On Optimal Outcomes of Negotiations over Resources

Ulrich Endriss, Nicolas Maudet, Fariba Sadri, and Francesca Toni
Department of Computing
Imperial College London
{ue,maudet,fs,ft}@doc.ic.ac.uk

ABSTRACT

We analyse scenarios in which self-interested agents negotiate with each other in order to agree on deals to exchange resources. We consider two variants of the framework, one where agents can use money to compensate other agents for disadvantageous deals, and one where this is not possible. In both cases, we analyse what types of deals are necessary and sufficient to guarantee an optimal outcome of negotiation. To assess whether a given allocation of resources should be considered optimal we borrow two concepts from welfare economics: maximal social welfare in the case of the framework with money and Pareto optimality in the case of the framework without money. We also show how conditions for optimal outcomes can change depending on properties of the utility functions used by agents to represent the values they ascribe to certain sets of resources.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent systems*; J.4 [**Social and Behavioral Sciences**]: Economics; K.4.4 [**Computers and Society**]: Electronic Commerce

General Terms

Theory

Keywords

Negotiation, Resource allocation, Welfare economics

1. INTRODUCTION

We analyse negotiation scenarios where self-interested agents exchange resources in order to increase their respective individual welfare. Negotiation in multiagent systems may be studied at various levels. One important line of research is concerned with the *strategies* that agents may use to determine their next move during negotiation. Good examples of work in this area are, for instance, the books by Rosenschein and Zlotkin [10] or Kraus [6]. Another important field of activity concerns the *protocols* agents may agree on when conversing with each other and, more generally, the broad area of *agent communication languages* as covered, for instance, in the collection edited by Dignum and Greaves [4].

Here, we are not concerned with any specific protocols or even strategies, but rather with the patterns of resource exchanges agents could *possibly* agree on and to what extent these patterns are sufficient or necessary to guarantee optimal outcomes of negotiations. One central assumption that we *do* make with respect to the strategies that agents follow is that they are *individually rational* in the sense that they will never accept a disadvantageous deal. In the first instance, we consider the outcome of a negotiation to be optimal whenever it results in an allocation of resources with *maximal social welfare*. Here we adopt a *utilitarian* notion of social welfare, that is, we identify the welfare enjoyed by a society of agents with the sum of the values ascribed by the individual agents in that society to the resources they hold in a particular situation.

A similar framework has been studied by Sandholm in [12] and elsewhere, mostly in the context of agents negotiating in order to reallocate *tasks*. In fact, part of the present paper is concerned with transferring the results of [12] to the domain of resource allocation problems. One of the central aspects of Sandholm's framework is that agents can use *money* to compensate other agents for accepting (otherwise) disadvantageous deals. We extend this framework here to also model negotiations over resources where no money changes hands. (This is what Rosenschein and Zlotkin call negotiation without "explicit utility transfer" [10].) A money-free framework may be more appropriate for certain applications for at least two reasons. Firstly, it could be the case that agents cannot put a precise price tag on every set of resources they may or may not hold, but they could still prefer certain sets of resources over others. Such a scenario excludes exchanges involving money, but money-free barter trade could still be a beneficial option. Secondly, as will be made precise later on, the framework described in [12] makes the implicit assumption that agents have sufficiently large amounts of money available to them to be able to agree on *any* contract that seems beneficial. This assumption is not realistic for many applications.

It is therefore important to investigate negotiation scenarios *without money*. However, a downside of the money-free approach is that we cannot always guarantee that a negotiation will result in an allocation with maximal social welfare, not even in theory. Instead, we are going to study necessary and sufficient conditions for outcomes that are at least *Pareto optimal*. (An outcome is Pareto optimal iff there is no other allocation of resources that is better for some agents without being worse for any of the other agents in the society.)

Both Sandholm's results for the framework with money as well as our new results for the framework without money are encouraging and daunting at the same time. On the one hand, it is possible to show that, assuming there are no limitations on time and computational resources, agents can always negotiate an allocation of resources that is optimal (in the sense appropriate for the respective framework). On the other hand, one can also show that any type of deal (any pattern of resource exchange), however complex, may be required in order to actually reach these optimal allocations. That is, we cannot assume that there are any limits on either the number of agents or the number of resources involved in a single deal. This second type of result, on the necessity of all deal types, stems from the generality of the framework. If we do not make any specific assumptions on the nature of agents' preferences (which we are going to model by means of utility functions over sets of resources), then we cannot exclude the need for potentially very complex types of deals. We therefore discuss a number of possible restrictions on the utility functions used by agents, which—at least in some cases—lead to more favourable results.

