Admissible Agreements among Goal-directed Agents

Guido Boella
Dipartimento di Informatica
Università di Torino - Italy
guido@di.unito.it

Luigi Sauro
Dipartimento di Informatica
Università di Torino - Italy
sauro@di.unito.it

Leendert van der Torre
CWI and TU Delft - Amsterdam
The Netherlands
torre@cwi.nl

Abstract

We study admissible coalitions in goal-directed multiagent systems. We define a qualitative criterion of admissibility in which a coalition has itself all the necessary information to check admissibility. We show also that, under some assumptions on preference relations of the agents, this admissibility criterion can be used to reduce the search space in a game theoretical approach.

1. Introduction

It is desirable that artificial agents can help each other when they cannot achieve their goals, or when they profit from cooperating. Cooperative game theory [6, 1] focuses on collusive behaviors, supported by enforced agreements, that involve the formation of coalitions. In collusive behaviors, agents have the possibility to decide how to coordinate themselves without imposition by anyone. An agreement is enforced if the involved parties cannot deviate from the agreement, once they decide to enter it.

Sandholm et al. [7] distinguish two phases to establish which coalitions can be formed. In the first phase a structure describing all the possible coalition configurations is defined. In the second phase a quantitative method is used to prune those configurations that cannot occur, under the assumption of self-interested agents. It is reasonable to use game theoretical criteria as pruning method, but, unfortunately, it has been shown that several solution criteria defined in cooperative game theory are computationally intractable [8, 9]. Sandholm et al. [7] therefore define some approximation algorithms to search the space of possible coalitions.

In [3] we have introduced the do-ut-des property as a qualitative criterion of admissibility for coalition formation in goal-directed multiagent systems. This criterion has been defined by means of a balance between the set of goals of an agent achieved in a coalition and the tasks it is burdened to perform if it agree to enter the coalition. In this way the formalization of do-ut-des property is based on a not compelling mixture of two aspects, the goals achieved and the tasks executed to achieve them, that should correspond to two different level of abstractions. In this paper we consider an alternative approach that removes this weakness. Moreover if in [3] the do-ut-des property was defined starting from a multiagent system representation that directly describes the achievement power of sets of agents. Here we face the problem to define a typology of these achievement powers starting from the capabilities of the single agents. This notion of power presents an analogy with the one we developed in [2]. The main difference is that in [2] we defined a notion of power requiring to a set A of agents to be minimal with respect the achievement of a goal g, in this way we formalized the fact that all the agents in A have to be necessary for the satisfaction of g. In this work the notion of power requires a minimalization of the tasks assigned to A and not a minimalization on the set itself. Therefore in this case we formalize the notion of relevance of the tasks executed by A with respect to the achievement of g.

In this work we face the problem to cut off from the space of all possible coalitions the ones that cannot occur by using a qualitative admissibility criterion to be applied before a quantitative game theoretical criterion. The methodology used is on one hand to abstract from the specific preference relations of the agents by focusing on the goals representing the advantages an agent gains entering a coalition. On the other hand we do not simply represent a coalition by means of the goals it can attain, as done in Dunne et al. [9], we represent a coalition as an agreement describing for each set of agents the goals it is burdened to achieve.

In Section 2 we define a multiagent system and provide the notion of goals assurable by a set of agents. In Section 3 we define the cooperative game relative to a multiagent system. In Section 4 the do-ut-des property is defined. Section 5 shows under which conditions the do-ut-des coalitions can be employed as a qualitative reasoning on profitability of coalitions. Section 6 shows the relation between solution concept of core and the do-ut-des property.

2. The Multiagent System

In order to describe the coalitions that can be formed in a multiagent system we define in this section a framework to formalize agents with their individual characteristics, as the goals they have or the tasks they can embark on. Tasks are distinguished from strategies. A strategy is a complete schedule of actions an agent can adopt in a game, so two different strategies cannot be adopted at the same time. Instead, tasks are partial schedules and an agent can execute more tasks at the same time if these tasks are compatible with each other. As in Shehory and Kraus [8], an agent can possibly joint two different coalitions if the required tasks are compatible. We model compatibility among tasks as a relation *comp*, given an agent *a* and a non-empty set of tasks T it can execute. $(a, T) \in comp$ means that *a* can execute all the tasks in T at the same time, therefore $(a, \{t\}) \in comp$ means that the agent *a* can execute the task t. Since we consider goal-directed agents, we describe the effect of a task only with respect to the goals it achieves. Formally we have the following definition of a multiagent system:

Definition 1 (Multiagent System) *A multiagent system, MaS, is a tuple $\langle Ag, Gl, Tk, gl, comp, ach\rangle$ with the following elements. Ag is a set of agents. Gl is a set of goals. Tk is a set of tasks. $comp \subseteq Ag \times 2^{Tk}$ is a relation describing the sets of tasks an agent can execute at the same time. $gl : Ag \rightarrow 2^{Gl}$ is a function from agents to the sets of their own goals such that $\bigcup_{a \in Ag} gl(a) = Gl$. Finally, $ach \subseteq 2^{Tk} \times Gl$ describes the results of executing a set of tasks.*

Example: As example we consider a simple version of the game described in Grosz *et al.* [4]. There are four players and each of them has two colored chips and two colored boxes, both the chips and the boxes can be colored as follows: red r, green g or blue b. Every player has only one goal, to fill each of his boxes with a chip of the same color of the box. The players have the possibility to exchange their chips and, for simplicity, we assume that the game is made only of one turn.

