In the IJCAI-01 workshop on Autonomy, Delegation, and Control: Interacting with Autonomous Agents, 2001.

Quantifying Relative Autonomy in Multiagent Interaction

Sviatoslav Brainov
State University of New York at Buffalo
226 Bell Hall
Buffalo, NY 14260-2000

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

Abstract

In the paper we introduce a quantitative measure of autonomy in multiagent interaction. We quantify and analyze different types of agent autonomy with respect to (a) an agent's user, (b) the other agents, and (c) the other groups of agents. We also introduce a measure of group autonomy that accounts for the degree with which one group depends on another group. We analyze the question of finding a multiagent group with maximum overall autonomy and we prove that this problem is NP-complete. Therefore, the problem of finding the optimal group or agent with whom to share a task (or to whom to delegate a task) is computationally hard in general. This prompts for developing approximation algorithms for measuring and adjusting autonomy.

1. Introduction

The concept of autonomy plays an important role in multiagent interaction. It relates to an individual or collective ability 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. Autonomy could be relative to an individual or a group. Autonomy can be in regards to either acting or decision-making. Depending on the context we can have autonomy with respect to the physical environment or with respect to the social environment. A usual way to look at autonomy is to see it as self-control, which is an ambiguous concept, because it might be referring to either the *capacity or competency* to control oneself, or to the *actual condition* of self-control, or to the *authority* to control oneself.

It is clear that the main problem of autonomy is to explain how autonomous agents make decisions and how they act upon them. In this regard it is worthwhile to distinguish between action autonomy and decision autonomy. Action autonomy relates to the way an agent acts in the environment. For example, an agent may have partial control over the environmental events that affect the outcome of an action. A digital financial assistant may be authorized to buy a stock in an electronic exchange. The agent may, however, not have control over the transaction price. The price may depend on the actions of other financial assistants and on the network communication delay.

Decision autonomy is concerned with the ability of an agent to make consistent choices. A decision autonomous agent must have knowledge about the user's preferences and the potential alternatives. We may be reluctant to classify a digital financial assistant as autonomous, if it keeps asking its user what stock to buy and at what price.

Throughout this paper we assume a distributed problem-solving environment. There is a single user interested in a single task, which can be achieved by deploying one or more agents. A multiagent interaction, where different agents act on behalf of different users, is beyond the scope of this paper.

In general, autonomy can be considered to have the following constituents:

- The subject/agent of autonomy: the entity (a single agent or a group of agents), which has to be considered autonomous.
- The influencer of autonomy: the entity that influences the autonomy of the subject./agent. It could be the human user, physical environment, another agent or group of agents.
- The scope of autonomy: the specific means by which the influencer can affect autonomy. This can include actions, resources, information, indirect means, etc.
- The object of autonomy: includes all elements with respect to which the subject/agent can be autonomous. This could be a single action, goal, task, etc.

- The degree of autonomy: measures that capture the extent to which influencer can affect autonomy.

There may be several objects of autonomy and each object may contribute in a different way to the overall autonomy. At the same time there could be several influencers; each influencer can have different means and each means may affect one or more objects of autonomy in different ways.

Depending on the influencer we can have autonomy with respect to an agent's user (or users), autonomy with respect to the physical environment, and autonomy with respect to other agents (both human and artificial). 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. In the context of interaction with the physical environment, autonomy is one's ability to act independently in the physical environment. This kind of environmental autonomy usually presupposes (a) some kind of control or mastery over the environmental events and objects, and (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 interaction.

In this paper we focus on the last constituent of autonomy: the degree of autonomy. The degree of autonomy measures to what extent the influencer can affect the object of autonomy and how the change relates to the overall autonomy. In our case the object of autonomy is a task. It is evident, that the task is a complex object that is made up of several other objects (actions, plans, etc.). In order to simplify the analysis we consider the task as a single object and assume that during the task execution agents always make optimal decisions.