The remainder of this paper is structured as follows. In Section 2 we are going to formally introduce the resource allocation problems studied in this paper, as well as discuss the notion of a deal type in some detail and remark on the utilitarian approach to welfare economics taken in this paper. Section 3 analyses what deal types are sufficient and necessary to guarantee an optimal outcome of a negotiation process for scenarios with money, and Section 4 does the same for scenarios without money. Results for specific utility functions are discussed in Section 5. Finally, the concluding section discusses a number of potential extensions as well as applications of our frameworks.

2. RESOURCE ALLOCATION

The basic scenario we study in this paper is that of a society inhabited by a number of agents, each of which initially holds a number of resources. Agents will usually ascribe different values to different sets of resources. They may engage into a negotiation in order to agree on the reallocation of some of the resources and thereby increase their respective individual welfare. 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.

We begin our analysis by giving formal definitions of the various parameters relevant to such negotiation scenarios. In the first instance, all definitions refer to the framework *with* money. Necessary adjustments for the framework *without* money will be discussed in Section 4.

2.1 Main Definitions

Negotiations over resources take place in a *system* $(\mathcal{A}, \mathcal{R})$, where $\mathcal{A}$ is a finite set of (at least two) *agents* and $\mathcal{R}$ is a finite set of (discrete) *resources*.

A particular *allocation* is a partitioning of the available resources $\mathcal{R}$ amongst the agents in $\mathcal{A}$.

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

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 (subsets of $\mathcal{R}$) to real numbers. This could really be *any* such function, that is, 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 utility may strongly depend on context, i.e. what other resources the agent holds at the same time. (We are going to discuss more specific classes of utility functions in Section 5.)

Definition 2 (Utility functions). *Let $\mathcal{R}$ be a set of resources and let i be an agent. The utility function u_i for agent i (over $\mathcal{R}$) is a function from $2^{\mathcal{R}}$ to $\mathbb{R}$.*

We are now in a position to define a *resource allocation problem* as consisting of a set of agents, a set of resources, a collection of utility functions, and an initial allocation.

Definition 3 (Resource allocation problems). *A resource allocation problem is a quadruple $(\mathcal{A}, \mathcal{R}, \mathcal{U}, A_0)$ where $\mathcal{A}$ is a finite set of (at least two) agents, $\mathcal{R}$ is a finite set of resources, $\mathcal{U} = \{u_i : 2^{\mathcal{R}} \to \mathbb{R} \mid i \in \mathcal{A}\}$ is a collection of utility functions over $\mathcal{R}$, and A_0 is an initial allocation of resources for the system $(\mathcal{A}, \mathcal{R})$.*

Throughout this paper, $\mathcal{A}$ will stand for the set of agents in the negotiation system under consideration and $\mathcal{R}$ for the set of available resources. Furthermore, every agent $i \in \mathcal{A}$ is assumed to be equipped with a utility function u_i ranging over subsets of $\mathcal{R}$. Any specific allocations mentioned (such as A or A') are understood to refer to this system $(\mathcal{A}, \mathcal{R})$.

Agents can negotiate *deals* to exchange resources in order to improve their respective welfare. An example would be: "I give you r_1 if you give me r_2 and r_3 to John". 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 4 (Deals). *A deal is a pair $\delta = (A, A')$ where A and A' are allocations of resources with $A \neq A'$.*

The intended interpretation of this 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, respectively.

Our agents are self-interested in the sense that they will only propose or accept deals that strictly increase their own welfare. A deal may be accompanied by a payment to compensate some of the partners for accepting a loss in utility. Rather than specifying for each pair of agents how much money the former pays to the latter, we simply say how much money each single agent either pays or receives. This will be modelled using a *payment function*.

Definition 5 (Payments). *A payment function is a function p from $\mathcal{A}$ to $\mathbb{R}$ such that $\sum_{i \in \mathcal{A}} p(i) = 0$.*

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 society does not change.

We call a deal *individually rational* iff it increases the welfare of all the agents involved in it. Recall that, given an allocation A, $A(i)$ is the set of resources held by agent i in that situation. We are going to abbreviate $u_i(A) = u_i(A(i))$ for the utility value assigned to that set by agent i.

DEFINITION 6 (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 δ iff it has to pay less than its gain in utility or iff 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. (A concrete example may be found in the next subsection.) For any given deal, there will usually be a range of possible 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.