The initial configuration is:

players	boxes	chips
p_1	$r\ r$	$r\ g$
p_2	$r\ g$	$b\ b$
p_3	$b\ g$	$r\ b$
p_4	$b\ b$	$r\ g$

This table shows how, in the initial configuration, there is no a player that is self sufficient to see to its own goal, so they have to exchange chips with each other in order to satisfy their goals.

We formalize this multiagent system by the following *MaS*. *Ag* is the set of the four player $\{p_1, p_2, p_3, p_4\}$, the set of the goals *Gl* is $\{rr, rg, bg, bb\}$, according to the boxes owned by the agents. The set of tasks *Tk* is the set of pair (x, p_j), where x can be r, g or b and j denotes one of the four players: executing this task means to provide a chip colored of x to the player p_j.

$comp$ describes compatible tasks, a set of distinct tasks $\{(x^1, p_j^1), \ldots, (x^n, p_j^n)\}$ is compatible for a player p_i, $(p_i, \{(x^1, p_j^1), \ldots, (x^n, p_j^n)\}) \in comp$, if and only if p_i has n chips, the first of color x^1, the second of color x^2 and so on. Since every agent has only two chips, any set of tasks with more then two tasks is not compatible for a player. *gl* associates, to any player, his relative goal, formally $gl = \{(p_1, \{rr\}), (p_2, \{rg\}), (p_3, \{bg\}), (p_4, \{bb\})\}$. *ach* describes the goals achieved by sets of tasks; a goal is achieved when the two boxes of a player are filled with the right chip, so, for example, the set of tasks $\{(g, p_3), (b, p_3)\}$ achieves the goal bg of p_3. ■

Since agents act together, and possibly collaborate in the achievement of goals, we introduce tasks distributions, i.e., the tasks that can be executed by sets of agents. A tasks distribution of a set of agents A associates to each agent $a \in A$ a nonempty set of tasks T such that $(a, T) \in comp$ and to all the other agents the empty set.

Definition 2 *Given a set of agents $A \subseteq Ag$, a tasks distribution of A is a function $\tau_A : Ag \rightarrow 2^{Tk}$ such that $\tau_A \subseteq comp$ and $\tau_A(a) \neq \emptyset$ iff $a \in A$.*

We denote with $\bigsqcup(\tau_A)$ the union for $a \in A$ of $\tau_A(a)$.

Now we define a notion of compatibility among tasks distributions.

Definition 3 *We denote with $\tau_{A_1} \sqcup \tau_{A_2}$ a function that, for each agent $a \in Ag$, associates $\tau_{A_1}(a) \cup \tau_{A_2}(a)$. Moreover we say that $\tau_{A'} \sqsubseteq \tau_A$ if there exists a $\tau_{A''}$ such that $\tau_A = \tau_{A'} \sqcup \tau_{A''}$, as usual, $\tau_{A'} \sqsubset \tau_A$ is and only if $\tau_{A'} \sqsubseteq \tau_A$ and $\tau_{A'} \not\sqsupseteq \tau_A$. The tasks distributions $\tau_{A_1}, \ldots, \tau_{A_n}$ is compatible iff $\tau_{A_1} \sqcup \ldots \sqcup \tau_{A_n} \subseteq comp$.*

Assume that there exists a tasks distribution τ_A of the set of agents A such that it achieves a goal g no matter which tasks distribution the other agents execute. If this is the case, then A can assure, by means of τ_A, the achievement of g.

We also define a minimality condition on assurable goals that is satisfied when a set of the agents A can assure a goal g by means of τ_A and all the tasks in τ_A play a role in the achievement of g, that is, there is no a subset $A' \subseteq A$ and a tasks distribution τ'_A such that A' can assure g by means of τ'_A and $\tau'_A \sqsubset \tau_A$.

Definition 4 (Assurable goals) *Given a task distribution τ_A and $g \in Gl$, we say that A can assure g by means of τ_A, $(\tau_A, g) \in \alpha$-eff, iff*

$$\forall \overline{A} \subseteq [Ag \setminus A]\ \forall \tau_{\overline{A}}\ \left(\bigsqcup(\tau_{\overline{A}} \sqcup \tau_A), g\right) \in ach$$

We say that (τ_A, g) can minimally assure g by means of τ_A, $(\tau_A, g) \in$ min-α-eff, iff $(\tau_A, g) \in \alpha$-eff and there does not exist an $A' \subseteq A$ and a $\tau_{A'} \sqsubset \tau_A$ such that $(\tau_{A'}, g) \in \alpha$-eff.

