

Chapter 4

Quantifying Relative Autonomy in Multiagent Interaction

SVIATOSLAV BRAYNOV* and HENRY HEXMOOR**

**State University of New York at Buffalo, 210 Bell Hall, Buffalo, NY 14260-2000*

***Computer Science & Computer Engineering Department, Engineering Hall, Room 313, Fayetteville, AR 72701*

Key words: multiagent systems, utility-based decision making

Abstract: In the paper we introduce a quantitative measure of autonomy in multiagent interactions. We quantify and analyse different types of agent autonomy: (a) decision autonomy versus action autonomy, (b) autonomy with respect to an agent's user, (c) autonomy with respect to other agents and groups of agents, and (d) a measure of group autonomy that accounts for the degree with which one group depends on another group. We analyse the problem of composing multiagent group with maximum overall autonomy and we prove that this problem is NP-complete. Therefore, finding the optimal group or agent with whom to share a task (or to whom to delegate a task) is in general computationally hard.

1. INTRODUCTION

The concept of autonomy plays an important role in multiagent interactions. It captures capability of an individual or collective to decide and act consistently without outside control or intervention. Autonomy has been a subject of continuous interest in different research areas including multiagent systems (Castelfranchi, 1995 and 2000, Hexmoor, 2000a; Hexmoor and Kortenkamp 2000), sociology (Dworkin, 1988), and philosophy (Mele, 1995; Schneewind, 1997).

The notion of autonomy has been used in a variety of senses and has been studied in different contexts. Depending on the context we can have autonomy with respect to the physical environment or with respect to the social environment. In a social context we can differentiate between individual or group autonomy, individual autonomy with respect to a group, group autonomy with respect to another group, group autonomy with respect to an individual, etc.

In the context of physical environment, autonomy is one's ability to act independently of physical forces and influences. This kind of environmental autonomy usually presupposes (a) some kind of control or mastery over the environmental events and objects, (b) imperviousness or liberty from uncertainties, and (c) robustness against environmental changes. Autonomy in the context of social multiagent interaction is concerned with variations in an agent's ability when other agents are involved. It might be desirable for an agent's performance to be invariant to interactions i.e., stable and independent of other agents.

Autonomy could refer to acting or to decision-making. In this case we can distinguish between action autonomy and decision autonomy. Action autonomy relates to an agent's ability to perform successful actions despite the influence of environmental factors and the presence of other agents. If an agent has only partial information or partial control over the environmental factors or other agents, the agent may not be completely autonomous. For example, a digital financial assistant, looking for a stock in an electronic exchange may not have complete control over the transaction price. The price may depend on the actions of other agents, market conditions, network communication delays, etc.

While action autonomy applies to one's ability to behave as one wishes to, decision autonomy refers to one's ability of making efficient decisions. That is, the ability to find, enable, and choose the most preferred option from a set of available options. To make decisions agents need to know what they want and what their choices are. In other words, agents must have preferences over a set of choices. While human agents could change their preferences or generate new ones, software agents are not free to have independent preferences. This results from the fact that a software agent always acts on behalf of its user, thereby reflecting the user's preferences and desires. As a proxy of its user, a software agent is required to decide in the way the user would have decided. Therefore, agent autonomy is always relative to what the user wants and expects from the agent.

Autonomy in the context of agent-user interaction captures the notion of an agent's ability to act efficiently without the user's intervention. An agent may not be autonomous with respect to its user if it needs permission for certain actions. An agent that has full permission may still not be autonomous, if it has partial knowledge about the user's preferences and the ways in which these preferences could be met.

Autonomy can also have several modes. We may have *capacity or competency* for autonomy, or the *actual condition* of autonomy, or the *authority* to act or decide autonomously. The capacity mode deals with one's potential of being autonomous. An agent could be fully autonomous in certain circumstances without being given the chance to act and generate its own goals. For example, a user who doubts an agent's ability to act autonomously might restrict the agent's power to perform tasks or make decisions even when the agent is completely autonomous.

Throughout this paper we assume a distributed problem-solving environment. That is all agents serve the same user. An interaction where different agents act on behalf of different users, is beyond the scope of this paper.

In this paper we focus on the degree of autonomy. It is obvious that between the lack of autonomy and the complete autonomy there is a wide range of intermediate states that describe an agent's ability to act and decide independently. The degree of autonomy might change over time as an agent acquires more knowledge and more expertise. The degree of autonomy could also be a random variable if there is uncertainty about one owns' abilities, the user preferences, the environment, etc. In multiagent teams where an agent has a partial knowledge about the state of the group, the degree of autonomy could be a distributed knowledge, i.e., a combined knowledge that no single agent knows.

The ability to measure the degree of agents' autonomy is of significant importance for designing robust, scalable and reliable multiagent systems. First, it facilitates task allocation by providing a criterion for choosing the most autonomous candidate for a particular task. Second, a measure of team autonomy could facilitate team organization and configuration. By measuring the autonomy of various agent groups and group configurations, a system designer can choose the most autonomous group. Third, agents could keep track of the degree of their autonomy, and dynamically reorganize themselves in order to increase it.