Finally, we have to fix a notion of optimality for society as a whole. Adopting a utilitarian view, we define the *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. 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.

DEFINITION 7 (SOCIAL WELFARE). *The social welfare $sw(A)$ of an allocation of resources A is defined as follows:*

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

We say that an allocation A has *maximal social welfare* for a given system $(\mathcal{A}, \mathcal{R})$ iff there is no other allocation A' for that system with $sw(A) < sw(A')$. Maximal social welfare is our first optimality criterion (a second one will be discussed in Section 4).

2.2 Deal Types

Following Sandholm [12], we can distinguish a number of *deal types*. The simplest deals are *one-resource-at-a-time deals* where a single resource is passed on from one agent to another one. This corresponds to the 'classical' form of a contract typically found in the *contract net protocol* [13]. Clearly, if the agent parting with the resource in question does ascribe to it any value at all, then such a one-resource-at-a-time deal can only go ahead if the two agents can agree on an appropriate price.

As the following example will show, one-resource-at-a-time deals alone are not always sufficient to guarantee the optimal outcome of a negotiation. Take a system with two agents, agent 1 and agent 2, and a set of two resources $\{r_1, r_2\}$. The following table specifies for each of the two agents what resources they hold in the initial allocation A_0 as well as the values of the two utility functions u_1 and u_2 for every subset of $\{r_1, r_2\}$:

Agent 1			Agent 2		
$A_0(1)$	$=$	$\{r_1, r_2\}$	$A_0(2)$	$=$	$\{\}$
$u_1(\{\})$	$=$	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

Social welfare for the initial allocation A_0 is 7, but it could be 8, namely if agent 2 had both resources. However, 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.

Deals where one agent passes a *set* of resources on to another agent are called *cluster deals*. In the above example, passing $\{r_1, r_2\}$ from agent 1 to 2, would be individually rational if agent 2 paid agent 1 an amount of, say, 7.5.

Deals where one agent gives a single item to another agent who returns another single item are called *swap deals*. A swap deal may be beneficial for both of the agents involved, even if no money is exchanged and even if both agents ascribe some value to either resource.

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. 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 in the next section. Concrete examples are given in [12].

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. In other words, *every* deal δ (in the sense of Definition 4) is a combined deal. The ontology of deal types presented here is, of course, not exhaustive. It may, for instance, also be of interest to consider the class of deals that involve exactly two agents but any number of resources.

2.3 Remarks on Social Welfare Functions

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 agent i prefers 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 even 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 with a monetary component, i.e. where agents need to be able to agree on prices, we actually *need* access to specific numerical utility values.

A *social welfare function* is a mapping from the set of preference profiles of the agents in a society (here represented in terms of utility functions) to a preference profile of the society as a whole [2]. In this paper, we have adopted a *utilitarian* social welfare function: maximising the function sw from Definition 7 amounts to maximising the "sum of all pleasures" enjoyed by members of the society. It should be stressed, however, that this is by no means the only way of characterising social welfare.

For instance, 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 currently worst off even further. (This is Rawls' *difference principle* [9].) Therefore, an egalitarian society will (in the first instance) aim at maximising the utility of the agent that is currently worst off. That is, a suitable egalitarian social welfare function would be $sw_e(A) = min\{u_i(A) \,|\, i \in \mathcal{A}\}$. An allocation A that maximises this function is an allocation that, from an egalitarian point of view, maximises social welfare. Other more sophisticated functions would also take into account the utility of other (unhappy) agents in the society [8].

Given the distributed character of multiagent systems, particularly when having in mind a commercial setting of some sort, intuitively, the utilitarian view on social welfare seems more appropriate than the egalitarian approach. We are going to make this intuition more precise later on. In the framework discussed in this paper, the system characteristics are embodied in the definition of individual rationality: agents have no responsibilities towards other agents and their only interests lie in increasing their own individual welfare. In Lemma 1 in the next section we are going to show that, in fact, individually rational deals are precisely those deals that increase utilitarian social welfare. Therefore, if we wish to design agent societies that maximise egalitarian social welfare by means of negotiation over resources, we have to define a new criterion for the acceptability of deals to individual agents first. If we perceive the notion of multiagent systems in a broader sense and do not wish to restrict the range of potential applications to purely commercial transactions, then this may indeed be an interesting direction of research. We shall leave this issue for another occasion [5]. The technical results reported in *this* paper all pertain to the utilitarian framework set out earlier.

