social welfare of such an agent society. In this paper, we study an abstract negotiation framework where agents can agree on multilateral deals to exchange bundles of discrete resources. We then analyse how these deals affect social welfare for different instances of the basic framework and different interpretations of the concept of social welfare itself. In particular, we show how certain classes of deals are both sufficient and necessary to guarantee that a socially optimal allocation of resources will be reached eventually.

1 Introduction

A multiagent system may be thought of as an artificial society of autonomous software agents. Negotiation over the distribution of resources (or tasks) amongst the agents inhabiting such a society is an important area of research in artificial intelligence and computer science [8, 25, 31, 36]. A number of variants of this problem have been studied in the literature. Here we consider the case of an artificial society of agents where, to begin with, each agent holds a bundle of discrete (*i.e.* non-divisible) resources to which it assigns a certain utility. Agents may then negotiate with each other in order to agree on the redistribution of some of these resources to benefit either themselves or the agent society they inhabit.

Rather than being concerned with specific strategies for negotiation [25, 31] or concrete communication protocols to enable negotiation in a multiagent system [1, 39], we analyse how the redistribution of resources by means of negotiation affects the well-being of the agent society as a whole. To this end, we make use of formal tools for measuring social welfare developed in welfare economics and social choice theory [4, 28]. 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 (and are going to argue in this paper) that it is worthwhile to also consider some of the other social welfare orderings that have been studied in the social sciences. As we shall argue, also notions such as egalitarian social welfare [38], Lorenz optimality [28], or envy-freeness [7] may be usefully exploited when designing multiagent systems.

In this paper, we study the effect that negotiation over resources has on society for a number of different interpretations of the concept of social welfare. In particular, we show how certain classes of deals regarding the exchange of resources allow us to guarantee that a socially optimal allocation of resources will be reached eventually. These results may be interpreted as the emergence of a particular global behaviour (at the level of society) in reaction to local behaviour governed by the negotiation strategies of individual agents (which determine the kinds of deals agents are prepared to accept). The work described here is complementary to the large body of literature on mechanism design and game-theoretical models of negotiation in multiagent systems (see e.g. [20, 25, 31]). While such work is typically concerned with negotiation at the local level (how can we design mechanisms that provide an incentive to individual agents to adopt a certain negotiation strategy?), we address negotiation at a global level by analysing how the actions taken by agents locally affect the overall system from a social point of view.

We also show that truly multilateral deals involving any number of agents as well as any number of resources may be necessary to be able to negotiate socially optimal allocations of resources. This is true as long as we use arbitrary utility functions to model the preferences of individual agents. In some application domains, however, where utility functions may be assumed to be subject to certain restrictions (such as being additive), we are able to obtain stronger results and show that also structurally simpler classes of deals (in particular, deals involving only a single resource at a time) can be sufficient to negotiate socially optimal allocations.

Our approach to multiagent resource allocation is of a *distributed* nature. In general, the allocation procedure used to find a suitable allocation of resources could be either centralised or distributed. In the centralised case, a single entity decides on the final allocation of resources amongst agents, possibly after having elicited the agents' preferences over alternative allocations. Typical examples are combinatorial auctions [12]. Here the central entity is the auctioneer and the reporting of preferences takes the form of bidding. In truly distributed approaches, on the other hand, allocations emerge as the result of a sequence of local negotiation steps. Both approaches have their advantages and disadvantages. Possibly the most important argument in favour of auction-based mechanisms concerns the simplicity of the communication protocols required to implement such mechanisms. Another reason for the popularity of centralised mechanisms is the recent push in the design of powerful algorithms for combinatorial auctions that, for the first time, perform reasonably well in practice [22, 37]. Of course, such techniques

are, in principle, also applicable in the distributed case, but research in this area has not yet reached the same level of maturity as for combinatorial auctions. An important argument *against* centralised approaches is that it may be difficult to find an agent that could assume the role of an "auctioneer" (for instance, in view of its computational capabilities or in view of its trustworthiness). The distributed model seems also more natural in cases where finding optimal allocations may be (computationally) infeasible, but even small improvements over the initial allocation of resources would be considered a success. While centralised approaches tend to operate to an "all or nothing" policy (either an optimal allocation can be found in the time available or no solution at all will be computed), step-wise improvements over the *status quo* are naturally modelled in a distributed negotiation framework.

The line of research pursued in this paper has been inspired by Sandholm's work on sufficient and necessary contract (*i.e.* deal) types for distributed task allocation [35]. Since then, it has been further developed by the present authors, their colleagues, and others in the context of resource allocation problems [6, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19]. In particular, we have extended Sandholm's framework by also addressing negotiation systems without compensatory side payments [18], as well as agent societies where the concept of social welfare is given a different interpretation to that in the utilitarian programme [16, 19]. The present paper provides a comprehensive overview of the most fundamental results, mostly on the convergence to an optimal allocation with respect to different notions of social welfare, in a very active and timely area of ongoing research.

The remainder of this paper is organised as follows. Section 2 introduces the basic negotiation framework for resource reallocation we are going to consider. It gives definitions for the central notions of *allocation*, *deal*, and *utility*, and it discusses possible restrictions to the class of admissible deals (both structural and in terms of acceptability to individual agents). Section 2 also includes a short introduction to the concept of a socially optimal allocation of resources. Subsequent sections analyse specific instances of the basic negotiation framework (characterised, in particular, by different criteria for the acceptability of a proposed deal) with respect to specific notions of social welfare. In the first instance, agents are assumed to be *rational* (and "myopic") in the sense of never accepting a deal that would result in a negative payoff. Section 3 analyses the first variant of this model of rational negotiation, which allows for monetary side payments to increase the range of acceptable deals. As we shall see, this model facilitates negotiation processes that maximise *utilitarian social welfare*. If side payments are not possible, we cannot guarantee outcomes with maximal social welfare, but it is still possible to negotiate *Pareto optimal* allocations. This variant of the rational model is studied in Section 4. Section 5 investigates how restrictions to the range of admissible utility functions may affect the results obtained earlier.

At the beginning of Section 6, drawing on Rawls' famous *veil of ignorance* [30], we argue why we believe that multiagent systems research could benefit from considering notions of social welfare that go beyond the utilitarian agenda. The remainder of that section analyses our framework of resource allocation by negotiation in the context of

egalitarian agent societies. Section 7 discusses a variant of the framework that combines ideas from both the utilitarian and the egalitarian programme and enables agents to negotiate *Lorenz optimal* allocations of resources. This section also includes a discussion of the idea to use an *elitist* model of social welfare for applications where societies of agents are merely a means to enable at least one agent to achieve their goal. Finally, Section 7 also reviews the concept of *envy-freeness* and discusses ways of measuring different degrees of envy. Section 8 summarises our results and concludes with a brief discussion of the concept of *welfare engineering*, *i.e.* with the idea of choosing tailor-made definitions of social welfare for different applications and designing agents' behaviour profiles accordingly.

2 Resource Allocation by Negotiation

The basic scenario of *resource allocation by negotiation* studied in this paper is that of an artificial society inhabited by a number of agents, each of which initially holds a certain number of resources. These agents will typically ascribe different values (utilities) to different bundles of resources. They may then engage in negotiation and agree on the reallocation of some of the resources, for example, in order to improve their respective individual welfare (*i.e.* to increase their utility). Furthermore, we assume that it is in the interest of the system designer that these distributed negotiation processes —somehow— also result in a positive payoff for society as a whole.

2.1 Basic Definitions

An instance of our abstract negotiation framework consists of a finite set of (at least two) *agents* $\mathcal{A}$ and a finite set of discrete (*i.e.* non-divisible)[1] *resources* $\mathcal{R}$. An *allocation of resources* is a partitioning of $\mathcal{R}$ amongst the agents in $\mathcal{A}$.

Definition 1 (Allocations) *An allocation of resources is a function A from $\mathcal{A}$ to subsets of $\mathcal{R}$ such that $A(i) \cap A(j) = \{\}$ for $i \neq j$ and $\bigcup_{i \in \mathcal{A}} A(i) = \mathcal{R}$.*

For example, given an allocation A with $A(i) = \{r_3, r_7\}$, agent i would own resources r_3 and r_7. Given a particular allocation of resources, agents may agree on a (multilateral) *deal* to exchange some of the resources they currently hold. An example would be: "I give you r_1 if you give r_2 to me and r_3 to John". Note that, in the most general case, any numbers of agents and resources could be involved in a single deal. From an abstract point of view, a deal takes us from one allocation of resources to the next. That is, we may characterise a deal as a pair of allocations.

Definition 2 (Deals) *A deal is a pair $\delta = (A, A')$ where A and A' are allocations of resources with $A \neq A'$.*

[1] Other authors have also considered *continuous* resources (e.g. [25]). Our results, on the other hand, specifically apply to scenarios with discrete resources, where the number of distinct bundles an agent could hold is finite.

The intended interpretation of the above definition is that the deal $\delta = (A, A')$ is only applicable in situation A and will result in situation A'. It specifies for each resource in the system whether it is to remain where it is or where it is to be moved to. When referring to a *sequence* of deals, we implicitly mean that any deal in such a sequence applies to the allocation that has been reached by implementing its predecessor in the sequence. That is, for a sequence of deals $\delta_1, \delta_2, \ldots, \delta_n$, there are allocations $A_0, A_1, A_2, \ldots, A_n$ such that $\delta_1 = (A_0, A_1)$, $\delta_2 = (A_1, A_2)$, $\ldots$, $\delta_n = (A_{n-1}, A_n)$. We write $\mathcal{A}^\delta$ for the set of agents involved in δ. Observe that the agents involved in a deal are exactly those that hold a different bundle of resources after the deal has been implemented.

Definition 3 (Involved agents) *The set of agents involved in a deal $\delta = (A, A')$ is given by $\mathcal{A}^\delta = \{i \in \mathcal{A} \mid A(i) \neq A'(i)\}$.*

The *composition* of two deals is defined as follows: If $\delta_1 = (A, A')$ and $\delta_2 = (A', A'')$, then $\delta_1 \circ \delta_2 = (A, A'')$. If a given deal is the composition of two deals that concern disjoint sets of agents, then that deal is said to be *independently decomposable*.

Definition 4 (Independently decomposable deals) *A deal δ is called independently decomposable iff there exist deals δ_1 and δ_2 such that $\delta = \delta_1 \circ \delta_2$ and $\mathcal{A}^{\delta_1} \cap \mathcal{A}^{\delta_2} = \{\}$.*

By Definitions 3 and 4, if $\delta = (A, A')$ is independently decomposable then there exists an intermediate allocation B different from both A and A' such that the intersection of $\{i \in \mathcal{A} \mid A(i) \neq B(i)\}$ and $\{i \in \mathcal{A} \mid B(i) \neq A'(i)\}$ is empty, *i.e.* such that the union of $\{i \in \mathcal{A} \mid A(i) = B(i)\}$ and $\{i \in \mathcal{A} \mid B(i) = A'(i)\}$ is the full set of agents $\mathcal{A}$. Hence, $\delta = (A, A')$ *not* being independently decomposable implies that there exists no allocation B different from both A and A' such that either $B(i) = A(i)$ or $B(i) = A'(i)$ for all agents $i \in \mathcal{A}$ (we are going to use this fact in the proofs of our "necessity theorems" later on).

The value an agent $i \in \mathcal{A}$ ascribes to a particular set of resources R will be modelled by means of a *utility function*, that is, a function from sets of resources to real numbers.[2]

Definition 5 (Utility functions) *Every agent $i \in \mathcal{A}$ is equipped with a utility function u_i mapping subsets of $\mathcal{R}$ to real numbers.*

That is, we do not model any *externalities*; the utility an agent derives from an allocation A only depends on the bundle of resources it receives in that allocation, but not on any other aspects of the allocation (or indeed factors that lie outside the allocation itself). Recall that, given an allocation A, the set $A(i)$ is the bundle of resources held by agent i in that situation. We are usually going to abbreviate $u_i(A) = u_i(A(i))$ for the utility value assigned by agent i to that bundle.