In order to apply a qualitative measure of autonomy we need a scale and some criterion for distinguishing between autonomous, semi-autonomous and non-autonomous behavior. Since autonomy is a relative concept depending on what a user (or another agent) expects from an agent, we use the agent's performance as a scale. In other words, the degree of autonomy could be measured as a distance from some standard of achievement that the user applies and that depends on the task the agent performs. If an agent errs and fails continuously while performing a task, we may be reluctant to call it autonomous no matter how self-directing and independent it is (Meyers, 1989). Moreover, it is possible for an agent to be autonomous and non-autonomous at the same time with respect to the same task, if different users apply different performance standards.

In our previous research we investigated efficiency as a basis for teaming among agents and presented a performance-based teaming algorithm (Hexmoor and Duchschere, 2001). In this paper we consider an agent's performance relative to its context as an indicator of autonomy. Autonomy is understood in the context of the environment that can be made up of events, object, and other agents. In other words, in order to evaluate the degree of an agent's autonomy we have to put the agent in touch with objects, events, and other agents, and compare the agent's performance with our expectations. If an agent can perform *in the presence* of other agents at least as well as it performs in isolation, then the other agents are not restricting the agent's autonomy. By putting agents together, we make no assumptions about cooperation or coordination or other inter-agent attitudes. We also do not make any assumptions about psychological influences among agents. When in presence of other agents, these other agents are considered as a distinguished part of the environment. For example, a factory worker who gets part for a widget and assembles it may work alone or alongside other agents who do the same. This factory worker might experience gains or losses in its productivity in the presence of these other workers, who are a special part of its environment.

Autonomy is by no means identical to efficiency. Later in this paper we will present the *autonomy-efficiency dilemma* and we will show that (a) autonomous behavior could be inefficient, as well as (b) efficient behavior might not necessarily be autonomous. Relative performance i.e., the performance in a context, however, could be a good indicator of the degree to which the context restricts or extends individual autonomy. The comparative analysis of autonomy allows us to define and differentiate between different kinds of autonomy relationships: an agent in the context of a group, a group in the context of another group, and a group in the context of an agent. In this paper we emphasize relative autonomy in the context of a user, environmental factors, and in a social setting.

Barber and Martin (1999) proposed another quantitative measure of agent autonomy. They defined the degree of autonomy as an agent's relative voting weight in decision-making. This approach has several advantages. For example, it allows for explicit representation and adjustment of the agents' autonomy. To our knowledge, it has been the first attempt to describe an agent's autonomy from a decision-theoretic point of view. There are several indexes of agents' voting power (Banzhaf, 1965; Shapley and Shubik, 1954). The game-theoretic research, however, reveals that an agent's relative voting weight does not correlates well with the voting power, since it does not take into account the frequency with which an agent's vote is pivotal (Banzhaf, 1965).

The concept of autonomy is closely connected to the concepts of power, control and dependence (Brainov and Sandholm, 1999; Castelfranchi, 2000). An agent is autonomous with respect to another agent, if it is beyond the scope of

power of that agent. In other words, autonomy presupposes some independence or at least restricted dependence. Further exploration of the relationship between power, control, and autonomy is beyond the scope of this paper.

The paper is organized as follows. In the next section we discuss autonomy as a relation including several constituents. We argue that the notion of autonomy is usually relative and depends on what an agent's user wants and expects from an agent. Section 3 defines decision-autonomy as a combination of preference-autonomy and choice-autonomy. In Section 4 we study autonomy in the context of user-agent interaction. We propose a measure of autonomy that indicates the degree to which an agent is independent of its user. In Section 5 we analyse autonomy in social multiagent interaction and introduce quantitative measures of group autonomy. We study the problem of finding an agent group with maximum overall autonomy and prove that this problem is NP-complete.

2. AUTONOMY AS A RELATION

We consider autonomy as a relation including four constituents:

- The *subject of autonomy*: the entity (a single agent or a group of agents), which acts or makes decisions.
- The *object of autonomy*. The object of autonomy could be a goal or a task that the subject wants to perform or to achieve. It could also be a decision that the subject wants to make.
- The *affector of autonomy*: the entity that has an impact on the subject's decisions and actions, thereby affecting the final outcome of the subject's behavior. The affector could be the physical environment, another agent (including the user), or group of agents. The affector could either increase or decrease the autonomy of the subject.
- *Performance measure*: a measure of how successful the subject is with respect to the object of autonomy.

Viewing autonomy as a relation suggests that autonomy is a relative concept. It depends on who is the subject of autonomy, what the subject is trying to achieve, who is affecting the subject's behavior, and who evaluates the subject's performance. The subject of autonomy, for example, could be autonomous with respect to particular object (task, goal, decision, etc.), particular affector (the user, the other agents, the environment), and a particular moment of time. As a result, an agent may be autonomous with respect to one task and non-autonomous with respect to another task, autonomous by the judgement of one user, and non-autonomous by the judgement of another user.

The agent's user plays an important role in defining autonomy. It is the user who specifies the agent's performance measure. That is, the criterion that specifies how successful an agent is. Without such a measure it is not possible to

determine whether the agent has succeeded in performing its task or achieved its goal. Obviously, there is no universal performance measure applicable to all agents, all users, and all situations. A goal could be achieved with various degrees of completeness, some of which could be considered successful in a particular context. This suggests that the notion of autonomy is relative and depends on what the user expects from an agent. For example, an agent could have been autonomous in the past and non-autonomous in the present, if the user raises his performance standard over time.