It can be seen that our framework satisfies the property of super-additivity, i.e., given two disjoint sets of agents A_1 and A_2, if A_1 can assure g_1 by means of τ_{A_1} and A_2 can assure g_2 by means of τ_{A_2}, then $A_1 \cup A_2$ can assure for both g_1 and g_2 by means of $\tau_{A_1} \sqcup \tau_{A_2}$.

Example: Defined the multiagent system relative to the boxes-chips game, tasks distributions as well as the sets α-*eff* and *min*-α-*eff* can be derived following the relative definitions. Since it is not possible here to describe them extensively, we provide only some examples and considerations. In the following, for readability reasons, we denote an element $(p_i, \{(x^1, p_j^1), \dots, (x^n, p_j^n)\})$ of a tasks assignment with

$$p_i \xrightarrow{x^1} p_j^1, \dots, p_i \xrightarrow{x^n} p_j^n$$

An example of tasks distribution is $\tau_{\{p_1,p_3\}} = \{p_1 \xrightarrow{g} p_3, p_3 \xrightarrow{b} p_3\}$, indeed both $p_1 \xrightarrow{g} p_3$ and $p_3 \xrightarrow{b} p_3$ are in *comp*. When a player obtains, from some players, the chips needed to satisfy his goal, the other players cannot obstruct him. α-*eff* reflects the relation *ach*, i.e., if $(\bigsqcup(\tau_A), g) \in$ *ach*, then $(\tau_A, g) \in \alpha$-*eff*.

This in general is not true, for example, when two persons with the same strength try respectively to open and to close a door. Their tasks, pull the door and push the door, would be successful if performed separately, but together they obstruct with each other.

Nevertheless in our game *min*-α-*eff* is strictly contained in α-*eff*, in fact, given the tasks distribution $\tau_{\{p_1,p_2,p_3\}} = \{p_1 \xrightarrow{g} p_3, p_3 \xrightarrow{b} p_3, p_2 \xrightarrow{b} p_4\}$, since p_3 obtains the needed chips b and g, his goal is satisfied and, hence, $(\tau, bg) \in \alpha$-*eff*. Nevertheless also the restriction of τ to $\{p_1, p_3\}$ satisfies the goal bg, so $p_2 \xrightarrow{b} p_4$ is not useful for the achievement of bg. ∎

3. Cooperative games

Definition 4 establishes the potentials for cooperation in a multiagent system. To see which of them can actually occur, we consider cooperative games defined starting from goal-directed agents. A coalition describes the possibility for a set of agents to help each other in the satisfaction of their goals. In this way it is not associated to a coalition a single value that can be divided among the members of the coalition as much as they like, but a set of *consequences* that describes the goals achieved and the burdens sustained by each agent. This typology of problems are faced by cooperative games without transferable payoffs whose definition we borrows from Osborne et al. [6].

Definition 5 (Cooperative games) *A cooperative game without transferable payoffs (NTU) is a tuple*

$$\langle Ag, Cs, att, \precsim_1, \dots, \precsim_n \rangle$$

with the following elements. Ag is a set of agents. Cs is a set of consequences. att : $2^{Ag} \setminus \emptyset \rightarrow 2^{Cs}$ maps each nonempty set A of agents in the set of consequences that are attainable by A. For each agent a_i, $\precsim_i \subseteq Cs \times Cs$ is a preference relation over Cs, i.e. it is complete, reflexive and transitive. Given $c_1, c_2 \in Cs$, with $c_1 \precsim_i c_2$ we denote that the agent a_i prefers c_2 at least as c_1.

Following Sandholm et al. [7], we consider the notion of core as solution criteria of a cooperative game. The notion of core is based on a dominance relation over the set of possible consequences attainable by the gran coalition *Ag*. $c \in att(Ag)$ is dominated if there exists a set of agents such that it can attain a consequence that is strictly preferred by all its members to c.

Definition 6 (Core) *Let $\langle Ag, Cs, att, \precsim_1, \dots, \precsim_n \rangle$ be a cooperative game without transferable payoffs, the core is the set of consequences $c \in att(Ag)$ such that there does not exist a group of agents $A \subseteq Ag$ and a consequence $c' \in att(A)$ such that for all $a_i \in A$, $c \prec_i c'$.*

We define a cooperative game without transferable payoffs relative to a multiagent system *MaS*. First of all, we consider the set of possible consequences that a set of agents can attain if they form a coalition.

The set *min*-α-*eff* defined in Definition 4, describes the set of all potentials for cooperation. If a set of agents A decides to form a coalition and there exists a $(\tau_{A'}, g) \in$ *min*-α-*eff* such that $A' \subseteq A$, then the goal g is in the achievement potential of A, i.e. can be a part of a consequence attainable by A.