[2]This could really be *any* such function; in particular, the utility ascribed to a set of resources is not just the sum of the values ascribed to its elements. The interesting aspect of this is that we can model the fact that the utility assigned to a single resource may strongly depend on context, *i.e.* on what other resources the agent in question holds at the same time. Still, for certain applications, restricted classes of utility functions may be more appropriate choices. Some of these will be discussed in Section 5.

An agent's utility function induces a *preference ordering* over the set of alternative allocations of resources for that agent. For instance, if $u_i(A_1) > u_i(A_2)$ then a purely rational agent i would prefer allocation A_1 over allocation A_2. In some cases, the particular values of the utility function are not relevant and we are only interested in the agent's preference profile. In fact, from a cognitive point of view, one may argue that qualitative (non-numerical) preference orderings are more appropriate than the quantitative approach where we associate specific numbers with alternative situations. Still, technically it will often simply be more convenient to describe an agent's preferences in terms of a (utility) function from the resources it holds in a given situation to numerical values. But only when we are working in a framework that requires utilities to be intercomparable (as in systems where agents have to agree on prices), we actually *need* access to specific numerical utility values.

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. 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. Another assumption that we make is that the composition of an agent society does not change while a negotiation process is going on, *i.e.* it is not possible that agents leave or join the system during negotiation. The reason for this assumption is that the comparison of the social welfare associated with different states is (typically) only meaningful if it is based on the same set of agents.

2.2 Types of Deals

Following Sandholm [35], we can distinguish a number of structurally different *deal types*. The simplest deals are *one-resource-at-a-time deals* where a single resource is passed on from one agent to another. This corresponds to the "classical" form of a contract typically found in the *contract net protocol* [39].

Deals where one agent passes a *set* of resources on to another agent are called *cluster deals*. Deals where one agent gives a single item to another agent who returns another single item are called *swap deals*. Sometimes it can also be necessary to exchange resources between more than just two agents. A *multiagent deal* is a deal that could involve any number of agents, where each agent passes at most one resource to each of the other agents taking part. Finally, deals that combine the features of the cluster and the multiagent deal type are called *combined deals*. These could involve any number of agents and any number

of resources. Therefore, *every* deal δ, in the sense of Definition 2, is a combined deal. In the remainder of this paper, when speaking about deals without further specifying their type, we are always going to refer to combined deals (without any structural restrictions).

The ontology of deal types discussed here is, of course, not exhaustive. It would, for instance, also be of interest to consider the class of bilateral deals, *i.e.* the class of deals involving exactly two agents but any number of resources.

2.3 Acceptable Deals

An agent may or may not find a particular deal $\delta = (A, A')$ *acceptable*. Here are some examples for possible acceptability criteria that agents may choose to use during negotiation:

- 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 every deal it is asked to participate in: $u_i(A) + 10 \leq u_i(A')$.
- A *masochist* agent may insist on losing utility: $u_i(A) > u_i(A')$.

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 *competitive* agent i may only accept deals $\delta = (A, A')$ that improve its own welfare with respect to the welfare enjoyed by its arch-rival (say, agent j): $u_i(A) - u_j(A) < u_i(A') - u_j(A')$.
- A *team worker* may require the collective utility of a particular group of agents to increase:

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

Note that the above examples are for illustration purposes only. For each of the instances of our negotiation framework discussed in later sections of this paper, we are going to give formal definitions of suitable acceptability criteria.

We should stress that, in this paper, we are not concerned with concrete negotiation mechanisms that allow agents to agree on particular deals. We only assume that, whatever concrete strategies agents may use, their behaviour is constrained by an acceptability criterion such as any of those discussed above.

2.4 Socially Optimal Allocations of Resources

As already mentioned in the introduction, we may think of a multiagent system as a *society* of autonomous software agents. While agents make their *local* decisions on what deals to propose and to accept, we can also analyse the system from a *global* or *societal* point of view and may thus prefer certain allocations of resources over others. To this end, *welfare economics* provides formal tools to assess how the distribution of resources amongst the members of a society affects the well-being of society as a whole (see, for instance, [28, 38]). A typical example would be the notion of *Pareto optimality:* A state of affairs (such as a particular allocation of resources) is called Pareto optimal iff there is no other state that would make at least one of the agents in the society better off 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. As we shall argue in Section 6.1, we believe that some of the other approaches to social welfare studied in the literature on welfare economics (or even entirely new definition of social welfare) are also relevant to multiagent systems.

In some cases, social welfare can be defined in terms of a collective utility function (we are going to see several examples in this paper) [28]. An allocation of resources A is then considered to be socially preferred over another allocation A' iff the value of the collective utility function is higher for A than it is for A'. In other cases it is not possible to express social preferences in terms of a function, but instead an ordering over alternative allocations representing the intended notion of social preference is given directly (again, we are going to see several examples). For a given notion of social preference, an allocation of resources is then called *socially optimal* iff there is no other allocation that is socially preferred to the former. For a number of different variants of the basic negotiation framework defined earlier in this section, we are going to investigate under what circumstances trading resources will lead to an allocation that is socially optimal with respect to a measure of social welfare suitable for that framework. In particular, we are going to be concerned with *utilitarian social welfare* (Sections 3 and 5), *Pareto optimality* (Section 4), *egalitarian social welfare* (Section 6), and *Lorenz optimality* (Section 7.1). We are also briefly going to discuss the idea of *elitist agent societies* (Section 7.2) and the concept of *envy-freeness* (Section 7.3).

2.5 Relation to Auctions

The most widely studied mechanisms for the reallocation of resources in multiagent systems are probably *auctions*. We should point out 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 [24]. 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. On top of that, with one exception, the concrete instances of our negotiation framework, do in fact not involve a monetary component at all, *i.e.* there is no notion of a price as such either.

Typically, an auction involves a single auctioneer selling goods to a number of potential buyers. In contrast to this, our negotiation scenario is *symmetric* (there is no distinction between sellers and buyers) and we specifically address the issue of multiple agents negotiating over multiple goods at the same time. The latter aspect has, to a certain degree, also been addressed in work on more complex auction mechanisms, in particular simultaneous auctions [32], combinatorial auctions [12], and sequential auctions [5]. While it may be possible to use a combination of such auction mechanisms to negotiate the precise conditions accompanying a deal in our scenario (at least if we include a monetary component), in the present paper we are only concerned with the structure of these deals themselves.

Nevertheless, at a more abstract level there is an important connection between our distributed negotiation framework and *combinatorial auctions*. If we view the problem of finding a socially optimal allocation as an algorithmic problem faced by a central authority (rather than as a problem of designing suitable negotiation mechanisms), then we can observe an immediate relation to the so-called *winner determination problem* in combinatorial auctions [12]. In a combinatorial auction, bidders can put in bids for different bundles of items (rather than just single items). After all bids have been received, the auctioneer has to find an allocation for the items on auction amongst the bidders in a way that maximises his revenue. If we interpret the price offered for a particular bundle of items as the utility the agent in question assigns to that set, then maximising revenue (*i.e.* the sum of prices associated with winning bids) is equivalent to finding an allocation with a maximal sum of individual utilities (which is one of the notions of social optimality considered in this paper). This equivalence holds, at least, in cases where the optimal allocation of items in an auction is such that *all* of the items on auction are in fact being sold (so-called *free disposal*).

Winner determination in combinatorial auctions is known to be an NP-hard optimisation problem, even for bidding languages that are less expressive (in terms of describing agent preferences) than the general utility functions used here [33]. These complexity results easily transfer also to our framework [9, 15]. Despite such negative theoretical results, we should stress that, in recent years, several algorithms for combinatorial auction winner determination have been proposed and applied successfully [22, 33, 37]. Given the close relationship between the two models, we would expect similarly positive results to be attainable also for distributed negotiation mechanisms. However, this is an issue for future research. In the present paper, we are only concerned with the theoretical feasibility of negotiating socially optimal allocations of resources.

3 Rational Negotiation with Side Payments

In this section, we are going to discuss a first instance of the general framework of *resource allocation by negotiation* set out earlier. This particular variant, which we shall refer to as the model of *rational negotiation with side payments* (or simply *with money*), is equivalent to a framework put forward by Sandholm where agents negotiate in order to reallocate *tasks* [35].

3.1 Additional Definitions

We first have to fix a notion of optimality for society as a whole. Adopting a utilitarian view, we define the *utilitarian social welfare* of an agent society for a given allocation A as the sum of the values the agents in that society ascribe to the sets of resources they hold in situation A [28].

Definition 6 (Utilitarian social welfare) *The utilitarian social welfare $sw_u(A)$ of an allocation of resources A is defined as follows:*

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

We say that an allocation A has *maximal utilitarian social welfare* (for a given system defined by a set of agents $\mathcal{A}$ with their utility functions and a set of resources $\mathcal{R}$) iff there is no other allocation A' for that system with $sw_u(A) < sw_u(A')$. Observe that maximising the collective utility function sw_u amounts to maximising the *average utility* enjoyed by the agents in the system.[3]

In this instance of our negotiation framework, a deal may be accompanied by a number of monetary *side payments* to compensate some of the agents involved for accepting a loss in utility. Rather than specifying for each pair of agents how much money the former is supposed to pay to the latter, we simply say how much money each agent either pays out or receives. This can be modelled by using what we call a *payment function*.

Definition 7 (Payment functions) *A payment function is a function p from $\mathcal{A}$ to real numbers satisfying the following condition:*

$$\sum_{i \in \mathcal{A}} p(i) \quad = \quad 0$$

Here, $p(i) > 0$ means that agent i *pays* the amount of $p(i)$, while $p(i) < 0$ means that it *receives* the amount of $-p(i)$. By definition of a payment function, the sum of all payments is 0, *i.e.* the overall amount of money present in the system does not change.[4]

[3] This is true as long as the size of society does not change. Throughout this paper, we assume that the set of agents making up a society remains constant. If this is not the case, then average utility would indeed provide a better measure of utilitarian social welfare than the sum of individual utilities.

[4] As the overall amount of money present in the system stays constant throughout the negotiation process, it makes sense not to take it into account for the evaluation of social welfare.

In the rational negotiation model, agents are self-interested in the sense that they will only propose or accept deals that strictly increase their own welfare. This "myopic" notion of *individual rationality* may be formalised as follows (see [35] for a justification of this approach).

Definition 8 (Individual rationality) *A deal $\delta = (A, A')$ is called individually rational iff there exists a payment function p such that $u_i(A') - u_i(A) > p(i)$ for all $i \in \mathcal{A}$, except possibly $p(i) = 0$ for agents i with $A(i) = A'(i)$.*

That is, agent i will be prepared to accept the deal δ iff it has to pay less than its gain in utility or it will get paid more than its loss in utility, respectively. Only for agents i not affected by the deal, *i.e.* in case $A(i) = A'(i)$, there may be no payment at all. For example, if $u_i(A) = 8$ and $u_i(A') = 5$, then the utility of agent i would be reduced by 3 units if it were to accept the deal $\delta = (A, A')$. Agent i will only agree to this deal if it is accompanied by a side payment of more than 3 units; that is, if the payment function p satisfies $-3 > p(i)$.

For any given deal, there will usually be a range of possible side payments. How agents manage to agree on a particular one is not a matter of consideration at the abstract level at which we are discussing this framework here. We assume that a deal will go ahead as long as there exists *some* suitable payment function p. We should point out that this assumption may not be justified under all circumstances. For instance, if utility functions are not publicly known and agents are risk-takers, then a potential deal may not be identified as such, because some of the agents may understate their interest in that deal in order to maximise their expected payoff [29]. Therefore, the theoretical results on the reachability of socially optimal allocations of resources reported below will only apply under the assumption that such strategic considerations will not prevent agents from making mutually beneficial deals.

3.2 An Example

As an example, consider a system with two agents, agent 1 and agent 2, and a set of two resources $\mathcal{R} = \{r_1, r_2\}$. The following table specifies the values of the utility functions u_1 and u_2 for every subset of $\{r_1, r_2\}$:

$u_1(\{\,\}) = 0$	$u_2(\{\,\}) = 0$
$u_1(\{r_1\}) = 2$	$u_2(\{r_1\}) = 3$
$u_1(\{r_2\}) = 3$	$u_2(\{r_2\}) = 3$
$u_1(\{r_1, r_2\}) = 7$	$u_2(\{r_1, r_2\}) = 8$

Also suppose agent 1 initially holds the full set of resources $\{r_1, r_2\}$ and agent 2 does not own any resources to begin with.