In order to apply a qualitative measure of autonomy we need a scale and some criterion for distinguishing between autonomous and non-autonomous behavior. Since the object of autonomy is a task and it is expected to be executed efficiently, we use an agent's performance as a scale. Since agents are performing a task on behalf of other agents, autonomy is related to some standard of achievement that derives from the function an agent serves. If an agent's performance is error-prone and continuously fails during a task, we may be reluctant to call it autonomous regardless of how self-directing and independent it might be (Meyers, 1989). Autonomy depends on others' expectation and is relative to those expectations. For example, an agent may be autonomous with respect to one task and not autonomous with respect to another task. 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. The relativeness of autonomy becomes more prominent in environments where the user cannot predict all contingencies upfront.

In our previous research we investigated efficiency as a basis for teaming among agents and presented a performance-based teaming algorithm (Hexmoor and Duchscherer, 2001). In this paper we consider an agent's performance relative to its context as an indicator of autonomy. We consider autonomy as a relative notion. It 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. 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 working in the presence of other agents, we make no assumptions about explicit cooperation or coordination or other interagent 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 parts 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 found with autonomous agents. 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 define 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. Several indexes of agents' voting power have been proposed (Banzhaf, 1965; Shapley and Shubik, 1954). The game-theoretic research, however, reveals that an agent's relative voting weight is not always a good measure of 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 related 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 influences of control and 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 analyze 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. Section 3 presents autonomy in the context of environmental factors and defines a corresponding autonomy measure. In Section 4, we analyze autonomy in social multiagent interaction and introduce quantitative measures of group autonomy. We explore the idea of finding an agent group with maximum overall autonomy and prove that this problem is NP-complete.

2. A Measure of Autonomy in the Context of User-Agent 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 provides identification for the agent. Whenever the agent identifies itself, exchanges digital certificates, or carries out a transaction, it acts on behalf of the user. Under the right circumstances we assume the user can activate or deactivate the agent at her will. The user is typically a human and typically the owner of the agent, which takes legal responsibility for the agent. However, 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.

An interesting aspect of autonomy is an agent's ability to maintain a sense of self and identity i.e., an agent's ability to keep its relationship with the user. An agent's identity becomes important given the agent's code, sensitive information (financial information, for example), access control rights, passwords, digital certificates can be accessed or altered by malicious third parties.

An interesting case arises when an agent simultaneously serves multiple users. By multiple users we mean users with different identities (the case when different users interact with the agent sharing the same identity is considered as a single user). If the users do not have predetermined priorities for the agent, the agent may exhibit autonomy by following the first-come first-served rule. The agent has to determine allocation of resources among competing users. In this case we assume that there is always a single administrator among the users with distinguished control permissions.

An agent may complete the 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 the extent to which the user's supervision is helpful for the agent. We assume that the agent's performance can be measured by some criterion of performance ν. The criterion ν may be thought of as a criterion of partial success, optimization function, index of satisfaction, utility function, etc. The user determines the criterion of performance ν. For the same task, different users may use different performance criteria.

With every agent **i** we can associate at least two performance measures[1]. The first measure ν^i is agent **i**'s performance in the case where it acts autonomously, i.e., without the user's supervision. The second measure ν^i_U is agent **i**'s performance with the user's supervision. ν^i_U does not measure the performance of the user and the agent collectively. Intuitively, this is the agent's own performance with the user's supervision. However, we make no assumptions about improved performance and, in fact, performance degradation is quite possible. 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 ν^i_U and ν^i.

Definition 1. By a *degree of individual autonomy* A_i (autonomy with respect to the user) we mean the ratio $\frac{\nu^i}{\nu^i_U}$.

The degree of individual autonomy indicates the extent to which an agent may act well independently of the user i.e., what part of an agent's performance must be attributed only to the agent's capabilities. In general, individual autonomy varies between $-\infty$ and $+\infty$. The degree of individual autonomy can be interpreted as the degree of independence from the user's supervision. The combined user and agent performance is not necessarily the maximal performance. For example, if the user is not competent enough, the agent may be more efficient by acting autonomously.

Definition 1 characterizes autonomy as 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, autonomy estimates may vary across different users. This means that different users could consider a pattern of behavior as either autonomous or non-autonomous. Suppose, for example, that an agent autonomously fulfills only 90% of a given task. A user may consider a 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 regard the agent as autonomous.