It is worth mentioning that an agent (not necessarily human) could also apply a performance standard with respect to another agent if the first agent delegates a task (or a portion of a task) to the second agent. In this case the performance standard is a derivative from the original performance standard set up by the agent's user.

The user's performance measure is usually represented as a preference relation between world states. That is, the user considers some states more desirable than others. The simplest way to represent the user's preferences is to use a binary satisfied or not satisfied criterion applied to the world states prevailing at the end of the agent execution. For example, some world states can be labelled as success and others as failure. Such an approach, however, cannot account for several factors directly related to goal achievement like efficiency, timeliness, partial success, etc. It considers any two successes equivalent without regard to how efficient they are. Analogously, any two failures are considered equivalent despite the fact that an agent could fail at the final stage where the goal is almost achieved. More realistic preference relations should account for various degrees of goal achievement. Such preferences could be applied to both intermediate and final stages of agent activity. When applied to intermediate stages, user's preferences can help an agent plan and search efficiently in the space of possible states of the world.

User preferences are usually represented as a real-valued utility function, i.e., an order preserving function over the space of possible world states. Such a function assigns a single number to a state in a way that a more desirable state receives a higher number. A utility maximizing agent is trying to choose an action that maximizes its utility function. That is, the agent chooses the action, which the user would have chosen if he was to make the choice. Suppose, for example, that an agent has to arrange a holiday vacation on behalf of its user. The user is looking for a hotel that is cheap and close to the beach. The user's preferences could be specified by a utility function weighing the price against the distance to the beach:

$$\begin{aligned}U(\text{Hotel-1}, \$90, 1 \text{ mile}) &= 1 \\U(\text{Hotel-2}, \$70, 2 \text{ miles}) &= 2 \\U(\text{Hotel-3}, \$60, 2.5 \text{ miles}) &= 1\end{aligned}$$

According to this utility function, the user prefers Hotel-2 to Hotel-1 and at the same time is indifferent between Hotel-1 and Hotel-3. The same preferences can be represented by the following binary preference relation: $P(\text{Hotel-2}, \text{Hotel-1})$, $P(\text{Hotel-2}, \text{Hotel-3})$, $P(\text{Hotel-1}, \text{Hotel-3})$, $P(\text{Hotel-3}, \text{Hotel-1})$, where $P(x,y)$ means that the user prefers x to y . The indifference between x and y is defined as both $P(x,y)$ and $P(y,x)$. That is, an agent is indifferent between x and y , if he prefers both x over y , and y over x .

3. DECISION AND ACTION AUTONOMY

We assume that at any moment in time an agent has a set, C , of available options, and a preference relation P defined on C . An option c_1 is preferred to an option c_2 iff $P(c_1, c_2)$. Every agent can be viewed as an abstract optimizer. An agent picks c_1 , $c_1 \in C$, if there is no c_2 , $c_2 \in C$, such that $P(c_2, c_1)$. That is, an agent chooses an option, if there is no better alternative in the space of available options C .

Both the set of available choices C , and the preference relation P could affect the decision-making of an optimizing agent. The set of available alternatives affects the agent's decision-making in two ways. First, an agent may not be able to choose the best possible option because he might be unaware of it. For example, an agent could be limited in its reasoning, or may have incomplete information. Second, an agent could be aware of the existence of a particular option and yet consider it infeasible at the moment of decision-making. For example, an option could require additional resources, an action might need a preparation, or the action takes too long and does not meet a deadline.

Another factor that affects an agent's decision-making is the preference relation P . To be successful, an agent needs to choose among several options in a way that best meets the user's preferences. Since the agent is going to be evaluated by the user, the agent needs to look at the world from the user's perspective. The problem is, however, that a concise presentation of user's preferences like a utility function seldom exists (Fishburn, 1970). If the space of all available choices C is large, it may not be computationally efficient to express user's preferences for all possible choices. Another problem is that the user usually cannot predict all possible contingencies, eventualities, and choices which the agent encounters. Suppose that in the example above the agent has come across a new Hotel-4, (Hotel-4, \$75, 1 mile), the user did know about. Since the difference in price between Hotel-2 and Hotel_4 is small, it is quite possible that the user would prefer Hotel-4 to Hotel-2. This leaves the agent with incomplete (and sometimes uncertain) information about the user's preferences.

Since both the choice set, C , and the preference relation, P , affect an agent's decision-making, it is natural to consider decision autonomy as a function of

both of them. An agent is *preference-autonomous*, if the agent has complete knowledge of the user's preferences. This means that the agent has the same likes and dislikes as the user, and the user does not need to tell the agent how to make the right choice. That is, in every choice situation, the agent can evaluate all available options and choose the most desirable one. A preference-autonomous agent can act as a perfect copy of its user whenever the agent and the user face the same choice problem.

Let $F(C,P)$ denote the agent's choice function. That is, given the set of available choices C and the preference relation P , the agent will choose an option $F(C,P) \in C$.

An agent's degree of preference-autonomy could be measured by the degree of efficiency of its choice. Let P' is the agent's preference relation (P' is usually a subset of the user's actual preference relation P). Then $F(C,P')$ is the choice that the agent makes, and $F(C,P)$ is the choice that the agent would have made if he had complete information about user's preferences. The degree of preference-autonomy is defined as the ratio of the efficiency of its choice to the maximum possible efficiency the agent would achieve if he knew the user's preferences.