Since not all of the elements of *min*-α-*eff* can be carried out at the same time, this means that the consequences attainable by a set of agents A derives from all the subsets $\{(\tau_{A_1}, g_1), \dots, (\tau_{A_n}, g_n)\} \subseteq$ *min*-α-*eff* in which the tasks are compatible, see Definition 3, and for all $1 \leq i \leq n$, $A_i \subseteq A$. Following Kraus et al. [5], we consider a consequence of the formation of a coalition A the whole set $c = \{(\tau_{A_1}, g_1), \dots, (\tau_{A_n}, g_n)\}$, and not simply the set of all goals achieved in c, as done in Dunne et al. [9]. This enables us to define in the next section the do-ut-des property.

In the following, instead of saying that c is a consequence attainable by the set A of agents if they form a coalition, we simply say that c is a possible coalition of A.

Definition 7 (Possible coalitions) *Given a multiagent system $MaS = \langle Ag, Gl, Tk, gl, comp, ach \rangle$ and a nonempty set of agents $A \subseteq Ag$, we say that $c = \{(\tau_{A_1}, g_1), \dots, (\tau_{A_n}, g_n)\} \subseteq$ min-α-eff is a possible coalition of A iff $\tau_{A_1}, \dots, \tau_{A_n}$ are compatible and for all $1 \leq i \leq n$, $A_i \subseteq A$.*

We say that c is a possible coalition if there exists a set of agent A such that c is a possible coalition of A, moreover we denote with Gl_c the set of all the goals achieved in c:

$$Gl_c = \{g \in Gl \mid \exists A \subseteq Ag \text{ s.t. } (\tau_A, g) \in c\}$$

We define a cooperative game relative to a multiagent system as follows.

Definition 8 (Cooperative game of a *MaS*) *Given a multiagent system MaS, a cooperative game relative to MaS, $NTU[MaS]$, is a tuple $\langle Ag, Cs, att, \precsim_1, \dots, \precsim_n \rangle$ with the following elements. Ag is the set of agents of MaS. Cs is the set of all possible coalitions of MaS. att : $2^{Ag} \setminus \emptyset \to 2^{Cs}$ maps for each nonempty set of agents A the set of all possible coalitions of A. For each agent $a_i \in Ag$, $\precsim_i \subseteq Cs \times Cs$ is a preference relation over Cs.*

Example: Given the *MaS* relative to the boxes-chips game, we consider a cooperative game $NTU[MaS] = \langle Ag, Cs, att, \precsim_1, \dots, \precsim_n \rangle$ as follows. A player p_i first of all prefers a coalition that satisfies his goal with respect to a coalition that does not satisfies it. In the case two coalitions both satisfy his goal or both do not satisfy it, the p_i prefers the one that minimize the number of chips he gives to the other agents.

More formally, given a possible coalition $c = \{(\tau_{A_1}, g_1), \dots, (\tau_{A_n}, g_n)\}$ and an agent p_i, we first denote with $\tau(p_i, c)$ the set of all tasks, (x, p_j), assigned to p_i in c such that $p_i \neq p_j$, i.e. $\tau(p_i, c)$ is the set of tasks (x, p_j) such that $p_j \neq p_i$ and there exists $(\tau_{A_i}, g_i) \in c$ with $(x, p_j) \in \tau_{A_i}(p_i)$. The preference relation $\precsim_i$ of each player p_i such that $c_1 \precsim_i c_2$ if and only one of the either (1) $(gl(a_i) \cap Gl_{c_1}) \subset (gl(a_i) \cap Gl_{c_2})$ or (2) $(gl(a_i) \cap Gl_{c_1}) = (gl(a_i) \cap Gl_{c_2})$ and $|\tau(a_i, c_1)| \leq |\tau(a_i, c_2)|$. ■

4. Do-ut-des Coalitions

In this section we provide a criterion of admissibility for coalitions, the *do-ut-des* property. This property is based on a dominance relation between a possible coalition c and a one of its sub-coalitions $c' \subset c$. So, against the usual quantitative notions of profitability as the core, in which the coalition c has to be compared with all the other possible coalitions to see if it belongs to the core, a coalition c contains all the information to check if it satisfies the do-ut-des property. First we define a preference relation between two coalitions by means of the functions *adv* and *obl* that describe for each agent its advantages and burdens in c.

Definition 9 (adv function) *Given a coalition c, adv$[c]$: $Ag \to 2^{Gl}$ maps, for each agent $a \in Ag$, the set of goals achieved in c that are goals of a: adv$[c](a) = Gl_c \cap gl(a)$.*

Instead, the function *obl* is defined as the set of goals in which achievement an agent is involved:

Definition 10 (obl function) *Given a coalition c, obl$[c]$: $Ag \to 2^{Gl}$ is such that for all $a \in Ag$*

$$obl[c](a) = \{g \in Gl_c \mid \exists (\tau_A, g) \in c \text{ s.t. } a \in A\}$$

Now we are able to define, for each agent a_i, a qualitative preference relation $\leq_i$ between two coalitions. In contrast with the quantitative preference relations introduced in the Definition 5, our preference relation in not complete, so it could be the case that, given an agent a_i and two coalitions c_1 and c_2, $c_1 \not\leq_i c_2$ and $c_2 \not\leq_i c_1$.