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

3. A Measure of Autonomy in the Context of Environment Interaction

In this section we analyze an agent's autonomy with respect to a set of environmental factors, which may contain uncertainty or unreliability. These environmental factors might be tools, instruments, electro-mechanical devices, or perishable resources. Let's imagine the environmental factors have a known probability of reliability or uncertainty. In general, device reliabilities are represented as percentages over reliability ranges. For example, a light bulb might be 90% of the time 99% reliable and 10% of the time unreliable. This can be extended to several ranges, say 90% of the time 95% reliable (i.e., fairly reliable), 5% of the time 99% reliable (i.e., highly reliable), and %5 of the time 10% reliable (i.e., unreliable). The probabilities add up to 1.0 but the ranges are open. After access to the knowledge of these probabilities of reliability, the agent may or may not decide to use the environmental factors or make a decision about the environmental factors. Let's imagine n ranges each with α_i probabilities. The agent's ability to act and decide is contrasted in each range of reliability of the environmental factors in light of the known probabilities. ν^{i1} is the agent **i**'s performance in the case where it acts without the use of a set of environmental factors knowing that they are the most reliable (i.e., the best). ν^{i2} is the agent **i**'s performance in the case where it acts without the use of a set of environmental factors knowing that it may have access to second best environmental factors. Continue this until ν^{in} where the agent **i**'s performance in the case where it acts without the use of a set of environmental factors knowing that it is has the least reliable (i.e, the worst) environmental factors. ν_{i1} is agent **i**'s performance with the use of the most reliable (i.e., the best) set of environmental factors. ν_{i2} is agent **i**'s performance with the use of the second best reliable set of environmental factors. Continue this until ν_{in} where the agent **i**'s performance with the use of the least reliable (i.e., the worst) set of environmental factors. We make no assumptions about improved performance and in fact performance degradation is quite possible.

Definition 2. By a *degree of t autonomy* with respect to an *unreliable* environmental element we mean the ratio $\sum_r \alpha_r v^{ir} \ / \ \sum_r \alpha_r v_{ir}$.

The degree of environment autonomy indicates the extent to which an agent may act well independently of the environmental factors i.e., what part of an agent's performance must be attributed only to the agent's capabilities. In general, environment autonomy varies between $-\infty$ and $+\infty$.

Environmental uncertainty implies that an agent has to decide between relying and not relying on the environmental factors. The choice has to be made in complete ignorance about the actual level of reliability. If an agent chooses to rely, its expected performance will be $\sum_r \alpha_r v^{ir}$. If it does not rely, then the expected performance is $\sum_r \alpha_r v_{ir}$. The ratio between the expected performances measures agent's **i** autonomy with respect to an uncertain environment.

In Definition 2 we assume that an agent is free to decide whether to use the environmental factors. This implies that the agent has at least partial control over the environment. In many situations, however, an agent cannot go around the environment factors and have to use them. In this case we view an agent's autonomy as the ability to choose the most favorable environmental factors.

Definition 3. By a *degree of autonomy* with respect to an *uncertain* environmental element we mean the ratio $\sum_r \alpha_r v_{ir} \ / \ v_{i1}$.

According to definition 3 the degree of autonomy is the ratio between the average (or expected) performance $\sum_r \alpha_r v_{ir}$ and the most successful performance v_{i1}. Since the agent does not have control over the environmental factors, it cannot choose among them. If it had control, it would choose the most favorable ones.

Consider the following example. Both the user and the agent are uncertain about the current conditions in the environment. However, they share their levels of uncertainty as common knowledge. With probability 2/3 the conditions are favorable and with probability 1/3 they are inauspicious. The agent gets 6 if the conditions are favorable and 3 if they are inauspicious. The agent expected performance is 5 and the most desirable outcome is 6. Therefore, the agent's degree of autonomy is 5/6. The situation is shown in Fig.1

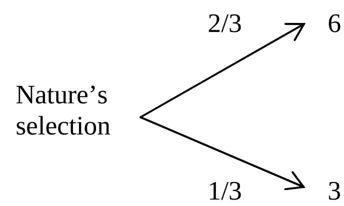

Fig.1: Individual autonomy in an uncertain environment.