Definition 1. The degree of *preference-autonomy* is the ratio:

$$\frac{U(F(C,P'))}{U(F(C,P))}$$

where $U(F(C,P))$ is the user's utility function.

If an agent has only partial knowledge of the user's preferences, then its preference relation P' is a subrelation of the user's relation P . That is, $P' \subseteq P$. Neither P nor P' need be complete relations. For example, P could be partial order if the user is not able to evaluate all available choices. Inability of the user to choose between different options could be a result of several reasons. First, the user might not feel difference between two options. Second, the user could be uncertain as to his preference between the options. Third, the user could find the comparison difficult, or even impossible. We assume that whenever one has several equally desirable alternatives, he will choose one of them at random. This rule applies to both the user and the agent. In other words, if an agent does not know the user's preferences on a set of choices, then the agent considers all choices equally desirable, and chooses one of them at random. The more incomplete the preference relation is, the more random choices an agent has to do, and the less effective his choices could be. In the hotel example the agent does not know the user's preferences for Hotel-2 and Hotel-4, and chooses each hotel with probability $\frac{1}{2}$. Suppose, further, that if the user knew about Hotel-4, he would have assigned $U(\text{Hotel-4}, \$75, 1 \text{ mile}) = 3$. Since the agent chooses at random between Hotel-4 and Hotel-2, the expected utility of its choice is 2.5. Hence, the agent has preference-autonomy of $2.5/3=0.83$. In other words, preference-autonomy measures the efficiency of the agent's decision-making

with respect to the user. This is a natural assumption, since it is the user who evaluates the agent's choices.

The situation when the agent has to ask the user for permission to make choices represents a special case of preference-autonomy. If the user is uncertain as to whether the agent has complete preference information, the user may impose restrictions on critical choices. For example, the user could prefer to make all critical choices by himself, thereby forbidding the agent to make important decisions. Another alternative is for the agent to report its choice to the user and wait for the user's confirmation. In both of these cases $U(F(C,P'))=0$ and the agent's preference-autonomy is zero. For example, the user might forbid the agent to reserve a hotel whose price exceeds \$100.

It is worth noting that having complete knowledge about the user's preferences could be insufficient for full decision autonomy. Although a preference-autonomous agent chooses the right choice from a set of choices, there is still no guarantee that the choice is optimal. In many cases an agent could be unaware of the existence of a better choice, or some choices could not be available to him. In other words, an agent could be a perfect optimizer and yet have an incomplete set of choices.

In the definition of preference-autonomy we focused on the completeness of preference relation. That is, we assumed that the choice set C was fixed and investigated how the completeness of P affects an agent's autonomy. Let us now turn our attention to the completeness of the choice set C .

Definition 2. The degree of choice-autonomy is the ratio:

$$\frac{U(F(C',P))}{U(F(C,P))}$$

Where C' is the agent's choice set, and C is the user's choice set.

According to Definition 2 the degree of choice-autonomy measures the completeness of the agent's choice set with respect to the user's choice set. Unlike preference-autonomy which is always less or equal to one, choice-autonomy could be greater than one. That is, an agent might find more choices than its user is able to find. In the example above, suppose that the user and the agent are independently searching for a hotel. Equipped with better searching capabilities the agent finds all the four hotels, while the user finds only the first three: Hotel-1, Hotel-2, and Hotel-3. Based on his incomplete choice set the user chooses Hotel-2, yielding him a utility of 2. On the other hand, the agent randomly chooses between Hotel-3 and Hotel-4, receiving an expected utility of 2.5. In this case the agent's choice-autonomy is $2.5/2=1.25$, meaning that the user is less choice-autonomous than the agent.

While choice-autonomy measures an agent's ability to find different paths and alternatives for achieving its goals, preference-autonomy relates to the ability of

choosing the right alternative. By bringing these two notions together we arrive at the definition of decision-autonomy.

Definition 3. The degree of *decision-autonomy* is the ratio:

$$\frac{U(F(C',P'))}{U(F(C,P))}$$

where C' is the agent's choice set, and C is the user's choice set, P' is the agent's preference relation, and P is the user's preference relation.

Decision autonomy measures an agent's ability to find perspective solutions, evaluate them according to the user's expectations, and choose the best solution. Depending on the degree of decision-autonomy, an agent could be:

- Completely decision-autonomous (degree of 1): the agent meets the user's expectations.
- Non-autonomous (degree less than 1): the agent falls short of the user's expectations.
- Super autonomous (degree more than 1): the agent exceeds the user's expectations.

4. A MEASURE OF AUTONOMY IN THE CONTEXT OF AGENT-USER INTERACTION

In this section we analyze autonomy with respect to an agent's user. We consider the user as the main agent who has the right to monitor and control an agent's performance. The user takes the responsibility for the agent's performance and gives identification to the agent. Whenever the agent identifies itself, exchanges digital certificates, or carries out a transaction, it acts on behalf of its user. We assume that the user can activate or deactivate the agent at his will (if circumstances allow). It is not necessary for the user to be a human agent. For example, a mobile agent can spawn a new agent and act as a user with respect to that agent. Since an agent acts on behalf of its user, user-agent interaction has greater priority for the agent than the interactions with other agents. Other interactions could be considered as instrumental with respect to the user-agent interaction.