The utilitarian social welfare for this initial allocation is 7, but it could be 8, namely if agent 2 had both resources. As we are going to see next, the simple class of *one-resource-at-a-time deals* alone are not always sufficient to guarantee the optimal outcome

of a negotiation process (if agents abide to the individual rationality criterion for the acceptability of a deal). In our example, the only possible one-resource-at-a-time deals would be to pass either r_1 or r_2 from agent 1 to agent 2. In either case, the loss in utility incurred by agent 1 (5 or 4, respectively) would outweigh the gain of agent 2 (3 for either deal), so there is no payment function that would make these deals individually rational.

The *cluster deal* of passing $\{r_1, r_2\}$ from agent 1 to 2, on the other hand, *would* be individually rational if agent 2 paid agent 1 an amount of, say, 7.5 units.

Similarly to the example above, we can also construct scenarios where swap deals or multiagent deals are necessary (*i.e.* where cluster deals alone would not be sufficient to guarantee maximal social welfare). This also follows from Theorem 2, which we are going to present later on in this section. Several concrete examples are given in [35].

3.3 Linking Individual Rationality and Social Welfare

The following result, first stated in this form in [18], says that a deal (with money) is individually rational iff it increases utilitarian social welfare. We are mainly going to use this lemma to give a simple proof of Sandholm's main result on sufficient contract types [35], but it has also found useful applications in its own right [13, 15, 17].

Lemma 1 (Individually rational deals and utilitarian social welfare) *A deal $\delta = (A, A')$ is individually rational iff $sw_u(A) < sw_u(A')$.*

Proof. '$\Rightarrow$': By definition, $\delta = (A, A')$ is individually rational iff there exists a payment function p such that $u_i(A') - u_i(A) > p(i)$ holds for all $i \in \mathcal{A}$, except possibly $p(i) = 0$ in case $A(i) = A'(i)$. If we add up the inequations for all agents $i \in \mathcal{A}$ we get:

$$\sum_{i \in \mathcal{A}} (u_i(A') - u_i(A)) \quad > \quad \sum_{i \in \mathcal{A}} p(i)$$

By definition of a payment function, the righthand side equates to 0 while, by definition of utilitarian social welfare, the lefthand side equals $sw_u(A') - sw_u(A)$. Hence, we really get $sw_u(A) < sw_u(A')$ as claimed.

'$\Leftarrow$': Now let $sw_u(A) < sw_u(A')$. We have to show that $\delta = (A, A')$ is an individually rational deal. We are done if we can prove that there exists a payment function p such that $u_i(A') - u_i(A) > p(i)$ for all $i \in \mathcal{A}$. We define p to be a function from $\mathcal{A}$ to the reals as follows:

$$p(i) \quad = \quad u_i(A') - u_i(A) - \frac{sw_u(A') - sw_u(A)}{|\mathcal{A}|} \quad \text{(for } i \in \mathcal{A}\text{)}$$

First, observe that p really *is* a payment function, because we get $\sum_{i \in \mathcal{A}} p(i) = 0$. We also get $u_i(A') - u_i(A) > p(i)$ for all $i \in \mathcal{A}$, because we have $sw_u(A') - sw_u(A') > 0$. Hence, δ must indeed be an individually rational deal. □

Lemma 1 suggests that the utilitarian collective utility function sw_u does indeed provide an appropriate measure of social well-being in societies of autonomous agents that use the notion of individual rationality (as given by Definition 8) to guide their behaviour during negotiation.

3.4 Maximising Utilitarian Social Welfare

Our next aim is to show that any sequence of deals in the rational negotiation model with side payments will converge to an allocation of resources with maximal utilitarian social welfare. We first prove that any such negotiation process is bound to terminate after a finite number of steps.

Lemma 2 (Termination) *There can be no infinite sequence of individually rational deals.*

Proof. Given that both the set of agents $\mathcal{A}$ as well as the set of resources $\mathcal{R}$ are required to be finite, there can be only a finite number of distinct allocations of resources. Furthermore, by Lemma 1, any individually rational deal will strictly increase utilitarian social welfare. Hence, negotiation must terminate after a finite number of deals. □

We are now ready to state the main result for the rational negotiation model with side payments, namely that the class of individually rational deals (as given by Definition 8) is *sufficient* to guarantee optimal outcomes for agent societies measuring welfare according to the utilitarian programme (Definition 6). This has originally been shown by Sandholm in the context of a framework where rational agents negotiate with each other in order to reallocate tasks and where the global aim is to minimise the overall costs of carrying out these tasks [35].

Theorem 1 (Maximal utilitarian social welfare) *Any sequence of individually rational deals will eventually result in a resource allocation with maximal utilitarian social welfare.*

Proof. By Lemma 2, any sequence of individually rational deals must terminate. For the sake of contradiction, assume that the terminal allocation A does *not* have maximal utilitarian social welfare, *i.e.* there exists another allocation A' with $sw_u(A) < sw_u(A')$. But then, by Lemma 1, the deal $\delta = (A, A')$ would be individually rational and thereby possible, which contradicts our earlier assumption of A being a terminal allocation. □

At first sight, this result may seem almost trivial. The notion of a multilateral deal without any structural restrictions is a *very* powerful one. A single such deal allows for any number of resources to be moved between any number of agents. From this point of view, it is not particularly surprising that we can always reach an optimal allocation (even in just a single step!). Furthermore, *finding* a suitable deal is a very complex task, which may not always be viable in practice. Therefore, one may question the relevance of this theorem. It *is* relevant. The true power of Theorem 1 is in the fine print: *any* sequence of deals will result in an optimal allocation. That is, whatever deals are agreed on in the early stages of the negotiation, the system will never get stuck in a local optimum and finding an allocation with maximal social welfare remains an option throughout (provided, of course, that agents are actually able to identify any deal that

is theoretically possible). Given the restriction to deals that are individually rational for all the agents involved, social welfare must increase with every single deal. Therefore, negotiation always pays off, even if it has to stop early due to computational limitations.

The issue of complexity is still an important one. If the full range of deals is too large to be managed in practice, it is important to investigate how close we can get to finding an optimal allocation if we restrict the set of allowed deals to certain simple patterns. Andersson and Sandholm [2], for instance, have conducted a number of experiments on the sequencing of certain contract/deal types to reach the best possible allocations within a limited amount of time. For a complexity-theoretic analysis of the problem of deciding whether it is possible to reach an optimal allocation by means of structurally simple types of deals (in particular one-resource-at-a-time deals), we refer to recent work by Dunne et al. [15].

It should be noted that Theorem 1 (as well as the other sufficiency results presented in later sections) is a theoretical result pertaining to an abstract negotiation framework. The theorem only claims that agents are guaranteed to eventually reach an optimal allocation of resources *provided* they are capable of identifying an individually rational deal whenever such a deal exists. This paper does, however, not address the (very important and complex) problem of actually finding these deals. In a concrete setting, both complexity issues [15] and strategic considerations [29] could prevent agents from successfully identifying individually rational deals. This is why the "anytime character" [35] of this framework is so important: even if the social optimum cannot be reached, every single deal that agents *do* agree on will result in an overall improvement from the social point of view.

3.5 Necessary Deals

The next theorem corresponds to Sandholm's main result regarding necessary contract types [35].[5] It states that for any system (consisting of a set of agents $\mathcal{A}$ and a set of resources $\mathcal{R}$) and any (not independently decomposable) deal δ for that system, it is possible to construct utility functions and choose an initial allocation of resources such that δ is *necessary* to reach an optimal allocation, if agents only agree to individually rational deals. All other findings on the insufficiency of certain types of contracts reported in [35] may be considered corollaries to this. For instance, the fact that, say, cluster deals alone are not sufficient to guarantee optimal outcomes follows from this theorem if we take δ to be any particular swap deal for the system in question.

Theorem 2 (Necessary deals with side payments) *Let the sets of agents and resources be fixed. Then for every deal δ that is not independently decomposable, there exist utility functions and an initial allocation such that any sequence of individually rational deals leading to an allocation with maximal utilitarian social welfare must include δ.*

[5] In fact, our theorem corrects a mistake in previous expositions of this result [18, 35], where the restriction to deals that are not independently decomposable had been omitted.

Proof. Given a set of agents $\mathcal{A}$ and a set of resources $\mathcal{R}$, let $\delta = (A, A')$ with $A \neq A'$ be any deal for this system. We need to show that there are a collection of utility functions and an initial allocation such that δ is necessary to reach an allocation with maximal social welfare. This would be the case if A' had maximal social welfare, A had the second highest social welfare, and A was the initial allocation of resources. As we have $A \neq A'$, there must be an agent $j \in \mathcal{A}$ such that $A(j) \neq A'(j)$. We now fix utility functions u_i for agents $i \in \mathcal{A}$ and sets of resources $R \subseteq \mathcal{R}$ as follows:

$$u_i(R) \quad = \quad \begin{cases} 2 & \text{if } R = A'(i) \text{ or } (R = A(i) \text{ and } i \neq j) \\ 1 & \text{if } R = A(i) \text{ and } i = j \\ 0 & \text{otherwise} \end{cases}$$

We get $sw_u(A') = 2 \cdot |\mathcal{A}|$ and $sw_u(A) = sw_u(A') - 1$. Because $\delta = (A, A')$ is not individually decomposable, there exists no allocation B different from both A and A' such that either $B(i) = A(i)$ or $B(i) = A'(i)$ for all agents $i \in \mathcal{A}$. Hence, $sw_u(B) < sw_u(A)$ for any other allocation B. That is, A' is the (unique) allocation with maximal social welfare and the only allocation with higher social welfare than A. Therefore, if we make A the initial allocation then $\delta = (A, A')$ would be the only deal increasing social welfare. By Lemma 1, this means that δ would be the only individually rational (and thereby the only possible) deal. Hence, δ is indeed necessary to achieve maximal utilitarian social welfare. □

By Theorem 2, any negotiation protocol that puts restrictions on the structure of deals that may be proposed will fail to guarantee optimal outcomes, even when there are no constraints on either time or computational resources. This emphasises the high complexity of our negotiation framework (see also [9, 13, 15, 17]).

To see that the restriction to deals that are not independently decomposable matters, consider a scenario with four agents and two resources. If the deal δ of moving r_1 from agent 1 to agent 2, and r_2 from agent 3 to agent 4 is individually rational, then so will be either one of the two "subdeals" of moving either r_1 from agent 1 to agent 2 or r_2 from agent 3 to agent 4. Hence, the deal δ (which is independently decomposable) cannot be necessary in the sense of Theorem 2 (with reference to our proof above, in the case of δ there *are* allocations B such that either $B(i) = A(i)$ or $B(i) = A'(i)$ for all agents $i \in \mathcal{A}$, *i.e.* we could get $sw_u(B) = sw_u(A')$).

3.6 Unlimited Amounts of Money

An implicit assumption made in the framework that we have presented so far is that every agent has got an *unlimited amount of money* available to it to be able to pay other agents whenever this is required for a deal that would increase utilitarian social welfare. Concretely, if A is the initial allocation and A' is the allocation with maximal utilitarian social welfare, then agent i may require an amount of money just below the difference $u_i(A') - u_i(A)$ to be able to get through the negotiation process. In the context of task contracting, for which this framework has been proposed originally [35], this may be

justifiable, at least if we are mostly interested in the reallocation of tasks and consider "money" merely a convenient way of keeping track of the utility transfers between friendly agents. For resource allocation problems, on the other hand, it seems questionable to make assumptions about the unlimited availability of one particular resource, namely money.

Note that the fact that money does not necessarily have to be a physical commodity does not defeat this argument. While, indeed, *during* negotiation agents may well use a virtual form of money and could arrange deals that involve higher amounts of money as side payments than they actually own, once negotiation has terminated the deals agreed upon have to be implemented and appropriate amounts of money (or their "real world" equivalents) have to be paid out.