3. SCENARIOS WITH MONEY

The two theorems we are going to present in this section correspond to the main results obtained by Sandholm in [12] on necessary and sufficient contract types in task-oriented domains. In what follows, we are going to give a formal account of these results, using the terminology of our resource allocation problems rather than that of task contracting. Instead of showing that the results of [12] really *are* applicable to our resource allocation scenarios, we choose to prove the relevant theorems here directly.

3.1 Rational Deals and Social Welfare

The following lemma, which states that a deal is individually rational iff it increases social welfare, will simplify the proofs of the subsequent theorems.

LEMMA 1 (RATIONAL DEALS). *A deal $\delta = (A, A')$ is individually rational iff $sw(A) < sw(A')$.*

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 social welfare, the left-hand side equals $sw(A') - sw(A)$. Hence, we really get $sw(A) < sw(A')$ as claimed.

'$\Leftarrow$': Now let $sw(A) < sw(A')$. 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 $\mathbb{R}$ as follows:

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

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 $sw(A') - sw(A') > 0$. Hence, δ must indeed be an individually rational deal. $\square$

Recall our discussion of different notions of social welfare in the previous section. In this context, a possible interpretation of Lemma 1 would be that utilitarian social welfare is indeed the appropriate notion for artificial societies where agents follow individually rational strategies in the sense of Definition 6.

3.2 Sufficient Deals (with Money)

The following theorem is, essentially, equivalent to Sandholm's result regarding the *sufficiency* of the combined contract type for finding a task allocation with maximal social welfare [12, Prop. 10].

It states that, for any sequence of (combined) deals such that each deal in the sequence is individually rational and after the final deal in the sequence no more individually rational deals are possible, the allocation reached at the end of the sequence must have maximal social welfare.

THEOREM 1 (SUFFICIENT DEALS WITH MONEY). *Any sequence of combined deals (with money) that are individually rational will eventually result in a resource allocation with maximal social welfare.*

PROOF. By Lemma 1, any individually rational deal will strictly increase social welfare. Hence, as the number of distinct allocations is finite, negotiation will terminate after a finite number of deals. Now, for the sake of contradiction, let us assume that negotiation terminates with a non-optimal allocation A, that is, there exists another allocation A' with $sw(A) < sw(A')$. 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. $\square$

At first sight, this result may seem almost trivial. The notion of a combined deal is a *very* powerful one. A single deal of this type 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 combined deal is a very complex task, which may not always be viable in practice. So, one may ask, is this kind of result really relevant?

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. 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 [1], for instance, have conducted a number of experiments on the sequencing of certain contract types to reach the best possible allocations within a limited amount of time.

3.3 Necessary Deals (with Money)

The next theorem corresponds to Sandholm's main result regarding *necessary* contract types [12, Prop. 11].

It states that for any given negotiation system $(\mathcal{A}, \mathcal{R})$ and any deal δ for that system there is an instance of the resource allocation problem (that is, there are particular utility functions and a particular initial allocation) such that δ is *necessary* to be able to reach an allocation of resources with maximal social welfare. All other findings on the insufficiency of certain types of contracts reported in [12] 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 MONEY). *Let the sets of agents and resources be fixed. Then for every deal δ there is a resource allocation problem with money such that δ is necessary to reach a resource allocation with maximal social welfare.*

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 $\mathcal{U}$ 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 $sw(A') = 2 \cdot |\mathcal{A}|$, $sw(A) = sw(A') - 1$, and $sw(B) < sw(A)$ 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 social welfare. □

3.4 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 social welfare. Concretely, if A is the initial allocation and A' is the allocation with maximal 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 [12], 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.

Sandholm [12] 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. 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 [12] and the corresponding results presented here, will not apply anymore.

4. SCENARIOS WITHOUT MONEY

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 [12] and the previous section still apply for resource allocation problems *without* money.

In scenarios without money, that is, if we do not allow for compensatory payments, we cannot always guarantee an outcome with maximal social welfare. This is, for instance, the case for the following simple problem:

Agent 1	Agent 2
$A_0(1) = \{r\}$	$A_0(2) = \{\,\}$
$u_1(\{\,\}) = 0$	$u_2(\{\,\}) = 0$
$u_1(\{r\}) = 4$	$u_2(\{r\}) = 7$

Here, passing resource r from agent 1 to agent 2 would increase social welfare by an amount of 3. For the framework *with* money, agent 2 could pay agent 1, say, the amount of 5.5 and the deal would be individually rational for both of them. Without money, however, no deal is possible and negotiation must terminate with a non-optimal allocation.

4.1 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 is Pareto optimal iff there is no other allocation where social welfare is higher while no single agent has lower utility.