One interesting aspect of autonomy is an agent's ability to keep its identity i.e., an agent's ability to keep its relationship with the user. The problem is that a malicious agent can gain control over another agent by changing the agent's code, accessing sensitive information like access control rights, passwords, private encryption keys, etc. The ability to keep one's identity is a special case of autonomy that does not have a direct counterpart in the human society. This type

of autonomy is central to all other autonomy types. For example, if a malicious agent changes an agent's preference relation, the agent will behave chaotically. In another words, there is no reason to try to measure an agent's action or decision autonomy if the agent is not autonomous enough to keep its identity. In this paper we assume that an agent is always capable of keeping its identity.

Another interesting case of autonomy arises when an agent simultaneously serves multiple users. By multiple users we mean different users with different identities (the case when different users interact with the agent using the same identity is considered as a single user). If the users do not have predetermined priorities, the agent may exhibit autonomy with respect to which user to serve first and how to allocate resources among competing users. In this case we assume that there is always a single administrator with distinguished control rights.

An agent may complete a task with or without the user's supervision. In order to measure the agent's autonomy with respect to its user, we have to know what the user expects from the agent. We assume that the agent's performance can be measured by some criterion of performance v . The criterion v may be thought of as a criterion of partial success, optimization function, index of satisfaction, utility function, etc. The criterion of performance v is determined by the user. For the same task different users may use different performance criteria.

With every agent i we can associate at least two performance measures¹. The first measure v_i^i is agent i 's performance in the case where it acts autonomously i.e., without the user's supervision. The second measure v_U^i is agent i 's performance with the user's supervision. Intuitively, v_U^i represents the maximum performance which the user expects from the agent. In other words, if the user could supervise and help the agent at any step of its action and decision making, the agent will get v_U^i . We make no assumptions, however, about improved performance and, in fact, performance degradation is quite possible. In other words, an agent could be super-autonomous and it can do better by acting alone.

In defining degree of autonomy we follow the standard assumption of keeping all other things equal. That is, the effect of other agents or the environmental events is the same for both measures v_U^i and v_i^i .

Definition 4. By a *degree of individual autonomy* A_U^i (autonomy with respect to the user) we mean the ratio:

$$A_U^i = \frac{v_i^i}{v_U^i}$$

The degree of individual autonomy indicates the extent to which an agent may act independently of the user i.e., what part of an agent's performance must

¹ In the next section we will introduce a complete description of relative performance.

be attributed only to the agent's capabilities. In general, individual autonomy varies between $-\infty$ and $+\infty$. The combined performance of the agent and the user is not necessarily the maximal performance. For example, if the user is not competent enough, the agent may be more efficient by acting autonomously.

According to Definition 4 autonomy is a relative concept. In order to evaluate an agent's autonomy the user must have some criterion of acceptable behavior or some expectation about the agent's behavior. Since different users may have different requirements for a task accomplishment, the same pattern of behavior could be considered as both autonomous and non-autonomous by different users. Suppose, for example, that an agent autonomously fulfills only 90% of its task. A user may consider 90% accomplished task as a success, and may be willing to classify the agent's performance as autonomous. At the same time, another user may consider the same performance as a failure, and may be reluctant to view the agent as autonomous.

5. GROUP AUTONOMY

In multiagent interaction where the agents' actions interfere with one another, an agent may affect the autonomy of other agents both directly and indirectly. For example, an agent might directly help or prevent another agent from achieving its goal. Indirect interaction could occur as a side effect of an agent's behavior. An agent's action may restrict or extend the autonomy of other agents by affecting the environmental conditions, the set of feasible goals, etc. In some cases the effects could be even more indirect. For example, an agent can affect another agent, which in turn may affect the autonomy of a third agent. This prompts for a quantitative measure of the degree of autonomy that takes into account various aspects of multiagent interaction: an agent in the context of a group, a group in the context of another group, and a group in the context of an agent.

We assume that agents can affect one another once they have been deployed in the environment. This is a natural assumption, since we cannot preclude agents from interfering with one another (in a positive or negative way) once they have been brought together. In other words, we assume that it is not possible to divide the environment into different mutually independent groups such that agents can affect one another if and only if they belong to the same group. Then, the question is which agents to deploy and how to structure their interaction? In other words, which subset of agents achieves maximum autonomy, maximum efficiency or some combination of them? This question is different from the problem of finding the optimal coalition structure (Sandholm et al., 1999). A coalition environment implies that agents can be divided into relatively independent groups called coalitions. Each coalition has its own

performance measure (value of the coalition) and the problem is to find a coalition partition that maximizes the sum of coalitions performances. In our case all agents serve the same user and there is no reason to separate them into different coalitions. That is, we assume that the user always deploys only one group of agents (the grand coalition). The problem is which group (or coalition) to deploy. The largest possible group of agents may not always be the most effective one.