Definition 11 (Qualitative preference relation) *Let c_1 and c_2 be two coalitions. We say that the agent a_i qualitatively prefers c_1 to c_2, $c_2 \leq_i c_1$, iff adv$[c_2](a_i) \subseteq$ adv$[c_1](a_i)$ and obl$[c_1](a_i) \subseteq$ obl$[c_2](a_i)$.*

As usual we say that a_i strictly qualitatively prefers c_1 to c_2, $c_2 <_i c_1$, if $c_2 \leq_i c_1$ and $c_1 \not\leq_i c_2$.

A possible coalition c satisfies the do-ut-des property, or it is a do-ut-des coalition, if and only if there does not exist an agent involved in c which strictly qualitatively prefers a coalition $c' \subset c$ and all the agents involved in c' qualitatively prefers c' at least as c. In this case *a* does not agree to c, considering that c' is better for him and that if all the agents involved in c' would have agreed to c, then they agree also to c'.

Definition 12 (Do-ut-des property) *Given a coalition c, we denote with Dom(c) the set of agents involved in c, i.e. Dom$(c) = \{a \in Ag \mid \exists (\tau_A, g) \in c \text{ s.t. } a \in A\}$. We say that $c' \subset c$ do-ut-des dominates c iff the following conditions hold:*

1. $\exists a_i \in Dom(c)\ \ c <_i c'$
2. $\forall a_j \in Dom(c')\ \ c \leq_j c'$

A coalition c is do-ut-des iff there does not exist a coalition $c' \subset c$ that do-ut-des dominates c.

Example: In our boxes-chips game we provide an example of a do-ut-des coalition and an example of a coalition that does not satisfy the do-ut-des property. As first example consider c described by:

τ_A	g
$p_1 \xrightarrow{r} p_1, p_3 \xrightarrow{r} p_1$	rr
$p_1 \xrightarrow{g} p_3, p_3 \xrightarrow{b} p_3$	bg

where each row of the table is an element of c. In this coalition every player obtains the satisfaction of his own goal, so no one of them qualitatively prefers the empty coalition to c. Moreover since both p_1 and p_3 are necessary for the fulfillment of rr and bg, in each one of the coalitions obtained considering only one row of the table there is a player that is disadvantaged. So c satisfies the *do-ut-des* property.

As second example consider the coalition c as shown in the following table:

τ_A	g
$p_1 \xrightarrow{r} p_1, p_3 \xrightarrow{r} p_1$	rr
$p_2 \xrightarrow{b} p_4, p_2 \xrightarrow{b} p_4$	$b\,b$

It can be seen that it does not satisfy the do-ut-des property. In fact the advantages of p_2 and p_3 in this coalition are empty, nevertheless they are involved respectively in the achievement of rr and bb. So both of them qualitatively prefer to c the empty coalition, in which they do not obtain anything but also do not provide anything. ■

5. Do-ut-des compatible NTU

The Definition 8 does not provide any restriction to the preference relations $\precsim_i$ of the agents. Nevertheless the notion of do-ut-des coalitions can be related to the quantitative notion of core only in the case the $\precsim_i$ are *compatible* with the Definition 11, i.e. for any agent a_i, $\leq_i$ implies $\precsim_i$ and $<_i$ implies $\prec_i$.

Definition 13 (Do-ut-des compatible NTU) *Let MaS a multiagent system, a cooperative game $NTU[MaS] = \langle Ag, Cs, att, \precsim_1, \ldots, \precsim_n\rangle$ is do-ut-des compatible iff for all $1 \leq i \leq n$ and for all possible coalitions $c_1, c_2 \in Cs$ (1) if $c_1 \leq_i c_2$, then $c_1 \precsim_i c_2$ and (2) if $c_1 <_i c_2$, then $c_1 \prec_i c_2$.*

Often the preference relation of an agent with respect to a set of consequences is represented by an utility function [1]. An utility function is real valued function over the set of consequences *Cs* that represents, for each consequence, the profitability of that consequence. We say that an utility function $utl : Cs \rightarrow \Re$ represents a preference relation $\precsim$ just in the case that for all $c_1, c_2 \in Cs$, $c_1 \precsim c_2$ if and only if $utl(c_1) \leq utl(c_2)$.