4. 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. Indirect interaction usually occurs 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 may affect one another once they have been deployed in the environment. Therefore, it is not possible to devide the environment into different mutually independent groups of agents such that agents can affect one another if and only if they belong to the same group. 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. Then, the question is which agents to deploy? 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 devided into relatively independent groups called coalitions. Each coalition has its own performance measure (value of the coalition). The problem is to find a coalition partition that maximizes the sum of coalitions performances. In our case all agents perform the same task and there is no reason to separate them into different coalitions. That is, we assume that the grand coalition (involving all active agents) always forms. In other words, the user always deploys one coalition of agents. The problem is which coalition to deploy.

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 performance[2] (ν^i, ν^i_j, ν^i_k, ν^i_{jk}). Here ν^i represents agent **i**'s performance acting alone i.e., by acting autonomously. ν^i_j 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. ν^i_j could be greater or smaller than ν^i depending on the type of interference. For example, if agent **i** depends positively on agent **j**, then $\nu^i_j \geq \nu^i$. ν^i_k is agent **i**'s performance in the presence of agent **k**. ν^i_{jk} 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.

The elements of the relative performance vector should be interpreted as guaranteed performance values. For example, agent **i** can always get ν^i_j 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 ν^i_j. In other words ν^i_j is the minimax performance that agent **i** can obtain in the company of agent **j** i.e., no matter how agent **j** behaves, agent **i** always gets at least ν^i_j.

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

Definition 4. The degree of agent **i**'s *autonomy with respect to agent* **j** is

$$A(i/j) = \frac{\nu^i_j}{\nu^i}$$

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 either produce positive or negative performance. The following dilemma states the problem with the interference.

Definition 5: *Autonomy–efficiency dilemma* arises when we compose a group of agents subject to the highest overall group performance with two classes of agents: (a) agents with low efficiency and high autonomy invariance agents (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. These 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(i/j), 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** is autonomous with respect to agent **i**. Agent **j**'s performance does not depend on agent **i** and it is always 2. If we are looking for maximum invariance in autonomies, then we have to deploy only agent **j**. This, however, is not an efficient solution since agent **j** has low performance. Completely autonomous, agent **j** is not as efficient as agent **i** is. 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 autonomous and vise versa; agents with autonomies unaffected by other agents may not be very efficient. To alleviate this dilemma, we suggest the following assumption that all agents have the same individual autonomy. The following definition gives this assumption a name.

---

[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.

Definition 6: *Equally-competent agents* is the assumption that all agents have the same individual performance. That is, $\nu^i=\nu$, for all agents **i**.

Under equally-competent agents assumption, each agent **i** by acting alone can achieve the same standard of performance ν. 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. In this paper by a group of agents we mean any set of agents that act concurrently.

Definition 7. The degree of agent **i**'s *autonomy with respect to a group* of agents (j,k) is:

$$A(i/jk) = \frac{\nu^i_{jk}}{\nu}$$

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 for a synergetic interaction. Consider the following example. Suppose that agent **i**'s vector of relative performance is (4,6,4,8). That is, $\nu=4$, $\nu^i_j=6$, $\nu^i_k=4$, and $\nu^i_{jk}=8$. This implies that agent **i** depends positively on agent **j** (A(i/j) = 6/4 = 1.5). At the same time it is autonomous with respect to agent **k** (A(i/k) = 4/4 = 1), and depends positively on the group of agents **j** and **k** (A(i/kj) = 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 8. The degree of *group autonomy* of the group of agents (i,j) under the equally-competent agents assumption is:

$$A(ij) = \frac{\nu^i_j + \nu^j_i}{\nu}$$

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 $\nu^i_j+\nu^j_i$. If only one agent (either one) is deployed, the performance is ν.

Proposition 1. Group autonomy under the equally-competent agents assumption equals the sum of individual autonomies. That is,

$$A(S) = \sum_{i \in S} A(i/S-\{i\})$$

Where S is a set of agents, and S-{i} is the set of agents S excluding agent **i**.