To answer this question we need to study the effect of multiagent interaction on individual and group autonomy. That is, how the presence of other agents or groups of agents affects the autonomy of a single agent or a group. In order to evaluate how well an agent is doing in the company of other agents we need some indicators of relative performance. With every agent i we associate a vector of relative performance² $(v_i^i, v_j^i, v_k^i, v_{jk}^i)$. Here v_i^i represents agent i 's performance when it acts alone. v_j^i is agent i 's performance in the company of agent j . In this case agents i and j can interfere with each other either negatively or positively. That is, v_j^i could be greater or smaller than v_i^i . For example, if agent i depends positively on agent j , then $v_j^i \geq v_i^i$. v_k^i is agent i 's performance in the presence of agent k . v_{jk}^i measures agent i 's performance if it acts concurrently with agents j and k . In the case of 3 agents the length of the vector of relative performance is 2^2 . In general, the vector's length is 2^{n-1} , where n is the number of agents. Figure 1 shows a matrix of relative performance for the case of three agents.

$$\begin{pmatrix} v_i^i & v_j^i & v_k^i & v_{jk}^i \\ v_i^j & v_j^j & v_k^j & v_{ik}^j \\ v_i^k & v_j^k & v_k^k & v_{ij}^k \end{pmatrix}$$

Figure 1. Relative performance matrix

The elements of the relative performance matrix should be interpreted as guaranteed performance values. For example, agent i can always get v_j^i in the company of agent j . The actual performance may differ depending on agent j 's behavior, but it is always greater or equal to v_j^i . In other words v_j^i is the minimax performance that agent i can obtain in the company of agent j . That is, no matter how agent j behaves, agent i always gets at least v_j^i .

² For the sake of simplicity we constrain our attention to the case of three agents i , j and k . The results can easily be generalized to an environment with an arbitrary number of agents.

The following definition introduces the concept of autonomy with respect to another agents.

Definition 5. The degree of agent i 's *autonomy with respect to agent j* is

$$A_j^i = \frac{v_j^i}{v_i^i}$$

The degree of agent i 's autonomy with respect to agent j is the ratio of agent i 's relative performance to its individual performance. In other words, the degree of autonomy indicates how well agent i performs in the presence of agent j . It is 1 when agent j does not affect agent i . It could also be 0, if agent j completely blocks agent i . In general, it varies between $-\infty$ and $+\infty$.

Group performance is highly affected by interference among agents. The interference might produce either positive or negative performance. The following dilemma states the problem with interference.

Definition 6: (*Autonomy–efficiency dilemma*) Maximum autonomy does not necessarily implies maximum efficiency, and vice versa.

The autonomy-efficiency dilemma arises when we bring together two types of agents: (a) agents with low efficiency and high autonomy invariance (agents that are impervious to interference from other agents), and (b) agents with high efficiency and high autonomy variance (agents whose performance is highly susceptible to interference with other agents).

The following two-agent example illustrates the autonomy-efficiency dilemma. Suppose that we have two agents i and j whose vectors of relative performance are (5,1) and (2,2) respectively. That is, $v_i^i=5$, $v_j^i=1$, $v_j^j=2$, $v_i^j=2$. The two agents have different levels of individual autonomy. By acting alone agent i gets 5, while agent j gets 2. If the agents are brought together, then agent i gets 1 and agent j gets 2. Therefore, the autonomy of agent i with respect to agent j , A_j^i , is 1/5. This indicates that agent j affects negatively agent i by reducing agent i 's performance 5 times. On the other hand, agent j 's performance does not depend on agent i and it is always 2. That is, agent j is autonomous with respect to agent i . If we are looking for maximum invariance in autonomy, then we have to deploy only agent j . This, however, is not an optimal solution since agent j is less efficient than agent i . Agent j gets 2, while agent i achieves 5. Therefore, if we are looking for maximum efficiency, we have to deploy only agent i . The dilemma autonomy-efficiency arises from the fact that efficient agents may be highly susceptible to interference from other agents and vice versa; agents with high autonomy invariance may not be very efficient. To alleviate this dilemma, we assume that all agents have the same individual autonomy.

Definition 7: *Equally-competent agents* is the assumption that all agents have the same individual performance. That is, $v_i^i = v$, for all agents i .

Under equally-competent agents assumption, each agent i by acting alone can achieve the same standard of performance v . Since all agents are equally competent, the dichotomy of autonomy-efficiency disappears.

The following definition introduces the concept of autonomy with respect to a group of agents.

Definition 8. The degree of agent i 's *autonomy with respect to a group* of agents (j,k) is:

$$A_{jk}^i = \frac{v_{jk}^i}{v}$$

The degree of autonomy with respect to a group measures to what extent the group can restrict or extend an agent's autonomy. A degree of 1 means independence from the group. A degree larger than 1 signals a synergetic interaction. Consider the following example. Suppose that agent i 's vector of relative performance is $(4,6,4,8)$. That is, $v_i^i = 4$, $v_j^i = 6$, $v_k^i = 4$, and $v_{jk}^i = 8$. This implies that agent i depends positively on agent j ($A_j^i = 6/4 = 1.5$). At the same time it is autonomous with respect to agent k ($A_k^i = 4/4 = 1$), and depends positively on the group of agents j and k ($A_{jk}^i = 8/4 = 2$).

In the next definition we introduce the concept of group autonomy. It measures how well agents are doing in a group.