Sandholm [35] also suggests to allow for a special cost value ∞ associated with tasks that an agent is *unable* to carry out. While this adds to the variety of scenarios that can be modelled in this framework, it also further aggravates the aforementioned problem of requiring unlimited amounts of money to bargain with. If we were to transfer this idea to our resource allocation scenarios, we could extend the domain of utility functions to include two special values ∞ and $-\infty$. The intended interpretation of, say, $u_i(R) = \infty$ would be that agent i would be prepared to pay just about *any price* in order to obtain the resources in R, while $u_i(R) = -\infty$ may be read as agent i having to get rid of the set of resources R, again, at all costs. Unfortunately, these are not just figures of speech. Indeed, if we were to include either ∞ or $-\infty$ into our negotiation framework, then we would have to make the assumption that agents have truly unlimited amounts of money at their disposal —otherwise the theoretical results of [35] and the corresponding results presented here will not apply anymore.

4 Rational Negotiation without Side Payments

As argued before, making assumptions about the unlimited availability of money to compensate other agents for disadvantageous deals is not realistic for all application domains. In this section, we investigate to what extent the theoretical results of [35] and the previous section still apply to negotiation processes *without* monetary side payments.[6]

4.1 An Example

In a scenario without money, that is, if we do not allow for compensatory payments, we cannot always guarantee an outcome with maximal utilitarian social welfare. To see this, consider the following simple problem for a system with two agents, agent 1 and agent 2, and a single resource r. The agents' utility functions are defined as follows:

$u_1(\{\,\}) = 0$	$u_2(\{\,\}) = 0$
$u_1(\{r\}) = 4$	$u_2(\{r\}) = 7$

[6]The results presented in this section have originally been published in [18].

Now suppose agent 1 initially owns the resource. Then passing r from agent 1 to agent 2 would increase utilitarian social welfare by an amount of 3. For the framework *with* money, agent 2 could pay agent 1, say, the amount of 5.5 units and the deal would be individually rational for both of them. Without money (*i.e.* if $p \equiv 0$), however, no individually rational deal is possible and negotiation must terminate with a non-optimal allocation.

4.2 Cooperative Rationality

As maximising social welfare is not generally possible, instead we are going to investigate whether a *Pareto optimal* outcome is possible in the framework without money, and what types of deals are sufficient to guarantee this. In the context of our utilitarian framework, an allocation of resources is Pareto optimal iff there is no other allocation where social welfare is higher while no single agent has lower utility.

Definition 9 (Pareto optimality) *An allocation A is called Pareto optimal iff there is no allocation A' such that $sw_u(A) < sw_u(A')$ and $u_i(A) \leq u_i(A')$ for all $i \in \mathcal{A}$.*

This formulation is equivalent to the more commonly used one: "An agreement is Pareto optimal if there is no other agreement [...] that is better for some of the agents and not worse for the others." (quoted after [25]).

As will become clear in due course, in order to get a sufficiency result, we need to relax the notion of individual rationality a little. For the framework without money, we also want agents to agree to a deal, if this at least maintains their utility (that is, no strict increase is necessary). This is a reasonable additional requirement for scenarios where agents can be assumed to be *cooperative*, at least to the degree of not being explicitly malicious. However, we are still going to require at least one agent to strictly increase their utility. This could, for instance, be the agent proposing the deal in question. (It would make little sense, even for a cooperative agent, to actively propose a deal that would not result in at least a small payoff.) We call deals conforming to this criterion *cooperatively rational*.

Definition 10 (Cooperative rationality) *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 $u_j(A) < u_j(A')$.*

In analogy to Lemma 1, we still have $sw_u(A) < sw_u(A')$ for any deal $\delta = (A, A')$ that is cooperatively rational, but *not* vice versa.

Lemma 3 (Cooperative rationality and utilitarian social welfare) *If a deal $\delta = (A, A')$ is cooperatively rational then $sw_u(A) < sw_u(A')$.*

Proof. This is an immediate consequence of Definitions 6 and 10. □

We call the instance of our negotiation framework where all deals are cooperatively rational (and hence do not include a monetary component) the model of *rational negotiation without side payments*.

4.3 Ensuring Pareto Optimal Outcomes

As the next theorem will show, the class of cooperatively rational deals is sufficient to guarantee a Pareto optimal outcome of money-free negotiation. It constitutes the analogue to Theorem 1 for the model of rational negotiation without side payments. We begin by establishing termination for this framework.

Lemma 4 (Termination) *There can be no infinite sequence of cooperatively rational deals.*

Proof. Every cooperatively rational deal strictly increases utilitarian social welfare (Lemma 3).[7] Together with the fact that there are only finitely many different allocations of resources, this implies that any negotiation process will eventually terminate. □

We are now ready to state the theorem.

Theorem 3 (Pareto optimal outcomes) *Any sequence of cooperatively rational deals will eventually result in a Pareto optimal allocation of resources.*

Proof. By Lemma 4, any sequence of cooperatively rational deals must eventually terminate. For the sake of contradiction, assume negotiation ends with allocation A, but A is not Pareto optimal.

The latter means that there exists another allocation A' with $sw_u(A) < sw_u(A')$ and $u_i(A) \leq u_i(A')$ for all $i \in \mathcal{A}$. If we had $u_i(A) = u_i(A')$ for all $i \in \mathcal{A}$, then also $sw_u(A) = sw_u(A')$; that is, there must be at least one $j \in \mathcal{A}$ with $u_j(A) < u_j(A')$. But then the deal $\delta = (A, A')$ would be cooperatively rational, which contradicts our assumption of A being a terminal allocation. □

Observe that the proof would not have gone through if deals were required to be strictly rational (without side payments), as this would necessitate $u_i(A) < u_i(A')$ for all $i \in \mathcal{A}$. Cooperative rationality means, for instance, that agents would be prepared to give away resources to which they assign a utility value of 0, without expecting anything in return. In the framework with money, another agent could always offer such an agent an infinitesimally small amount of money, who would then accept the deal.

Therefore, our proposed weakened notion of rationality seems indeed a very reasonable price to pay for giving up money.

[7] This is where we need the condition that at least one agent behaves *truly* individually rational for each deal.

4.4 Necessity Result

As our next result shows, also for the framework without side payments, deals of any structural complexity may be necessary in order to be able to guarantee an optimal outcome of a negotiation.[8]

Theorem 4 (Necessary deals without side payments) *Let the sets of agents and resources be fixed. Then for every deal δ that is not independently decomposable, there exist utility functions and an initial allocation such that any sequence of cooperatively rational deals leading to a Pareto optimal allocation would have to include δ.*

Proof. Let $\delta = (A, A')$ with $A \neq A'$. We try to fix utility functions u_i in such a way that A' has the highest and A has the second highest social welfare, and that $u_i(A) \leq u_i(A')$ for all agents $i \in \mathcal{A}$. As we have $A \neq A'$, there must be a $j \in \mathcal{A}$ such that $A(j) \neq A'(j)$. We now define utility functions as follows:

$$u_i(R) \quad = \quad \begin{cases} 2 & \text{if } R = A'(i) \text{ or } (R = A(i) \text{ and } i \neq j) \\ 1 & \text{if } R = A(i) \text{ and } i = j \\ 0 & \text{otherwise} \end{cases}$$

We get $sw_u(A') = 2 \cdot |\mathcal{A}|$ and $sw_u(A) = sw_u(A') - 1$. Because $\delta = (A, A')$ is not individually decomposable, there exists no allocation B different from both A and A' such that either $B(i) = A(i)$ or $B(i) = A'(i)$ for all agents $i \in \mathcal{A}$. Hence, $sw_u(B) < sw_u(A)$ for any other allocation B. We also have $u_i(A) \leq u_i(A')$ for all $i \in \mathcal{A}$. Hence, A is not Pareto optimal, but A' is. If we make A the initial allocation, then δ would be the only cooperatively rational deal (as every other deal would decrease social welfare), *i.e.* δ is indeed necessary to guarantee a Pareto optimal outcome. □

Observe that, while this proof has been very similar to the proof of Theorem 2, this time we also required the additional condition of $u_i(A) \leq u_i(A')$ for all $i \in \mathcal{A}$.

It is interesting to compare Theorems 3 and 4 with a recent result of McBurney et al. [27], which states, quite generally, that whenever agents, that are "purely self-interested and without malice, engage freely and without duress in a negotiation dialogue" using a protocol that is "inclusive" (no agent is prevented from participating), "transparent" (the rules of the game are known to all agents), and "fair" (all agents are treated equally), and whenever that dialogue "is conducted with neither time constraints nor processing-resource constraints", then the outcome reached will be Pareto optimal. All these side-constraints are fulfilled in our abstract framework, where the behaviour of agents is essentially governed by the notion of cooperative rationality. Therefore, Theorem 4 suggests that we have to interpret the quoted lack of "processing-resource constraints" at least in the following broad sense. Firstly, agents need sufficient computational resources to be able to propose and evaluate the required sequence of deals.

[8] This theorem corrects a mistake in the original statement of the result [18], where the restriction to deals that are not independently decomposable had been omitted.

(This is what is commonly understood by lack of processing-resource constraints.) Secondly, to be able to *communicate* proposals we also require a negotiation protocol based on an agent communication language that is rich enough to represent *every* possible deal. Amongst other things, this means that the protocol must allow for more than just two agents to agree on a transaction (namely in the case of multiagent deals).

5 Special Domains with Restricted Utility Functions

Theorems 2 and 4 are negative results in the sense that they show that deals of any complexity may be required in order to guarantee an optimal outcome of a particular negotiation. This is partly a consequence of the high degree of generality of our framework. In Section 2.1, we have defined utility functions as *arbitrary* functions from sets of resources to real numbers. For many application domains this may be unnecessarily general or even inappropriate and we may be able to obtain stronger results for specific classes of utility functions. In this section, we consider some examples.[9]

Clearly, the sufficiency results established in Theorems 1 and 3 will still apply, whatever restrictions we may put on utility functions. Interesting new results could be of two kinds: (i) either that a (structurally) weaker deal type, such as the class of one-resource-at-a-time deals, is sufficient for certain domains, or (ii) that the general deals are still necessary, *even* for a restricted class of utility functions.

5.1 Basic Restrictions

In general, there may be certain resources we would like to assign a negative utility to (e.g. "five tons of radioactive waste"), but in many domains *non-negative* utility functions will suffice.

Definition 11 (Non-negative utility) *We call a utility function u_i non-negative iff $u_i(R) \geq 0$ holds for every set of resources $R \subseteq \mathcal{R}$.*

An inspection of the particular utility functions used in the proofs of Theorems 2 and 4 reveals that all results on the necessity of deals still apply for scenarios where utility functions are required to be non-negative. As we shall see next, this will not be the case anymore if we add a further, seemingly innocent, restriction.

A slightly stronger requirement than non-negative utility would be to assign at least a small positive value to every non-empty set of resources. We are going to call utility functions of this description *positive* utility functions.

Definition 12 (Positive utility) *We call a utility function u_i positive iff it is non-negative and $u_i(R) \neq 0$ holds for all sets of resources $R \subseteq \mathcal{R}$ with $R \neq \{\}$.*

If we were to restrict ourselves to positive utility functions, then the result of Theorem 4 would *not* hold anymore. To see this, observe that any deal that would involve a particular

[9] Our discussion of special domains with restricted utility functions follows [18].

agent (with a positive utility function) giving away all its resources without receiving anything in return could never be cooperatively rational. Hence, such a deal could never be necessary to achieve a Pareto optimal allocation either, because this would contradict Theorem 3, which states that the set of cooperatively rational deals alone is sufficient to guarantee a Pareto optimal outcome.

Other natural classes of functions to consider would be *monotonic* utility functions (where any set of resources must be valued at least as high as any of its subsets) or *bounded* utility functions (where utility values may only fall within a certain interval). Bounded utility functions, for instance, are useful in systems where social welfare is measured using the utilitarian collective welfare function sw_u, as this prevents single agents from having an overly strong influence on social welfare (by assigning prohibitively high utility values to their most preferred bundles). However, we are not going to investigate either monotonic or bounded utility functions any further in this paper.