A way to effectively *calculate* the utility of a consequence is by means of a cost-benefit analysis. The idea underling the cost-benefit analysis is that any consequence represents a state of affairs that involves some advantages, but also has some contraindications. In our case a consequence is a possible coalition c and, given an agent a_i the advantages of a_i with respect to c derives from the of its goals that are achieved if c is formed, i.e. $adv[c](a_i)$; the contraindications of a_i with respect to c depends on many factors as, for example, the tasks it has to execute if c is formed as in [3], but also the costs implicit to the coalition formation process or to the fact that entering the coalition it binds itself to satisfy some goals. For all these reasons we consider that the costs relative to a coalition depends to the coalition as whole.

The utility of a consequence c, for the agent a_i, is represented by a function blc_i that balance an estimation of the advantages (the benefits) and an estimation of the contraindications (the costs) of c. We assume that the benefits and the costs of consequences are themselves real value functions, respectively bnf_i and $cost_i$. Therefore, given a possible coalition c, the utility of the agent a_i relative to c, is given by the formula $utl_i(c) = blc_i(bnf_i(adv[c](a_i), cost_i(c))$.

Now we show some sufficient conditions utl_i has to satisfy in order to be do-ut-des compatible. The first condition says that blc_i is a function strictly increasing in the first argument and strictly decreasing in the second argument. The second condition says that the more are the goals an agent has to execute the more are the relative costs. The third condition says that adding a desired goal to set of desired goals involves a not null increasing of the benefit function.

Definition 14 (Do-ut-des compatible utility functions) *An utility function of an agent a_i, $utl_i(c) = blc_i(bnf_i(adv[c](a_i), cost_i(obl[c](a_i)))$, is said to be do-ut-des compatible iff*

1. *for all fixed $\hat{x} \in \Re$ and $\hat{y} \in \Re$, $blc_i(x, \hat{y})$ and $blc_i(\hat{x}, y)$ are respectively strictly increasing and strictly decreasing.*
2. *for all possible coalitions $c_1, c_2 \in Cs$, $obl[c_1](a_i) \subseteq obl[c_2](a_i)$ iff $cost_i(c_1) \leq cost_i(c_2)$.*
3. *for all $c_1, c_2 \in Cs$, $adv[c_1](a_i) \subseteq adv[c_2](a_i)$ iff $bnf_i(adv[c_1](a_i)) < bnf_i(adv[c_2](a_i))$.*

The following theorem shows that if utl_i satisfies the Definition 14, then it is effectively compatible with the qualitative preference relation $\leq_i$.

Theorem 1 *If the utility function utl_i is do-ut-des compatible, then for all possible coalitions c and c', (1) if $c \leq_i c'$, then $utl_i(c) \leq utl_i(c')$, (2) $c <_i c'$, then $utl_i(c) < utl_i(c')$.*

proof: Assume that $c \leq_i c'$, then

$$adv[c](a_i) \subseteq adv[c'](a_i) \text{ and } obl[c'](a_i) \subseteq obl[c](a_i) \quad (1)$$

By hypothesis utl_i is do-ut-des compatible, so $cost_i(c') \leq cost_i(c)$ and $bnf_i(adv[c](a_i)) \leq bnf_i(adv[c'](a_i))$. But, since blc_i is a function strictly increasing in the first argument and strictly decreasing in the second argument, we have

$$\begin{aligned} &blc_i(bnf_i(adv[c](a_i)), cost_i(c)) \leq \\ &blc_i(bnf_i(adv[c'](a_i)), cost_i(c)) \leq \\ &blc_i(bnf_i(adv[c'](a_i)), cost_i(c')) \end{aligned} \quad (2)$$

and hence $utl_i(c) \leq utl_i(c')$.

If $c <_i c'$, then the previous inequality holds and also either $obl[c'](a_i) \subset obl[c](a_i)$, or $adv[c](a_i) \subset adv[c'](a_i)$. In the first case $cost_i(obl[c'](a_i)) < cost_i(obl[c](a_i))$, so

$$\begin{aligned} &blc_i(bnf_i(adv[c'](a_i)), cost_i(c)) < \\ &blc_i(bnf_i(adv[c'](a_i)), cost_i(c')) \end{aligned} \quad (3)$$

In the second case $bnf_i(adv[c](a_i)) < bnf_i(adv[c'](a_i))$, so

$$blc_i(bnf_i(adv[c](a_i)), cost_i(c)) < blc_i(bnf_i(adv[c'](a_i)), cost_i(c)) \quad (4)$$

In both the cases $utl_i(c) < utl_i(c')$. □

Example: In the boxes-chips game the preference relation $\precsim_i$ defined in Section 3 can be represented by the utility functions

$$utl_i(c) = \frac{1 + bnf_i(adv[c](p_i))}{1 + cost_i(c)} \quad (5)$$

where $bnf_i(adv[c](p_i))$ is equal to 1000 if $adv[c](p_i))$ is not empty, 0 otherwise. $cost(c)$ is equal to the sum of costs the tasks (x, p_j) that p_i has to perform under c, where the cost of a task (x, p_j) is 50 if $p_j = p_i$, 200 otherwise.