Proof. Follows immediately from Definitions 7 and 8.

If we apply Proposition 1 to the group of agents (i,j,k), we will get

$$A(ijk)=A(i/jk)+A(j/ki)+A(k/ij)$$

It is worth noting that Proposition 1 does not hold in general. The proposition depends 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 9. The degree of *autonomy of the group* of agents **i** and **k** with respect to agent **k** under the equally-competent agents assumption is:

$$A(ij/k) = \frac{\nu^i_{jk} + \nu^j_{ik}}{\nu^i_j + \nu^j_i}$$

The numerator in Definition 5 measures the group performance of agents **i** and **j** in the company of agent **k**. The denominator is the autonomous performance of the group. If we apply Proposition 1 to Definition 9, we will obtain the following proposition.

Proposition 2. $A(S/k) = \frac{\sum_{i \in S} A(i/S-\{i\}+\{k\})}{A(S)}$

where S-{i}+{k} is the group S excluding agent **i** and including agent **k**.

Proposition 2 says that the relative group autonomy (with respect to a third agent) depends positively on relative individual autonomies A(i/S-{i}+{k}) and negatively on the group autonomy A(S).

To illustrate all these notions, consider the following example. Let's assume we can deploy up to three agents **i**, **j** and **k** with the following vectors of relative performance:

- i's relative performance (ν^i, ν^i_j, ν^i_k, ν^i_{jk}) = (4, 3, 4, 3)
- j's relative performance (ν^j, ν^j_i, ν^j_k, ν^j_{ik}) = (4, 4, 2, 1)
- k's relative performance (ν^k, ν^k_i, ν^k_j, ν^k_{ij}) = (4, 4, 5, 5)

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.

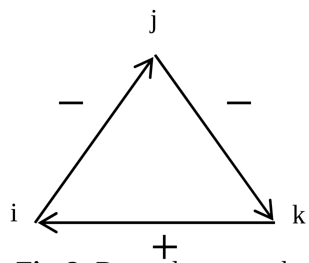

Fig. 2: Dependence graph.

This example shows that since A(ik/j)= (ν^i_{jk} + ν^k_{ij}) / (ν^i_k + ν^k_i) = (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(i/k) + A(k/i) = 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 A(ijk) = A(i/jk) + A(j/ik) + A(k/ij) = 3/4 + 1/4 + 5/4 = 2.25 is achieved when all agents are brought together. This is not apparent from the initial statement of the problem, since agent **j** relates negatively to agents **i** and **k**.

The problem of finding the group with maximum autonomy is of significant importance for multiagent interaction. Whenever a group of agents are deployed for solving a particular task, we have to know which group of agents has the maximum autonomy. Along the same line of reasoning, if an agent decides to share or delegate its task to other agents, it has to find the group with the most desirable autonomy. A related issue is finding a group of agents with minimum variance in their autonomy. This is important for fault tolerance reasons since if agents were allowed to come and go at will, we would not want the group's performance to be significantly affected. We have looked at this problem previously [Hexmoor, 2000b]. However, this is a different problem than seeking agents with maximum autonomy. 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 vectors, 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 at 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 ν^i. Let the relative performance vector of agent **i** be (1, ν^i, ν^i, ν^i,….). That is, agent **i**'s individual autonomy is 1 and its relative performance is always ν^i. Now, the elements of a set of numbers S' sum to K if and only if the set of agents S' that has a group autonomy K. Thus, our problem is NP-hard.

4. Conclusions

In this paper we introduced several quantitative measures of relative autonomy. The first measure defines individual autonomy with respect to user-agent interaction. The second measure relates to autonomy with respect to environmental factors. The third measure defines autonomy among groups and individuals. Our measures are domain independent and do not rely on specific interaction protocols. We also analyzed the question of finding a multiagent group with the maximum autonomy. We proved that this problem is NP-complete. Therefore, the problem of finding the optimal group or agent with whom to share a task (or to whom to delegate a task) is computationally hard in general. This suggests development of approximation algorithms for measuring and adjusting autonomy. Our future work includes looking into the relationship between maximum group autonomy and least variance in group autonomy.

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, Spinger-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.

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.