Definition 9. The degree of *group autonomy* of the group of agents (i,j) under the equally-competent agents assumption is:

$$A^{ij} = \frac{v_j^i + v_i^j}{v}$$

The degree of group autonomy compares individual performance with group performance and indicates whether it is worthwhile to put the agents together. If the agents are deployed in a group, the result is $v_j^i + v_i^j$. If only one agent (either one) is deployed, the performance is v .

The following definition is an analogue of Definition 9 in the case when agents are not equally competent.

Definition 10. The degree of *group autonomy* of the group of agents (i,j) is:

$$A^{ij} = \frac{v_j^i + v_i^j}{\max(v_i^i, v_j^j)}$$

Definition 10 measures group autonomy as the ratio of group performance, $v_j^i + v_i^j$, to the best individual performance, $\max(v_i^i, v_j^j)$. Intuitively, a negative

interaction between agents restricts group autonomy, while positive interaction improves it.

The next definition is a generalization of definition 9 to the case of more than three agents.

Definition 11. Group autonomy under the equally-competent agents assumption equals the sum of individual autonomies. That is,

$$A^S = \sum_{i \in S} A_{S-i}^i$$

Where S is a set of agents, and $S-i$ is the set of agents S excluding agent i . If we apply Definition 11 to the group of agents (i,j,k) , we will get

$$A^{ijk} = A_{jk}^i + A_{ki}^j + A_{ij}^k$$

It is worth noting that Definition 11 relies on the equally-competent agents assumption, i.e., that all agents have the same level of individual autonomy. In general, group autonomy is not linear with respect to individual autonomy.

Definition 12. The degree of *autonomy of the group* of agents i and k with respect to agent k under the equally-competent agents assumption is:

$$A_{k}^{ij} = \frac{v_{jk}^i + v_{ik}^j}{v_j^i + v_i^j}$$

The numerator in Definition 12 measures the group performance of agents i and j in the company of agent k . The denominator is the performance of the group without agent k . In the case of more than three agents, we have:

Definition 13. The degree of autonomy of group S with respect to agent k is:

$$A_{k}^S = \frac{\sum_{i \in S} A_{S-i+k}^i}{A^S}$$

where $S-i+k$ is the group S excluding agent i and including agent k .

Definition 13 says that the relative group autonomy (with respect to a third agent) depends positively on relative individual autonomies A_{S-i+k}^i and negatively on the group autonomy A^S .

To illustrate these notions, consider the following example. Suppose that the user can deploy up to three agents i , j and k with the following matrix of relative performance:

$$\begin{pmatrix} v_i^i & v_j^i & v_k^i & v_{jk}^i \\ v_i^j & v_j^j & v_k^j & v_{ik}^j \\ v_i^k & v_j^k & v_k^k & v_{ij}^k \end{pmatrix} = \begin{pmatrix} 4 & 3 & 4 & 3 \\ 4 & 4 & 2 & 1 \\ 4 & 5 & 4 & 4 \end{pmatrix}$$

In this situation agent **i** is autonomous with respect to agent **k**, and depends negatively on agent **j**. Agent **j** is autonomous with respect to agent **i**, and depends negatively on agent **k**. Finally, agent **k** is autonomous with respect to agent **j** and depends positively on agent **i**. The dependence graph is depicted in Fig. 2.

This example shows that since $A_j^{ik} = v_{jk}^i + v_{ij}^k / v_k^i + v_i^k = (3+5) / (4+4) = 1$, the group of agents **i** and **k** is independent from agent **j**. Moreover, the group autonomy of agents **i** and **k** is $A^{ik} = A_k^i + A_i^k = 4/4 + 4/4 = 2.0$. That is, by acting together they can increase their individual performance 2 times. It is easy to check that the maximum group autonomy is achieved when all agents are put together, $A^{ijk} = A(i/jk) + A(j/ik) + A(k/ij) = 3/4 + 1/4 + 5/4 = 2$. This is not apparent from the initial statement of the problem, since agent **j** relates negatively to agents **i** and **k**.

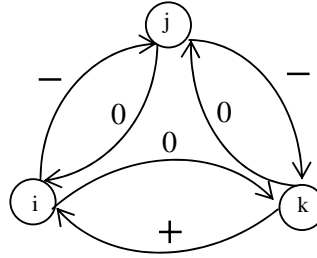


Figure 2: Dependence graph.

The problem of finding the group with maximum autonomy is of significant importance for multiagent interaction. A related issue is finding a group of agents with minimum variance in their autonomy. This is important for building fault tolerant multiagent groups with dynamic membership. If agents were allowed to join and leave groups, we would not want the group’s performance to be significantly affected. We have looked at this problem previously [Hexmoor, 2000b]. According to the following proposition the problem of finding the group with maximum autonomy is computationally hard. The problem is even more difficult if we have to account for the autonomy-efficiency dilemma.

Proposition 3. Finding a group with maximum autonomy is *NP-complete*.

Proof. The decision problem can be defined as follows. Given relative performance matrix, for some real number N , does there exist a group of agents whose group autonomy is N ?

The problem is in NP because verifying the degree of autonomy for a given group can be done in polynomial time. It involves summing the agents’ relative performance measures and dividing the result by the individual performance measure.

What remains to be shown is that the problem is NP-hard. We prove this by reducing the subset-sum problem to our problem. The subset-sum problem is the

following: given a finite set of natural numbers S and a number K , is there a subset S' , $S' \subseteq S$, whose elements sum to K ? This is a classic NP-complete problem (Cormen et al., 1990).