The characteristic function defined by the Formula 5 in the Section 3 is do-ut-des admissible. The first condition is evident since in the Formula 5 utl_i is directly proportional to $bnf_i(adv[c](p_i))$ and in inverse proportion to $cost(c)$. The second condition is also true since it is not possible for a player, providing one chip, to satisfy at the same time the goals of two different agents, so different goals require different tasks for the players. Concerning the third one, each player a has only one goal so $adv[c](p_i)$ can be either empty or equal to $gl(a)$; since the associated reward are 0, if it is empty, and 1, if it is $gl(a)$, the condition is satisfied. ■

6. Do-ut-des property and the core

In order to relate the qualitative do-ut-des property to the quantitative notion of core we consider a quantitative version of the *do-ut-des* property, we call it q-do-ut-des property. We use this property to relate the *do-ut-des* property to the notion of core, so we want that this property has a precise relationship with both of them.

Definition 15 (Q-do-ut-des) *Let* $NTU[MaS] = \langle Ag, Cs, att, \precsim_1, \ldots, \precsim_n\rangle$ *be the cooperative game of the multiagent system MaS. A possible coalition* c *is q-do-ut-des iff for all* $c' \neq c$ *there does not exist an agent* a_i *such that* $c \prec_i c'$ *and for all* $a_j \in c'$, $c \precsim_j c'$.

The following theorem shows that if a possible coalition c is q-do-ut-des, then it is in the core.

Theorem 2 *Let* $NTU[MaS] = \langle Ag, Cs, att, \precsim_1, \ldots, \precsim_n\rangle$ *be a cooperative game of the multiagent system MaS. If a possible coalition* c *is q-do-ut-des, then it is in the core.*

proof: If c is not in the core, then there exists a possible coalition c' such that for all the agents $a_i \in c'$, $c \prec_i c'$. Therefore there exists an agent that strictly prefers c' to c. Moreover, since if $c \prec_i c'$ holds, then also $c \precsim_i c'$ is true for all $a_i \in c'$. But this means that c' q-do-ut-des dominates c, and hence that c is not q-do-ut-des. □

Moreover, we show that, for do-ut-des compatible NTU, if a possible coalition c is q-do-ut-des, then it is do-ut-des.

Theorem 3 *Let* $NTU[MaS] = \langle Ag, Cs, att, \precsim_1, \ldots, \precsim_n\rangle$ *be a do-ut-des compatible cooperative game of a multiagent system MaS, if a possible coalition* c *is q-do-ut-des, then it is do-ut-des.*

proof: Assume that c is q-do-ut-des, but not do-ut-des. By definition this means that there exists a possible coalition $c' \subset c$ and an agent $a_i \in c$ such that $c <_i c'$ and for all $a_j \in c'$, $c \leq_j c'$. Since $NTU[MaS]$ is do-ut-des compatible, then it is also the case that $c \prec_i c'$ and for all $a_j \in c'$, $c \precsim_j c'$. But this means that c' q-do-ut-des dominates c against the hypothesis. □

The problem is to find q-do-ut-des coalitions of $NTU[PS]$. Since the set of q-do-ut-des coalitions is contained in the core, if we find a q-do-ut-des coalition, we are sure that it is also in the core.

The definition of q-do-ut-des coalitions requires to compare a possible coalition c with all the others. Due to the Theorem 3, we restrict the set of coalitions on which to check for q-do-ut-des coalitions to the set of do-ut-des coalitions. However, for each do-ut-des coalitions, we still have to compare it with of all the other possible coalitions.

Fortunately, the following theorem shows that if c is not do-ut-des, then there exists a do-ut-des coalition that dominates it. So we do not need to compare a do-ut-des coalition with all the others in order to see if it is q-do-ut-des, but only with the other do-ut-des coalitions.

Theorem 4 *Given a do-ut-des compatible cooperative game* $NTU[MaS]$ *of a multiagent system MaS, if a possible coalition* c *is not q-do-ut-des, then there exists a do-ut-des coalition* c' *such that* c' *q-do-ut-des dominates* c.

proof: Assume that c is not q-do-ut-des, and per absurdum that each possible coalition c' that q-do-ut-des dominates c is not do-ut-des. So, due to the Definitions 15 and 12, we have that (1) there exists a_i such that $c \prec_i c'$ and for all $a_j \in Dom(c')$, $c \precsim_j c'$. (2) c' is not do-ut-des: there exists a $c'' \subset c'$ and a $a_h \in Dom(c')$ such that $c' <_h c''$ and for all $a_k \in Dom(c'')$, $c' \leq_k c''$.