5.2 Additive Scenarios

We call a utility function u_i *additive* iff the value ascribed to a set of resources is always the sum of the values of its members. This corresponds to the notion of *modular* task-oriented domains discussed by Rosenschein and Zlotkin [31]. Additive utility functions are appropriate for scenarios where combining resources does not result in any synergy effects (in the sense of increasing an agent's welfare).

Definition 13 (Additive utility) *We call a utility function u_i additive iff the following holds for every set of resources $R \subseteq \mathcal{R}$:*

$$u_i(R) \quad = \quad \sum_{r \in R} u_i(\{r\})$$

We refer to systems where all agents have additive utility functions as *additive scenarios*. The following theorem shows that for these additive scenarios the simple one-resource-at-a-time deal type is sufficient to guarantee outcomes with maximal utilitarian social welfare in the framework of rational negotiation with side payments.[10]

Theorem 5 (Additive scenarios) *In additive scenarios, any sequence of individually rational one-resource-at-a-time deals will eventually result in an allocation with maximal utilitarian social welfare.*

Proof. Termination follows from Lemma 2. We are going to show that, whenever the current allocation does not have maximal social welfare, then there is still a possible one-resource-at-a-time deal that is individually rational.

In additive domains, the utilitarian social welfare of a given allocation may be computed by adding up the appropriate utility values for all the single resources in $\mathcal{R}$. For any allocation A, let f_A be the function mapping each resource $r \in \mathcal{R}$ to the agent $i \in \mathcal{A}$

[10] This has also been observed by T. Sandholm (personal communication, September 2002).

that holds r in situation A (that is, we have $f_A(r) = i$ iff $r \in A(i)$). The utilitarian social welfare for allocation A is then given by the following formula:

$$sw_u(A) = \sum_{r \in \mathcal{R}} u_{f_A(r)}(\{r\})$$

Now suppose that negotiation has terminated with allocation A and there are no more individually rational one-resource-at-a-time deals possible. Furthermore, for the sake of contradiction, assume that A is *not* an allocation with maximal social welfare, *i.e.* there exists another allocation A' with $sw_u(A) < sw_u(A')$. But then, by the above characterisation of social welfare for additive scenarios, there must be at least one resource $r \in \mathcal{R}$ such that $u_{f_A(r)}(\{r\}) < u_{f_{A'}(r)}(\{r\})$. That is, the one-resource-at-a-time deal δ of passing r from agent $f_A(r)$ on to agent $f_{A'}(r)$ would increase social welfare. Therefore, by Lemma 1, δ must be an individually rational deal, *i.e.* contrary to our earlier assumption, A cannot be a terminal allocation. Hence, A must be an allocation with maximal utilitarian social welfare. □

Additive functions that assign a positive utility to each single resource are a special case of the class of monotonic utility functions. We note here that Theorem 5 does *not* extend to this larger class of domains. This may be seen, for instance, by revisiting the example given in Section 3.2, where the utility functions are monotonic but not additive.

Before we move on to the case of negotiation without side payments, we briefly mention two recent results that both extend, in different ways, the result stated in Theorem 5 (a detailed discussion, however, would be beyond the scope of the present paper). The first of these results shows that rational deals involving at most k resources each are sufficient for convergence to an allocation with maximal social welfare whenever all utility functions are *additively separable* with respect to a common partition of $\mathcal{R}$ (*i.e.* synergies across different parts of the partition are not possible and overall utility is defined as the sum of utilities for the different sets in the partition [21]), and each set in this partition has at most k elements [10]. The second result concerns a *maximality property* of utility functions with respect to one-resource-at-a-time deals. Chevaleyre et al. [11] show that a class of utility functions that is only slightly more general than the class of additive functions considered here (namely, it is possible to assign a non-zero utility to the empty bundle) is maximal in the sense that for no class of functions strictly including that class it would still be possible to guarantee that agents using utility functions from that larger class and negotiating only individually rational one-resource-at-a-time deals will eventually reach an allocation with maximal social welfare in all cases.

5.3 0-1 Scenarios

The last special domain we are going to consider are scenarios where agents use additive utility functions that assign either 0 or 1 to every single resource. This may be sufficient if we simply wish to distinguish whether or not the agent *needs* a particular resource (to execute a given plan, for example). This is, for instance, the case for some of the agents defined in [34].

Definition 14 (0-1 utility) *We call a utility function u_i a 0-1 function iff it is additive and $u_i(\{r\}) = 0$ or $u_i(\{r\}) = 1$ for every single resource $r \in \mathcal{R}$.*

As the following theorem shows, for *0-1 scenarios* (*i.e.* for systems where all utility functions are 0-1 functions), the one-resource-at-a-time deal type is sufficient to guarantee maximal utilitarian social welfare, even in the framework *without* monetary side payments (where all deals are required to be cooperatively rational).

Theorem 6 (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 social welfare.*

Proof. Termination follows from Lemma 4. If an allocation A does not have maximal social welfare then it must be the case that some agent i holds a resource r with $u_i(\{r\}) = 0$ and there is another agent j in the system with $u_j(\{r\}) = 1$. Passing r from i to j would be a cooperatively rational deal, so either negotiation has not yet terminated or we are indeed in a situation with maximal utilitarian social welfare. □

This result may be interpreted as a formal justification for some of the negotiation strategies proposed in [34].

6 Egalitarian Agent Societies

The utilitarian programme, *i.e.* the idea of trying to maximise the sum of all utilities of the members of a society, is often taken for granted in the multiagent systems literature (see e.g. [27, 31, 36, 40]). This is not the case in welfare economics and social choice theory, where different notions of social welfare are being studied and compared with each other. Here, the concept of *egalitarian social welfare* takes a particularly prominent role [4, 28, 38]. In an egalitarian system one would consider any differences in individual welfare unjust, unless removing these differences would inevitably result in reducing the welfare of the agent who is currently worst off even further. This is Rawls' so-called *difference principle* [30]. In other words, the first and foremost objective of such a society would be to maximise the welfare of its weakest member.

In this section, we are going to argue that egalitarian principles may be of interest to multiagent systems research as well, and we are going to develop the framework of resource allocation by negotiation for *egalitarian agent societies*.[11]

6.1 The "Veil of Ignorance" in Multiagent Systems

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,

[11] The concept of egalitarian agent societies has been introduced in [19]. This is also where the technical results of this section have first been published.

addressed the respective benefits and drawbacks of utilitarianism on the one hand and egalitarianism on the other [23, 30, 38]. 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 (a precise definition will be given in Section 6.2).

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* [30]. 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 themselves 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. According to Rawls, 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.[12] 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 studied by Lemaître et al. [26], 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

[12] A possible counter-argument would be, for instance, that a society governed by utilitarian principles would maximise their members' *expected* utility, which may be considered a rational reason for opting for a utilitarian system [23].

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. In summary, what definition of social welfare is appropriate depends on the application under consideration. While, so far, we have focussed on systems governed by utilitarian principles, in the remainder of this paper we are going to consider several alternative approaches, starting with egalitarian systems in this section.

6.2 Egalitarian Welfare Orderings

Given the preference profiles of the individual agents in a society (which, in our framework, are represented by means of their utility functions), a *social welfare ordering* over alternative allocations of resources formalises the notion of a society's preferences (an example would be the ordering induced by the utilitarian collective utility function sw_u) [3]. Next we are going to introduce two further such orderings, the so-called egalitarian maximin- and the leximin-orderings, both of which are standard concepts in social choice theory and welfare economics [28].

The first goal of an egalitarian society should be to increase the welfare of its weakest member [28, 30, 38]. 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 15 (Egalitarian social welfare) *The egalitarian social welfare $sw_e(A)$ of an allocation of resources A is defined as follows:*

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

The egalitarian collective utility function sw_e gives rise to a social welfare ordering over alternative allocations of resources: A' is strictly preferred over A iff $sw_e(A) < sw_e(A')$. This ordering is sometimes called the *maximin-ordering*. An allocation A is said to have *maximal egalitarian social welfare* iff there is no other allocation A' such that $sw_e(A) < sw_e(A')$.

The maximin-ordering only takes into account the welfare of the currently weakest agent, but is insensitive to utility fluctuations in the rest of society. To allow for a finer distinction of the social welfare of different allocations we introduce the so-called *leximin-ordering*.

For a society with n agents, let $\{u_1, \ldots, u_n\}$ be the set of utility functions for that society. Then every allocation A determines a utility vector $\langle u_1(A), \ldots, u_n(A)\rangle$ 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 ith element in such a vector (for $1 \leq i \leq |\mathcal{A}|$). That is, $\vec{u}_1(A)$ for instance, is the utility value assigned to allocation A by the currently weakest agent. We now declare a *lexicographic ordering* over vectors of real numbers (such as $\vec{u}(A)$) in the usual way: $\vec{x}$ lexicographically precedes $\vec{y}$ iff $\vec{x}$ is a (proper) prefix of $\vec{y}$ or $\vec{x}$ and $\vec{y}$ share a common (proper) prefix of length k (which may be 0) and we have $\vec{x}_{k+1} < \vec{y}_{k+1}$.

Definition 16 (Leximin-ordering) *The leximin-ordering $\prec$ over alternative allocations of resources is defined as follows:*

$$A \prec A' \quad \textit{iff} \quad \vec{u}(A) \textit{ lexicographically precedes } \vec{u}(A')$$

We write $A \preceq A'$ iff either $A \prec A'$ or $\vec{u}(A) = \vec{u}(A')$. An allocation of resources A is called *leximin-maximal* iff there is no other allocation A' such that $A \prec A'$.

Let us note some simple consequences of Definitions 15 and 16. It is easily seen that $sw_e(A) < sw_e(A')$ implies $A \prec A'$, because the former requires already the element at the *first* position in the ordered utility vector of A to be smaller than that of the ordered utility vector of A'. Also note that $A \preceq A'$ implies $sw_e(A) \leq sw_e(A')$. Finally, every leximin-maximal allocation has maximal egalitarian social welfare, but not vice versa.

6.3 An Example

We illustrate Definition 16 and the use of ordered utility vectors by means of an example. Consider a society with three agents and two resources, with the agents' utility functions given by the following table:

$u_1(\{\}) = 0$	$u_2(\{\}) = 0$	$u_3(\{\}) = 0$
$u_1(\{r_1\}) = 5$	$u_2(\{r_1\}) = 4$	$u_3(\{r_1\}) = 2$
$u_1(\{r_2\}) = 3$	$u_2(\{r_2\}) = 2$	$u_3(\{r_2\}) = 6$
$u_1(\{r_1, r_2\}) = 8$	$u_2(\{r_1, r_2\}) = 17$	$u_3(\{r_1, r_2\}) = 7$

First of all, we observe that the egalitarian social welfare will be 0 for any possible allocation in this scenario, because at least one of the agents would not get any resources at all. Let A be the allocation where agent 2 holds the full bundle of resources. The corresponding utility vector is $\langle 0, 17, 0\rangle$, *i.e.* $\vec{u}(A) = \langle 0, 0, 17\rangle$. (Note that A is the allocation with maximal utilitarian social welfare.) Furthermore, let A' be the allocation where agent 1 gets r_1, agent 2 gets r_2, and agent 3 has to be content with the empty bundle. Now we get an ordered utility vector of $\langle 0, 2, 5\rangle$. The initial element in either vector is 0, but $0 < 2$, *i.e.* $\vec{u}(A)$ lexicographically precedes $\vec{u}(A')$. Hence, we get $A \prec A'$, *i.e.* A' would be the socially preferred allocation with respect to the leximin-ordering.

6.4 Pigou-Dalton Transfers and Equitable Deals

Our next aim is to identify a suitable criterion that agents inhabiting an egalitarian agent society may use to decide whether or not to accept a particular deal. Clearly, cooperatively rational deals, for instance, would not be an ideal choice, because Pareto optimal allocations will typically not be optimal from an egalitarian point of view [28].