We use the following reduction. Let $N=K$. Let the S be the set of all agents. We associate every agent i with some natural number v^i . Let the relative performance vector of agent i be $(1, v^i, v^i, v^i, \dots)$. That is, agent i 's individual autonomy is 1 and its relative performance is always v^i . Now, the elements of a subset of numbers S' sum to K if and only if the subset of agents S' that has a group autonomy K . Thus, our problem is NP-hard.

6. CONCLUSIONS

In this paper we view autonomy as a relation between several constituents and we argue that autonomy can be usefully seen as relative. It depends on who is the subject of autonomy, what the subject is trying to achieve, who is affecting the subject's behaviour, and who evaluates the subject's autonomy. The paper proposes and studies several measures of autonomy. The first measure defines decision-autonomy as a combination of preference-autonomy and choice-autonomy. While choice-autonomy measures an agent's ability to find different paths and alternatives for achieving its goals, preference-autonomy relates to the ability of choosing the right alternative. The second measure relates to autonomy with respect to agent-user interaction. The third measure defines autonomy among groups and individuals. Our measures are domain independent and do not rely on specific interaction protocols.

The ability to measure the degree of agents' autonomy is of significant importance for designing robust, scalable and reliable multiagent systems. First, it facilitates task allocation by providing a criterion for choosing the most autonomous candidate for a particular task. Second, a measure of team autonomy could facilitate team organization and configuration. By measuring the autonomy of various agent groups and group configurations, a system designer can choose the most autonomous group. Third, agents could keep track of the degree of their autonomy, and dynamically reorganize themselves in order to increase it.

We also studied the problem of finding a multiagent group with the maximum autonomy. We proved that this problem is NP-complete. Therefore, it is in general computationally hard to find the optimal group with whom to share a task (or to whom to delegate a task). This suggests development of approximation algorithms for measuring and adjusting autonomy.

7. REFERENCES

- Banzhaf, J. 1965. Weighted Voting Doesn't Work: A Mathematical Analysis. *Rutgers Law Review* 19:317-343.
- Barber, S., Martin, C. 1999. Agent Autonomy: Specification, Measurement, and Dynamic Adjustment. In Proceedings of the *Autonomy Control Software Workshop, Agents '99*, pp. 8-15. May 1-5, 1999, Seattle, WA.
- Brainov S., Sandholm T. 1999. Power, Dependence and Stability in Multiagent Plans. In Proceedings of *AAAI'99*, pp. 11-16, Orlando, Florida.
- Castelfranchi, C. 1990. Social Power. In Demazeau Y. and Muller J.-P. eds. *Decentralized AI - Proceedings of the First European Workshop on Modeling Autonomous Agents in a Multi-Agent World*, pp. 49-62. Elsevier Science Publishers.
- Castelfranchi, C., 1995 Guaranties for Autonomy in Cognitive Agent Architecture. In N. Jennings and M. Wooldridge (eds.) *Agent Theories, Architectures, and Languages*, pp. 56-70, Springer-Verlag.
- Castelfranchi, C. 2000. Founding Agent's Autonomy on Dependence Theory, In proceedings of *ECAI'01*, pp. 353-357, Berlin.
- Cormen, T., Leiserson, C., Rivest, R. 1990. *Introduction to Algorithms*. MIT Press.
- Dworkin, G. 1988. *The Theory and Practice of Autonomy*. Cambridge.
- Fishburn, P. 1970. *Decision Theory for Decision Making*. John Wiley and Sons.
- Hexmoor, H. 2000a. A Cognitive Model of Situated Autonomy. In Proceedings of *PRICAI-2000 Workshop on Teams with Adjustable Autonomy*, Australia.
- Hexmoor, H. 2000b. Towards Empirical Evaluation of Tradeoffs between Agent Qualities, In *PRIMA-2000*, (C. Zhang and V. Woo, eds), LNAI Volume 1881, Australia.
- Hexmoor, H., Duchscherer, H. 2001. Efficiency as Motivation for Teaming, In *Proceedings of FLAIRS 2001*, AAAI press.
- Hexmoor, H., Kortenkamp, D. 2000. Autonomy Control Software, An introductory article of the special issue of *Journal of Experimental and Theoretical Artificial Intelligence*, Kluwer.
- Jennings, N., Sycara, K., Wooldridge, M. 1998. A Roadmap of Agent Research and Development. *Autonomous Agents and Multi-Agent Systems*, 1:7-38.
- Meyers, D. *Self, Society and Personal Choice*. Columbia University Press, 1989.
- Mele, A. 1995. *Autonomous Agents: From Self-Control to Autonomy*. Oxford University Press.
- Sandholm, T., Larson, K., Anderson, M., Schehory, O., Tohme, F. 1999. Coalition Structure Generation with Worst Case Guarantees. *Artificial Intelligence*, 111: 209-238.
- Schneewind, J. B., 1997. *The Invention of Autonomy: A History of Modern Moral Philosophy*, Cambridge Univ. Press.
- Shapley, L., Shubik, M. 1954. A Method for Evaluating the Distribution of Power in a Committee System. *American Political Science Review*, 48:787-792.