DEFINITION 8 (PARETO OPTIMALITY). *An allocation A is called Pareto optimal iff there is no allocation A' such that $sw(A) < sw(A')$ 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 [6]).

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 of this type *cooperatively rational*.

DEFINITION 9 (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 $u_j(A) < u_j(A')$.*

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

4.2 Sufficient Deals (without Money)

The following theorem shows that the class of cooperatively rational deals is sufficient to guarantee a Pareto optimal outcome of negotiations without money. It constitutes the analogue to Theorem 1 for the money-free framework.

THEOREM 3 (SUFFICIENT DEALS WITHOUT MONEY). *Any sequence of combined deals (without money) that are cooperatively rational will eventually result in a Pareto optimal allocation of resources.*

PROOF. Every cooperatively rational deal strictly increases social welfare.[1] Together with the fact that there are only finitely many allocations, this implies that any negotiation will 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 $sw(A) < sw(A')$ 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 $sw(A) = sw(A')$, 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. $\square$

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

Observe that the proof would not have gone through if deals were required to be strictly individually rational, 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 that they assign a utility value of 0 to, 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. So our proposed weakened notion of rationality seems indeed a very reasonable price to pay for giving up money.

4.3 Necessary Deals (without Money)

As our next result shows, also for the framework without money, each and every deal may be necessary in order to be able to guarantee an optimal outcome of a negotiation.

THEOREM 4 (NECESSARY DEALS WITHOUT MONEY). *Let the sets of agents and resources be fixed. Then for every deal δ there is a resource allocation problem without money such that δ is necessary to reach a Pareto optimal allocation of resources.*

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:

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

We get $sw(A') = 2 \cdot |\mathcal{A}|$, $sw(A) = sw(A') - 1$, and $sw(B) < sw(A)$ for all other allocations 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. $\square$

Observe that, while this proof has been very similar to the proof of Theorem 2, now 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, Parsons, and Wooldridge [7], 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. (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. SPECIFIC 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 two frameworks. In Section 2, 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 discuss some examples.

Clearly, the results on the sufficiency of the combined deal type established in Theorems 1 and 3 will still apply, whatever restrictions we may put on utility functions. Interesting new results could be either that a weaker deal type is sufficient for certain domains or that the combined deal type is 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 10 (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.

Definition 11 (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 $R \neq \{\,\}$.*

For positive utility functions, Theorem 4 does *not* hold anymore. To see this, first observe that any deal that would involve a particular 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.

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 [10]. 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 12 (Additive utility). *We call a utility function u_i additive iff we have $u_i(R) = \sum_{r\in R} u_i(\{r\})$ for every set of resources $R \subseteq \mathcal{R}$.*

The following theorem shows that for domains with additive utility functions the simple one-resource-at-a-time deal type is sufficient to guarantee outcomes with maximal social welfare in the framework with money.[2]

Theorem 5 (Additive scenarios). *If the utility functions of all agents are additive, then any sequence of one-resource-at-a-time deals (with money) that are individually rational will eventually result in a resource allocation with maximal social welfare.*

Proof. Termination is shown as for Theorem 1. We are going to show that whenever the current allocation does not have maximal social welfare, then there still is a possible one-resource-at-a-time deal that is individually rational.

In additive domains, the social welfare of a given allocation may be computed by adding up the appropriate utility values for all the single resources in $\mathcal{R}$ (the full set of resources present in the society). For any allocation A, let f_A be the function mapping each resource $r \in \mathcal{R}$ to the agent $i \in \mathcal{A}$ that holds r in situation A (that is, we have $r \in A(i)$). The social welfare for allocation A is then given by the following equation:

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

Now suppose that negotiation has terminated with allocation A. 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 $sw(A) < sw(A')$. 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. So by Lemma 1, δ must be individually rational, i.e. contrary to our earlier assumption, A cannot be a terminal allocation. Hence, A must be an allocation with maximal social welfare. $\Box$

5.3 0-1 Scenarios

An additive utility function u_i may assign either 0 or 1 to each 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 [11].

Definition 13 (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, the one-resource-at-a-time deal type is even sufficient to guarantee maximal social welfare in the framework *without* money.