When searching the economics literature for a class of deals that would benefit society in an egalitarian system we soon encounter *Pigou-Dalton transfers*. The Pigou-Dalton *principle* states that whenever a utility transfer between two agents takes place which reduces the difference in utility between the two, then that transfer should be considered socially beneficial [28]. In the context of our framework, a Pigou-Dalton transfer (between agents i and j) can be defined as follows.

Definition 17 (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 in this definition could be relaxed to postulate $u_i(A) + u_j(A) \leq u_i(A') + u_j(A')$, 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 negotiation outcomes with maximal egalitarian social welfare? Consider the following example:

$u_1(\{\,\}) = 0$	$u_2(\{\,\}) = 0$
$u_1(\{r_1\}) = 3$	$u_2(\{r_1\}) = 5$
$u_1(\{r_2\}) = 12$	$u_2(\{r_2\}) = 7$
$u_1(\{r_1, r_2\}) = 15$	$u_2(\{r_1, r_2\}) = 17$

The first agent attributes a relatively low utility value to r_1 and a high one to r_2. Furthermore, the value of both resources together is simply the sum of the individual utilities, *i.e.* agent 1 is using an additive utility function (no synergy effects). The second agent 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 $A(1) = \{r_1\}$ and $A(2) = \{r_2\}$. The "inequality index" for this allocation is $|u_1(A) - u_2(A)| = 4$. We can easily check that inequality is in fact minimal for allocation A (which means that there can be no inequality-reducing deal, and certainly no Pigou-Dalton transfer, given this allocation). However, allocation A' with $A'(1) = \{r_2\}$ and $A'(2) = \{r_1\}$ would result in a higher level of 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.

Intuitively, agents operating according to egalitarian principles should help any of their fellow agents that are worse off than they are themselves (as long as they can afford to do so without themselves ending up even worse). This means, the purpose of any exchange of resources should be to improve the welfare of the weakest agent involved in the respective deal. We formalise this idea by introducing the class of *equitable* deals.

Definition 18 (Equitable deals) *A deal $\delta = (A, A')$ is called equitable iff it satisfies the following criterion:*

$$min\{u_i(A) \mid i \in \mathcal{A}^\delta\} \quad < \quad min\{u_i(A') \mid i \in \mathcal{A}^\delta\}$$

Recall that $\mathcal{A}^\delta = \{i \in \mathcal{A} \mid A(i) \neq A'(i)\}$ denotes the set of agents involved in the deal δ. Given that for $\delta = (A, A')$ to be a deal we require $A \neq A'$, $\mathcal{A}^\delta$ can never be the empty set (*i.e.* the minima referred to in above definition are well-defined).

It is easy to see that any Pigou-Dalton transfer will also be an equitable deal, because it will always result in an improvement for the weaker one of the two agents concerned. The converse, however, does not hold (not even if we restrict ourselves to deals involving only two agents). In fact, equitable deals may even increase the inequality of the agents concerned, namely in cases where the happier agent gains more utility than the weaker does.

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).

6.5 Local Actions and their Global Effects

We are now going to prove two lemmas that provide the connection between the local acceptability criterion given by the notion of equitability and the two egalitarian social welfare orderings discussed earlier.

The first lemma shows how global changes are reflected locally. If a deal happens to increase (global) egalitarian social welfare, that is, if it results in a rise with respect to the maximin-ordering, then that deal will in fact be an equitable deal.

Lemma 5 (Maximin-rise implies equitability) *If A and A' are allocations with $sw_e(A) < sw_e(A')$, then $\delta = (A, A')$ is an equitable deal.*

Proof. Let A and A' be allocations with $sw_e(A) < sw_e(A')$ and let $\mathcal{A}^\delta$ be the set of agents involved in the deal $\delta = (A, A')$ (see Definition 3). Any agent with minimal utility for allocation A must be involved in δ, because egalitarian social welfare, and thereby these agents' individual utility, is higher for allocation A'. That is, we have $min\{u_i(A) \mid i \in \mathcal{A}^\delta\} = sw_e(A)$. Furthermore, because $\mathcal{A}^\delta \subseteq \mathcal{A}$, we certainly have $sw_e(A') \leq min\{u_i(A') \mid i \in \mathcal{A}^\delta\}$.

Together with our original assumption of $sw_e(A) < sw_e(A')$, we now obtain the inequation $min\{u_i(A) \mid i \in \mathcal{A}^\delta\} < min\{u_i(A') \mid i \in \mathcal{A}^\delta\}$. This shows that δ will indeed be an equitable deal. □

Observe that the converse does not hold; not every equitable deal will necessarily increase egalitarian social welfare. This is for instance not the case if only agents who are currently better off are involved in a deal. In fact, there can be no class of deals (that could be defined without reference to the *full* set of agents in a society) that will always result in an increase in egalitarian social welfare. This is a consequence of the fact that the maximin-ordering induced by sw_e is not separable.[13] To be able to detect changes in welfare

[13] A social welfare ordering is called *separable* iff the effect of a local welfare redistribution with respect to that ordering (rise or fall) is independent of non-concerned agents [28].

resulting from an equitable deal we require the finer differentiation between alternative allocations of resources given by the leximin-ordering. In fact, as we shall see next, any equitable deal can be shown to result in a strict improvement with respect to the leximin-ordering.

Lemma 6 (Equitability implies leximin-rise) *If $\delta = (A, A')$ is an equitable deal, then $A \prec A'$.*

Proof. Let $\delta = (A, A')$ be a deal that satisfies the equitability criterion and define $\alpha = min\{u_i(A) \mid i \in \mathcal{A}^\delta\}$. The value α may be considered as partitioning the ordered utility vector $\vec{u}(A)$ into three subvectors: Firstly, $\vec{u}(A)$ has got a (possibly empty) prefix $\vec{u}(A)^{<\alpha}$ where all elements are strictly lower than α. In the middle, it has got a subvector $\vec{u}(A)^{=\alpha}$ (with at least one element) where all elements are equal to α. Finally, $\vec{u}(A)$ has got a suffix $\vec{u}(A)^{>\alpha}$ (which again may be empty) where all elements are strictly greater than α.

By definition of α, the deal δ cannot affect agents whose utility values belong to $\vec{u}(A)^{<\alpha}$. Furthermore, by definition of equitability, we have $\alpha < min\{u_i(A') \mid i \in \mathcal{A}^\delta\}$, which means that all of the agents that *are* involved will end up with a utility value which is strictly greater than α, and at least one of these agents will come from $\vec{u}(A)^{=\alpha}$. We now collect the information we have on $\vec{u}(A')$, the ordered utility vector of the second allocation A'. Firstly, it will have a prefix $\vec{u}(A')^{<\alpha}$ identical to $\vec{u}(A)^{<\alpha}$. This will be followed by a (possibly empty) subvector $\vec{u}(A')^{=\alpha}$ where all elements are equal to α and which must be strictly shorter than $\vec{u}(A)^{=\alpha}$. All of the remaining elements of $\vec{u}(A')$ will be strictly greater than α. It follows that $\vec{u}(A)$ lexicographically precedes $\vec{u}(A')$, *i.e.* $A \prec A'$ holds as claimed. □

Again, the converse does not hold, *i.e.* not every deal resulting in a leximin-rise is necessarily equitable. Counterexamples are deals where the utility value of the weakest agent involved stays constant, despite there being an improvement with respect to the leximin-ordering at the level of society.

A well-known result in welfare economics states that every Pigou-Dalton utility transfer results in a leximin-rise [28]. Given that we have observed earlier that every deal that amounts to a Pigou-Dalton transfer will also be an equitable deal, this result can now also be regarded as a simple corollary to Lemma 6.

6.6 Maximising Egalitarian Social Welfare

Our next aim is to prove a sufficiency result for the egalitarian framework (in analogy to Theorems 1 and 3). We are going to show that systems where agents negotiate equitable deals always converge towards an allocation of resources with maximal egalitarian social welfare. We first prove the appropriate termination lemma.

Lemma 7 (Termination) *There can be no infinite sequence of equitable deals.*

Proof. By Lemma 6, any equitable deal will result in a strict rise with respect to the leximin-ordering $\prec$ (which is both irreflexive and transitive). Hence, as there are only a finite number of distinct allocations, negotiation will have to terminate after a finite number of deals. □

The proof of the following theorem shows that equitable deals are sufficient for agents to reach an allocation of resources with maximal egalitarian social welfare. As for our earlier sufficiency theorems, the result is even stronger than this: *any* sequence of equitable deals will eventually result in an optimal allocation. That is, agents may engage "blindly" into negotiation. Whatever their course of action, provided they restrict themselves to equitable deals, once they reach an allocation where no further equitable deals are possible, that allocation is bound to have maximal egalitarian welfare.

Theorem 7 (Maximal egalitarian social welfare) *Any sequence of equitable deals will eventually result in an allocation of resources with maximal egalitarian social welfare.*

Proof. By Lemma 7, negotiation will eventually terminate if all deals are required to be equitable. So suppose negotiation has terminated and no more equitable deals are possible. Let A be the corresponding terminal allocation of resources. The claim is that A will be an allocation with maximal egalitarian social welfare. For the sake of contradiction, assume it is not, *i.e.* assume there exists another allocation A' for the same system such that $sw_e(A) < sw_e(A')$. But then, by Lemma 5, the deal $\delta = (A, A')$ will be an equitable deal. Hence, there is still a possible deal, namely δ, which contradicts our earlier assumption of A being a terminal allocation. This shows that A will be an allocation with maximal egalitarian social welfare, which proves our claim. □

After having reached the allocation with maximal egalitarian social welfare, it may be the case that still some equitable deals are possible, although they would not increase social welfare any further (but they would still cause a leximin-rise). This can be demonstrated by means of a simple example. Consider a system with three agents and two resources. The following table fixes the utility functions:

$u_1(\{\,\}) = 0$	$u_2(\{\,\}) = 6$	$u_3(\{\,\}) = 8$
$u_1(\{r_1\}) = 5$	$u_2(\{r_1\}) = 7$	$u_3(\{r_1\}) = 9$
$u_1(\{r_2\}) = 0$	$u_2(\{r_2\}) = 6.5$	$u_3(\{r_2\}) = 8.5$
$u_1(\{r_1, r_2\}) = 5$	$u_2(\{r_1, r_2\}) = 7.5$	$u_3(\{r_1, r_2\}) = 9.5$

A possible interpretation of these functions would be the following. Agent 3 is fairly well off in any case; obtaining either of the resources r_1 and r_2 will not have a great impact on its personal welfare. The same is true for agent 2, although it is slightly less well off to begin with. Agent 1 is the poorest agent and attaches great value to r_1, but has no interest in r_2. Suppose agent 3 initially holds both resources. This corresponds to the ordered utility vector $\langle 0, 6, 9.5\rangle$. Passing r_1 to agent 1 would lead to a new allocation with the ordered utility vector $\langle 5, 6, 8.5\rangle$ and increase egalitarian social welfare to 5, which is

the maximal egalitarian social welfare that is achievable in this system. However, there is still another equitable deal that could be implemented from this latter allocation: agent 3 could offer r_2 to agent 2. Of course, this deal does not affect agent 1. The resulting allocation would then have the ordered utility vector $\langle 5, 6.5, 8\rangle$, which corresponds to the leximin-maximal allocation.

To be able to detect situations where a social welfare maximum has already been reached but some equitable deals are still possible, and to be able to stop negotiation (assuming we are only interested in maximising sw_e as quickly as possible), however, we would require a *global* criterion.[14] We could define a class of *strongly equitable* deals that are like equitable deals but on top of that require the (currently) weakest agent to be involved in the deal. This would be a sharper criterion, but it would also be against the spirit of distributivity and locality, because every single agent would be involved in every single deal (in the sense of everyone having to announce their utility in order to be able to determine who is the weakest).

From a purely practical point of view, Theorem 7 may be of a lesser interest than the corresponding results for utilitarian systems, because it does not refer to an 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.