Since $NTU[MaS]$ is do-ut-des compatible, it is the case that (1) $c' \precsim_k c''$ and hence, being $Dom(c'') \subseteq Dom(c')$, $c \precsim_k c''$, (2) $c' \prec_h c''$ and hence, being $a_h \in c'$, $c \prec_h c''$. But this means that c'' q-do-ut-des dominates c. Now let consider a sequence of possible coalitions $c_1, \ldots, c_m$, such that (1) $c_m = c''$, (2) for all $1 \leq i \leq m-1$, c_i do-ut-des dominates c_{i+1} (3) c_1 is do-ut-des. Such a sequence certainly exists, indeed let consider the set of all the sequences $c_1, \ldots, c_m = c''$ such that for all $1 \leq m-1$, c_i do-ut-des dominates c_{i+1}, and denote this set with $DUD[c'']$. In the case there is no a possible coalition that do-ut-des dominates c'', then there exists only one sequence in $DUD[c'']$ and it is composed by c'' itself. On the contrary there is at least a sequence in $DUD[c'']$ of length equal to 2. Now if $c_1, \ldots, c_m \in DUD[c'']$ and c_1 is not do-ut-des, then there exists a sequence of length $m+1$ in $DUD[c'']$. But his entails that if for all $c_1, \ldots, c_m \in DUD[c'']$, c_1 is not do-ut-des, then $DUD[c]$ has infinite cardinality, but this is impossible since there exists only a finite number of sequences of possible coalitions $c_1, \ldots, c_m = c''$ such that for all $1 \leq i \leq m-1$, $c_i \subset c_{i+1}$.

Following the proof done for c'', for each $1 \leq i \leq n$, c_i q-do-ut-des dominates c, but this means that c_1 is a do-ut-des coalition that q-do-ut-des dominates c. □

Example: In the our boxes-chips game the cardinality of the possible coalitions is 157. The do-ut-des coalitions are only 14, i.e. the 91% of all coalitions, and the cardinality of the q-do-ut-des coalitions is 4. An example of q-do-ut-des coalition is described in the following table.

τ_A	g
$p_1 \xrightarrow{r} p_1, p_3 \xrightarrow{r} p_1$	rr
$p_2 \xrightarrow{b} p_4, p_2 \xrightarrow{b} p_4$	bb
$p_1 \xrightarrow{g} p_3, p_3 \xrightarrow{b} p_3$	bg
$p_4 \xrightarrow{r} p_2, p_4 \xrightarrow{g} p_2$	rg

All the players satisfy their own goals. In the initial configuration p_1 and p_3 need respectively only one chip to satisfy their goals. Since no other player provides to them the needed chip without achieve his own goal, p_1 and p_3 have to provide at least one chip. But they provide only one chip, so they reach the best score they can obtain. Instead, p_2 and p_4, that initially have chips and boxes totally unmatched, have to provide both their chips in order to fill their boxes. ■

7. Conclusions

In this paper we have provided a description of a multiagent system as a collection of individual agents, each of them with their own goals and skills. Then we have defined the sets of goals assurable by a set of agents. Our definition considers with respect to the analogous definitions in Boella et al. [2] a minimalization on the sets of tasks assigned to each agent.

Starting from the notion of assurable goals we have defined a cooperative game associated to a multiagent system and considered the notion of core as solution criterion for it. We have defined the qualitative criterion of admissibility for the formation of a coalition called do-ut-des and proved that it can be used as a method to reduce the space on which to select a subset of the core.

This reducing process can be computationally profitable. The inclusion of a coalition c in the core, depends on a comparison of c all the other possible coalitions. This means that in order to establish if a coalition is in the core the entire cooperative game relative to a multiagent systems have to be given and hence all the achievement capabilities have to be calculated for each set of agents. In our approach, instead, a coalition c have all the necessary information to establish if it satisfies the do-ut-des property, therefore the do-ut-des check can be parallelized with respect to the process that calculate the cooperative game relative to the multiagent system.

References

[1] R.J. Aumann. The core of a cooperative game without side payments. *Transaction of the American Mathematical Society*, 98:539–552, 1961.

[2] G. Boella, L. Sauro, and L. van der Torre. Power and Dependence Relations in Groups of Agents. In *Procs. of IAT'04*, pages 246–252, 2004.

[3] G. Boella, L. Sauro, and L. van der Torre. Reducing Coalition Structures via Agreement Specification. In *Procs. of AAMAS'05*, 2005.

[4] B. Grosz, S. Kraus, S. Talman, B. Stossel, and M. Havlin. The influence of social dependencies on decision-making: Initial investigation with a new game. In *Procs. of AAMAS'04*, pages 782–789, 2004.

[5] S. Kraus, O. Shehory, and G. Taase. The advantages of compromising in coalition formation with incomplete information. In *Procs. of AAMAS'04*, pages 588–595, 2004.

[6] M. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1994.

[7] T. Sandholm and V. Lesser. Coalitions among computationally bounded agents. *Artificial Intelligence*, 94(1):99–137, 1997.

[8] O. Shehory and S. Kraus. Methods for Task Allocation via Agent Coalition Formation. *Artificial Intelligence*, 101(1-2):165–200, 1998.

[9] M. Wooldridge and P.E. Dunne. On the computational complexity of qualitative coalitional games. *Artificial Intelligence*, 158(1):27–73, 2004.