Theorem 6 (0-1 scenarios). *If the utility functions of all agents are 0-1, then any sequence of one-resource-at-a-time deals (without money) that are cooperatively rational will eventually result in a resource allocation with maximal social welfare.*

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

PROOF. Termination is shown as for Theorem 3. 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 in a situation with maximal social welfare. □

6. CONCLUSIONS

We believe that the main contribution of this paper lies in the transfer of Sandholm's results on necessary and sufficient conditions for optimal outcomes in negotiation scenarios *with* money (as reported in [12]) to a framework *without* money. This involved replacing the notion of (strict) individual rationality with the notion of cooperative rationality, and the optimality criterion of maximal social welfare with the weaker concept of Pareto optimality. The technical results here are Theorems 3 and 4 on the sufficiency of combined deals and the necessity of all deals, respectively. Other contributions include our results on the sufficiency of the one-resource-at-a-time deal type for additive scenarios with money and 0-1 scenarios without money.

In the remainder of the paper, we briefly discuss a number of possible directions for future research in this area.

6.1 Utility Functions and Deal Types

At this stage, theoretical results fall into two extremes: On the one hand, we know that in the general case only the very powerful combined deals are sufficient to guarantee optimal outcomes. On the other hand, we have examples for specific scenarios where the very simple one-resource-at-a-time deal type is sufficient. Future work should aim at establishing a clearer connection between utility functions and deal types. Given a particular class of utility functions, what types of deals would be sufficient to guarantee optimal outcomes? Similarly, given a particular set of deals, what would be the largest admissible class of utility functions?

6.2 Explicit Representation of Money

It is possible to use appropriate utility functions to model money explicitly. This can be achieved by forcing the utility functions of all the agents in the system to have the same *global* value for certain sets of resources, namely those that represent money. Such a framework would have the potential of avoiding the general problem of 'unlimited money' addressed earlier, while still allowing for negotiation results with maximal social welfare whenever there are sufficient amounts of money in the system.

6.3 Welfare Engineering

The *utilitarian* interpretation of social welfare is often taken for granted in the multiagent systems literature (e.g. [7, 12]). Lemma 1 shows that this *is* in fact the right notion for the kind of scenarios we have considered here. However, for different types of scenarios it may be of interest to investigate different types of measures studied in the literature on welfare economics. We have already mentioned the case of *egalitarian* welfare functions. Another option would be, for example, to search for conditions that guarantee allocations of resources that are *envy-free* [3]. In our framework, we would call an allocation A envy-free iff we have $u_i(A(i)) \geq u_i(A(j))$ for all $i, j \in \mathcal{A}$, that is, iff no agent would rather have the set of resources allocated to one of the other agents in the society.

For cooperatively rational agents, we cannot hope for either maximal egalitarian social welfare or envy-free outcomes in the general case. For applications where such outcomes would be desirable, alternative criteria for the acceptability of deals to individual agents need to be developed.

6.4 Protocol Design

Our framework may also provide practical guidelines for the design of concrete negotiation protocols. For example, if the application domain in question can be modelled in terms of additive or even 0-1 utility functions, then Theorems 5 and 6 tell us that it would be inappropriate to allow for dialogue moves for proposing, say, swap deals. At the other end of the spectrum, for domains where we cannot make any strong assumptions on the nature of utility functions, Theorems 2 and 4 show that, ideally, a good protocol should enable agents to agree on *any* kind of deal.

7. ACKNOWLEDGEMENTS

We would like to thank Jérôme Lang for his valuable comments on an earlier version of this paper. This research has been funded by the European Union as part of the SOCS project (Societies Of ComputeeS), IST-2001-32530.

8. REFERENCES

[1] M. Andersson and T. W. Sandholm. Contract Type Sequencing for Reallocative Negotiation. In *ICDCS-2000*, pages 154–160. IEEE, 2000.

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

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

[4] F. Dignum and M. Greaves, editors. *Issues in Agent Communication.* Springer-Verlag, 2000.

[5] U. Endriss, N. Maudet, F. Sadri, and F. Toni. Resource Allocation in Egalitarian Agent Societies. In *Secondes Journées Francophones sur les Modèles Formels d'Interaction (MFI-2003)*, 2003. To appear.

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

[7] P. McBurney, S. Parsons, and M. Wooldridge. Desiderata for Argumentation Protocols. In *AAMAS-2002*, pages 402–409. ACM Press, 2002.

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

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

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

[11] F. Sadri, F. Toni, and P. Torroni. Dialogues for Negotiation: Agent Varieties and Dialogue Sequences. In *ATAL-2001.* Springer-Verlag, 2001.

[12] T. W. Sandholm. Contract Types for Satisficing Task Allocation: I Theoretical Results. In *AAAI Spring Symposium: Satisficing Models*, 1998.

[13] 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.