6.7 Necessity Result

As our next theorem will show, if we restrict the set of admissible deals to those that are equitable, then every single deal δ (that is not independently decomposable) may be necessary to guarantee an optimal result (that is, no sequence of equitable deals excluding δ could possibly result in an allocation with maximal egalitarian social welfare).[15] This mirrors the necessity results for the two variants of the utilitarian negotiation framework (Theorems 2 and 4).

Theorem 8 (Necessary deals in egalitarian systems) *Let the sets of agents and resources be fixed. Then for every deal δ that is not independently decomposable, there exist utility functions and an initial allocation such that any sequence of equitable deals leading to an allocation with maximal egalitarian social welfare would have to include δ.*

Proof. Given a set of agents $\mathcal{A}$ and a set of resources $\mathcal{R}$, let $\delta = (A, A')$ be any deal for this system. As we have $A \neq A'$, there will be a (at least one) agent $j \in \mathcal{A}$ with $A(j) \neq A'(j)$. We use this particular j to fix suitable utility functions u_i for agents $i \in \mathcal{A}$

[14] This is again a consequence of the fact that the maximin-ordering is not separable. No measure that only takes the welfare of agents involved in a particular deal into account could be strong enough to always tell us whether or not the deal in question will result in an increase in egalitarian social welfare (see also our discussion after Lemma 5).

[15] This theorem corrects a mistake in the original statement of the result [19], where the restriction to deals that are not independently decomposable had been omitted.

and sets of resources $R \subseteq \mathcal{R}$ as follows:

$$u_i(R) \quad = \quad \begin{cases} 2 & \text{if } R = A'(i) \text{ or } (R = A(i) \text{ and } i \neq j) \\ 1 & \text{if } R = A(i) \text{ and } i = j \\ 0 & \text{otherwise} \end{cases}$$

That is, for allocation A' every agent assigns a utility value of 2 to the resources it holds. The same is true for allocation A, with the sole exception of agent j, who only assigns a value of 1. For any other allocation, agents assign the value of 0 to their set of resources, unless that set is the same as for either allocation A or A'. As δ is not decomposable, this will happen for at least one agent for every allocation different from both A and A'. Hence, for every such allocation at least one agent will assign a utility value of 0 to its allocated bundle. We get $sw_e(A') = 2$, $sw_e(A) = 1$, and $sw_e(B) = 0$ for every other allocation B, *i.e.* A' is the only allocation with maximal egalitarian social welfare.

The ordered utility vector of A' is of the form $\langle 2, \ldots, 2 \rangle$, that of A is of the form $\langle 1, 2, \ldots, 2 \rangle$, and that of any other allocation has got the form $\langle 0, \ldots \rangle$, *i.e.* we have $A \prec A'$ and $B \prec A$ for all allocations B with $B \neq A$ and $B \neq A'$. Therefore, if we make A the initial allocation of resources, then δ will be the only deal that would result in a rise with respect to the leximin-ordering. Thus, by Lemma 6, δ would also be the only equitable deal. Hence, if the set of admissible deals is restricted to equitable deals then δ is indeed necessary to reach an allocation with maximal egalitarian social welfare. □

This result shows, again, that there can be no simple class of deals (such as the class of deals only involving two agents at a time) that would be sufficient to guarantee an optimal outcome of negotiation.

6.8 Remarks on Restricted Utility Functions

In Section 5, we have seen that there are cases where the optimal outcome of a negotiation process may be guaranteed even when we admit only very specific types of deals, provided that we put suitable restrictions on the class of utility functions that agents may use to represent their valuation of different sets of resources. These results apply to the instances of our framework where social welfare is given a utilitarian interpretation. In the egalitarian setting, on the contrary, we have not been able to establish similar results. Even the (arguably) strongest restrictions used in the utilitarian case do not allow us to eliminate any type of deal in the egalitarian framework. Let us consider the example of *0-1 functions* (see Definition 14). By Theorem 6, this restriction guarantees an outcome with maximal utilitarian social welfare in the model of rational negotiation without side payments, even when only one-resource-at-a-time deals are possible.

This result does not hold anymore for egalitarian agent societies. Counterexamples can easily be constructed. Take, for instance, a scenario of three agents furnishing their flats. *Ann* needs a *picture* and has a *desk* which she does not need. *Bob* needs a *desk* and a *chair*, but only has the *chair*. *Carlos* needs a *picture*, a *chair*, and a *cushion*, and he only owns the *picture* and the *cushion* at the beginning of the negotiation process. The

ordered utility vector for this allocation is $\langle 0, 1, 2\rangle$. However, in the situation where *Ann* has the *picture*, *Bob* the *desk* instead of the *chair*, and *Carlos* the *chair* and the *cushion* is better; the corresponding ordered utility vector would be $\langle 1, 1, 2\rangle$. Unfortunately, only a very complex equitable deal (involving all three agents, namely *Carlos* giving the *picture* to *Ann*, *Ann* giving the *desk* to *Bob*, and *Bob* giving the *chair* to *Carlos*) would allow this agent society to reach the preferred allocation of resources.

7 Further Variations

In this section, we are going to consider three further notions of social welfare and discuss them in the context of our framework of resource allocation by negotiation.

In particular, we are going to analyse the case of so-called *Lorenz optimal* allocations of resources in detail. This is an approach for measuring social welfare that attempts to offer a compromise between the utilitarian and the egalitarian programmes discussed in previous sections. We also briefly consider *elitist* agent societies, where social welfare is tied to the welfare of the agent that is currently best off, and, finally, societies where *envy-free* allocations of resources are desirable.

7.1 Negotiating Lorenz Optimal Allocations

We are now going to introduce a social welfare ordering that combines utilitarian and egalitarian aspects of social welfare.[16] The basic idea is to endorse deals that result in an improvement with respect to utilitarian welfare without causing a loss in egalitarian welfare, and vice versa. An appropriate formalisation of social preference for this kind of agent society is given by the notion of Lorenz domination, which is another important social welfare ordering studied in the welfare economics literature [28].[17]

For the following definition, recall the concept of an ordered utility vector $\vec{u}(A)$ for an allocation A, which we have introduced in Section 6.2.

Definition 19 (Lorenz domination) *Let A and A' be allocations for a society with n agents. Then A is Lorenz dominated by A' iff*

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

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

[16] Another important approach to measuring social welfare that combines utilitarian and egalitarian views (which we are not going to discuss in this paper) is given by the so-called *Nash collective utility function*, which is defined as the *product* of individual utilities [28]. Note that using this function is only meaningful if utility functions are non-negative. Like for the utilitarian collective utility function, individual agents with high utility have a positive social effect, but inequality-reducing measures are also rated positively (e.g. $4 \cdot 4$ is better than $2 \cdot 6$).

[17] Our results pertaining to the notion of Lorenz optimality have originally been published in [16].

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 10 and 19 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 (see Definition 17) 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 (see Definition 18). 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 so 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 [28, Lemma 2.3], it is the case that any Lorenz improvement can be implemented by means of a sequence of Pareto improvements (*i.e.* cooperatively rational exchanges) and Pigou-Dalton transfers. It is important to stress that this seemingly general result does *not* apply to our negotiation framework. To see this, we consider the following example:

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

Let A be the allocation in which agent 3 owns both resources, *i.e.* $\vec{u}(A) = \langle 0, 0, 10\rangle$ and 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 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 the general result reported by Moulin is not applicable to our domain is that we cannot use Pigou-Dalton transfers to implement arbitrary utility transfers here. Any such transfer would have 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 (Definition 14). 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 20 (Simple Pareto-Pigou-Dalton deals) *A deal δ is called a simple Pareto-Pigou-Dalton deal iff $|\mathcal{A}^\delta| = 1$ and δ 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. We begin by showing that negotiation will always terminate if agents only agree on simple Pareto-Pigou-Dalton deals.

Lemma 8 (Termination) *There can be no infinite sequence of simple Pareto-Pigou-Dalton deals.*

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 are only a finite number of distinct allocations of resources, after a finite number of deals the system will have reached an allocation where no more simple Pareto-Pigou-Dalton deals are possible; that is, negotiation must terminate. □

Our sufficiency theorem for this instance of the framework follows.

Theorem 9 (Lorenz optimal outcomes) *In 0-1 scenarios, any sequence of simple Pareto-Pigou-Dalton deals will eventually result in a Lorenz optimal allocation of resources.*

Proof. By Lemma 8, the system must reach a terminal allocation A after a finite number of simple Pareto-Pigou-Dalton deals. 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 one-resource-at-a-time 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 19 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 agent i to agent 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. □

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.

7.2 Elitist Agent Societies

In Section 6.2, we have discussed the maximin-ordering induced by the egalitarian collective utility function sw_e. This ordering is actually a particular case of a class of social welfare orderings, sometimes called *k-rank dictators* [28], where a particular agent (the one corresponding to the kth element in the ordered utility vector) is chosen to be the representative of society. Amongst this class of orderings, another particularly interesting case is where the welfare of society is evaluated on the basis of the happiest agent (as opposed to the unhappiest agent, as in the case of egalitarian welfare). We call this the *elitist* approach to measuring social welfare.[18]

Definition 21 (Elitist social welfare) *The elitist social welfare $sw_{el}(A)$ of an allocation of resources A is defined as follows:*

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

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

[18]The concept of an elitist agent society has first been proposed in [19].

somewhat unethical as far as human society is concerned, we believe that it could indeed be very appropriate for certain societies of artificial agents. For some 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. 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 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. 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.

7.3 Reducing Envy amongst Agents

Our final example for an interesting approach to measuring social welfare in an agent society is the issue of *envy-freeness* [7]. 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 22 (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 $i, j \in \mathcal{A}$.*

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.

Furthermore, aiming at agreeing on an envy-free allocation of resources is not always compatible with, say, negotiating Pareto optimal outcomes. Consider the following example of two agents with identical preferences over alternative bundles of resources:

$u_1(\{\,\}) = 0$	$u_2(\{\,\}) = 0$
$u_1(\{r_1\}) = 1$	$u_2(\{r_1\}) = 1$
$u_1(\{r_2\}) = 2$	$u_2(\{r_2\}) = 2$
$u_1(\{r_1, r_2\}) = 0$	$u_2(\{r_1, r_2\}) = 0$

For this example, either one of the two allocations where one agent owns all resources and the other none would be envy-free (as no agent would prefer the other one's bundle over its own). However, such an allocation would not be Pareto optimal. On the other hand, an allocation where each agent owns a single resource would be Pareto optimal, but not envy-free (because the agent holding r_1 would rather have r_2).

We should stress that envy-freeness is defined on the sole basis of an agent's private preferences, *i.e.* 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. 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.

To be able to measure different *degrees* of enviness, we could, for example, count the number of agents that are envious for a given allocation. Another option would be to compute for each agent i that experiences any envy at all the difference between $u_i(A(i))$ and $u_i(A(j))$ for the agent j that i envies the most. Then the sum over all these differences would also provide an indication of the degree of overall envy (and thereby of social welfare). 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 any such a metric. This is a simple consequence of the fact that a deal may affect the degree of envy experienced by an agent not involved in the deal at all (because it could lead to one of the participating agents ending up with a bundle preferred by the non-concerned agent in question).

8 Conclusion

We have studied an abstract negotiation framework where members of an agent society arrange multilateral deals to exchange bundles of discrete resources, and analysed how the resulting changes in resource distribution affect society with respect to different social welfare orderings.

For scenarios where agents act rationally in the sense of never accepting a deal that would (even temporarily) decrease their level of welfare, we have seen that systems where side payments are possible can guarantee outcomes with maximal utilitarian social welfare, while systems without side payments allow, at least, for the negotiation of Pareto optimal allocations. We have then considered two examples of special domains with restricted utility functions, namely additive and 0-1 scenarios. In both cases, we have been able to prove the convergence to a socially optimal allocation of resources also for negotiation protocols that allow only for one-resource-at-a-time deals. In the case of agent societies where welfare is measured in terms of the egalitarian collective utility function, we have put forward the class of equitable deals and shown that negotiation processes where agents use equitability as an acceptability criterion will also converge towards an optimal state. Another result states 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 where social welfare is tied to the welfare of the most successful agent. And finally, 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.

Specifically, we have proved the following technical results:

- The class of *individually rational* deals[19] is sufficient to negotiate allocations with *maximal utilitarian social welfare* (Theorem 1).
- In domains with *additive utility functions*, the class of *individually rational one-resource-at-a-time* deals is sufficient to negotiate allocations with *maximal utilitarian social welfare* (Theorem 5).
- In domains with *0-1 utility functions*, the class of *cooperatively rational one-resource-at-a-time* deals is sufficient to negotiate allocations with *maximal utilitarian social welfare* (Theorem 6).
- The class of *cooperatively rational* deals is sufficient to negotiate *Pareto optimal* allocations (Theorem 3).
- The class of *equitable* deals is sufficient to negotiate allocations with *maximal egalitarian social welfare* (Theorem 7).
- In domains with *0-1 utility functions*, the class of *simple Pareto-Pigou-Dalton* deals (which are one-resource-at-a-time deals) is sufficient to negotiate *Lorenz optimal* allocations (Theorem 9).

For each of the three sufficiency results that apply to deals without structural restrictions (rather than to one-resource-at-a-time deals), we have also proved corresponding *necessity results* (Theorems 2, 4, and 8). These theorems show that any given deal (defined as a pair of allocations) that is not independently decomposable may be necessary to be able to negotiate an optimal allocation of resources (with respect to the chosen notion of social welfare), if deals are required to conform to the acceptability criterion in question. As a consequence of these results, no negotiation protocol that does not allow for the representation of deals involving any number of agents and any number of resources could ever enable agents (whose behaviour is constrained by our various acceptability criteria) to negotiate a socially optimal allocation in all cases.

A natural question that arises when considering our sufficiency results concerns the complexity of the negotiation framework. How difficult is it for agents to agree on a deal and how many deals are required before a system converges to an optimal state? The latter of these questions has recently been addressed in [17]. The paper establishes

[19]Recall that individually rational deals (unlike any other class of deals discussed in this paper) may include monetary side payments.

upper bounds on the number of deals required to reach any of the optimal allocations of resources referred to in the four sufficiency theorems for the model of rational negotiation (*i.e.* Theorem 1, 3, 5, and 6). It also discusses the different aspects of complexity involved at a more general level (such as the distinction between the *communication complexity* of the system, *i.e.* the amount of information that agents need to exchange to reach an optimal allocation, and the *computational complexity* of the reasoning tasks faced by every single agent). Dunne [13] addresses a related problem and studies the number of deals meeting certain structural requirements (in particular, one-resource-at-a-time deals) that are required to reach a given target allocation (whenever this is possible at all —recall that our necessity results show that excluding certain deal patterns will typically bar agents from reaching optimal allocations).

In earlier work, Dunne et al. [15] have studied the complexity of deciding whether one-resource-at-a-time trading with side payments is sufficient to reach a given allocation (with improved utilitarian social welfare). This problem has been shown to be NP-hard. Other complexity results concern the computational complexity of finding a socially optimal allocation of resources, independently from the concrete negotiation mechanism used. As mentioned earlier, such results are closely related to the computational complexity of the winner determination problem in combinatorial auctions [12, 33]. Recently, NP-hardness results for this optimisation problem have been derived with respect to several different ways of representing utility functions [9, 15]. Bouveret and Lang [6] also address the computational complexity of deciding whether an allocation exists that is both envy-free and Pareto optimal.

Besides presenting technical results, we have argued that a wide spectrum of social welfare orderings (rather than just those induced by the well-known utilitarian collective welfare function and the concept of Pareto optimality) can be of interest to agent-based applications. In the context of a typical electronic commerce application, where participating agents have no responsibilities towards each other, a system designer may wish to ensure Pareto optimality to guarantee that agents get maximal payoff whenever this is possible without making any of the other agents worse off. In applications where a *fair* treatment of all participants is vital (e.g. cases where the system infrastructure is jointly owned by all the agents), an egalitarian approach to measuring social welfare may be more appropriate. Many applications are in fact likely to warrant a mixture of utilitarian and egalitarian principles. Here, systems that enable Lorenz optimal agreements may turn out to be the technology of choice. Other applications, however, may require social welfare to be measured in ways not foreseen by the models typically studied in the social sciences. Our proposed notion of elitist welfare would be such an example. Elitism has little room in human society, where ethical considerations are paramount, but for a particular computing application these considerations may well be dropped or changed.

This discussion suggests an approach to multiagent systems design that we call *welfare engineering* [16]. It involves, firstly, the application-driven choice (or possibly invention) of a suitable social welfare ordering and, secondly, the design of agent behaviour profiles and negotiation mechanisms that permit (or even guarantee) socially optimal outcomes

of interactions between the agents in a system. As discussed earlier, designing agent behaviour profiles does not necessarily contradict the idea of the autonomy of an agent, because autonomy always has to be understood as being relative to the norms governing the society in which the agent operates. We should stress that, while we have been studying a *distributed* approach to multiagent resource allocation in this paper, the general idea of exploring the full range of social welfare orderings when developing agent-based applications also applies to centralised mechanisms (such as combinatorial auctions).

We hope to develop this methodology of welfare engineering further in our future work. Other possible directions of future work include the identification of further social welfare orderings and the definition of corresponding deal acceptability criteria; the continuation of the complexity-theoretic analysis of our negotiation framework; and the design of practical trading mechanisms (including both protocols and strategies) that would allow agents to agree on multilateral deals involving more than just two agents at a time.

Acknowledgements. We would like to thank Jérôme Lang and the anonymous referees of AAMAS-2003, MFI-2003 and ESAW-2003 for their valuable comments on earlier versions of (parts of) this paper [16, 18, 19]. This research has been supported by the European Commission as part of the SOCS project (IST-2001-32530).

References

[1] S. Aknine, S. Pinson, and M. F. Shakun. An extended multi-agent negotiation protocol. *Journal of Autonomous Agents and Multi-Agent Systems*, 8(1):5–45, 2004.

[2] M. Andersson and T. W. Sandholm. Contract type sequencing for reallocative negotiation. In *Proceedings of the 20th International Conference on Distributed Computing Systems (ICDCS-2000)*, pages 154–160. IEEE, 2000.

[3] K. J. Arrow. *Social Choice and Individual Values*. John Wiley and Sons, 2nd edition, 1963.

[4] K. J. Arrow, A. K. Sen, and K. Suzumura, editors. *Handbook of Social Choice and Welfare*, volume 1. North-Holland, 2002.

[5] C. Boutilier, M. Goldszmidt, and B. Sabata. Sequential auctions for the allocation of resources with complementaries. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI-1999)*, pages 527–533. Morgan Kaufmann Publishers, 1999.

[6] S. Bouveret and J. Lang. Efficiency and envy-freeness in fair division of indivisible goods: Logical representation and complexity. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-2005)*, 2005. To appear.

[7] S. J. Brams and A. D. Taylor. *Fair Division: From Cake-cutting to Dispute Resolution*. Cambridge University Press, 1996.

[8] A. Chavez, A. Moukas, and P. Maes. Challenger: A multi-agent system for distributed resource allocation. In *Proceedings of the 1st International Conference on Autonomous Agents (Agents-1997)*, pages 323–331. ACM Press, 1997.

[9] Y. Chevaleyre, U. Endriss, S. Estivie, and N. Maudet. Multiagent resource allocation with k-additive utility functions. In D. Bouyssou et al., editors, *Proceedings of the DIMACS-LAMSADE Workshop on Computer Science and Decision Theory*, volume 3 of *Annales du LAMSADE*, pages 83–100, 2004.

[10] Y. Chevaleyre, U. Endriss, J. Lang, and N. Maudet. Negotiating over small bundles of resources. In *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2005)*, 2005. To appear.

[11] Y. Chevaleyre, U. Endriss, and N. Maudet. On maximal classes of utility functions for efficient one-to-one negotiation. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-2005)*, 2005. To appear.

[12] P. Cramton, Y. Shoham, and R. Steinberg, editors. *Combinatorial Auctions*. MIT Press, 2005. To appear.

[13] P. E. Dunne. Extremal behaviour in multiagent contract negotiation. *Journal of Artificial Intelligence Research*, 23:41–78, 2005.

[14] P. E. Dunne, M. Laurence, and M. Wooldridge. Tractability results for automatic contracting. In *Proceedings of the 16th Eureopean Conference on Artificial Intelligence (ECAI-2004)*, pages 1003–1004. IOS Press, 2004.

[15] P. E. Dunne, M. Wooldridge, and M. Laurence. The complexity of contract negotiation. *Artificial Intelligence*, 164(1–2):23–46, 2005.

[16] U. Endriss and N. Maudet. Welfare engineering in multiagent systems. In A. Omicini et al., editors, *Engineering Societies in the Agents World IV*, volume 3071 of *LNAI*, pages 93–106. Springer-Verlag, 2004.

[17] U. Endriss and N. Maudet. On the communication complexity of multilateral trading: Extended report. *Journal of Autonomous Agents and Multiagent Systems*, 2005. To appear.

[18] U. Endriss, N. Maudet, F. Sadri, and F. Toni. On optimal outcomes of negotiations over resources. In J. S. Rosenschein et al., editors, *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2003)*, pages 177–184. ACM Press, 2003.

[19] U. Endriss, N. Maudet, F. Sadri, and F. Toni. Resource allocation in egalitarian agent societies. In A. Herzig et al., editors, *Secondes Journées Francophones sur les Modèles Formels d'Interaction (MFI-2003)*, pages 101–110. Cépaduès-Éditions, 2003.

[20] S. S. Fatima, M. Wooldridge, and N. R. Jennings. An agenda-based framework for multi-issues negotiation. *Artificial Intelligence*, 152(1):1–45, 2004.

[21] P. C. Fishburn. *Utility Theory for Decision Making*. John Wiley and Sons, 1970.

[22] Y. Fujishima, K. Leyton-Brown, and Y. Shoham. Taming the computational complexity of combinatorial auctions: Optimal and approximate approaches. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI-1999)*. Morgan Kaufmann Publishers, 1999.

[23] J. C. Harsanyi. Can the maximin principle serve as a basis for morality? *American Political Science Review*, 69:594–609, 1975.

[24] 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.

[25] S. Kraus. *Strategic Negotiation in Multiagent Environments*. MIT Press, 2001.

[26] 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 (IJCAI-1999)*, pages 206–211. Morgan Kaufmann Publishers, 1999.

[27] P. McBurney, S. Parsons, and M. Wooldridge. Desiderata for argumentation protocols. In C. Castelfranchi and W. L. Johnson, editors, *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 402–409. ACM Press, 2002.

[28] H. Moulin. *Axioms of Cooperative Decision Making*. Cambridge University Press, 1988.

[29] R. B. Myerson and M. A. Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29(2):265–281, 1983.

[30] J. Rawls. *A Theory of Justice*. Oxford University Press, 1971.

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

[32] M. H. Rothkopf. Bidding in simultaneous auctions with a constraint on exposure. *Operations Research*, 25(4):620–629, 1977.

[33] M. H. Rothkopf, A. Pekeč, and R. M. Harstad. Computationally manageable combinational auctions. *Management Science*, 44(8):1131–1147, 1998.

[34] F. Sadri, F. Toni, and P. Torroni. Dialogues for negotiation: Agent varieties and dialogue sequences. In *Proceedings of the 8th International Workshop on Agent Theories, Architectures, and Languages (ATAL-2001)*, pages 405–421. Springer-Verlag, 2001.

[35] T. W. Sandholm. Contract types for satisficing task allocation: I Theoretical results. In *Proceedings of the AAAI Spring Symposium: Satisficing Models*, 1998.

[36] 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.

[37] T. W. Sandholm. Algorithm for optimal winner determination in combinatorial au ctions. *Artificial Intelligence*, 135:1–54, 2002.

[38] A. K. Sen. *Collective Choice and Social Welfare*. Holden Day, 1970.

[39] R. G. Smith. The contract net protocol: High-level communication and control in a distributed problem solver. *IEEE Transactions on Computers*, C-29(12):1104–1113, 1980.

[40] M. Wooldridge. *An Introduction to MultiAgent Systems*. John Wiley and Sons, 2002.