IEEE Transactions on Man, Machine, and cybernetics- Part C (SMC-C), Vol. 31, No. 4, Pages 509-517, November 2001

Stages of Autonomy Determination

Henry Hexmoor

Abstract-- **We discuss stages of autonomy determination for software agents that manage and manipulate knowledge in organizations that house other software agents and human knowledge workers. We suggest recognition of potential autonomies in belief, desire, and intention paradigm and actual reasoning about autonomy choices decision theoretically. We show how agents might revise their autonomies in light of one another's autonomy and might also experience new, derived autonomies. We discuss the conditions under which an entire group of agents might have a collective autonomy attitude toward agents outside their group. We believe group attitudes are a novel concept and form a strong basis for developing theories of dynamic organizational structure. We will briefly sketch outline of a case study that motivates reasoning about autonomies.**

Index Terms—**multiagent systems, autonomy**

I. INTRODUCTION

Knowledge management (KM) focuses on the processes for promoting, growing, communicating, and preserving knowledge [16, 19]. In organizations, it is desirable that KM works with distributed knowledge and without much intervention [20, 25]. A technological development that is being used in modeling, managing, and using distributed knowledge in an organization is agent-oriented programming [14] and increasingly knowledge workers use with agents. In this paradigm, agents model processes that manipulate knowledge. For independent and automated interaction, agents that process knowledge require abilities to reason about their autonomy. We are designing agents that can reason about their own autonomies and reason about others autonomies. This autonomy determination might lead agents to crucial decision about collaboration: autonomous action, delegation, or collaboration.

Autonomy is defined and used in multiagency and other disciplines [6, 7, 10, 12], sociology [9], and philosophy [17, 18, 23]. Autonomy is important in multiagent interaction since it relates abilities in a self or a group to the individual's freedoms and choices. Agent centered understanding of autonomy is required for coherent interagent interaction. The notion of autonomy has been used in a variety of senses and has been studied in different contexts. The concept of autonomy is closely related to the concepts of power, control and dependence [5, 7].

Computer Science & Computer Engineering Department, Engineering Hall, Room 313, Fayetteville, AR 72701 Email: Hexmoor@uark.edu

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. Biologically, it is said that an organism's ability to for self-organization and handling perturbations is material autonomy. Beyond that, it is argued that living organisms possess the ability for stable integration of self-reference and other-reference, known as syntactic autonomy [25]. This syntactic view of autonomy supports the relativistic sense of autonomy we are pursuing. Autonomy is defined in [6] as the agent's degree to which its decisions depend on external sources including other agents. This can be called a cognitive autonomy. This has been explored further in [7]. This work also promotes the relativistic view of autonomy we have developed in [4]. It is possible to differentiate autonomy into dynamic and deterministic types. Dynamic autonomy might capture the agent's initiate and self-start whereas deterministic autonomy might capture the agent's ability to refrain from actions it can perform.

Adjustable autonomy is a related notion that captures the idea of a human operator intervening and guiding actions of a machine [8]. Another example of the work on adjustable autonomy is [2]. A quantitative measure of agent autonomy is proposed in [1]. 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. Our own elsewhere introduces another measure of relative autonomy [4].

In this paper we take an agent's autonomy as a relative sense of its individual preference over the intender or desirer of goals over which it has nontrivial abilities. This preference is social in that it includes consideration of other agents, how they contribute to the agent's sense of freedom to choose and performance and how it prefers to work with others. The upshot of an agent's autonomy consideration will lead it to be self-directed, other directed (as in delegation), shared with other agents (as in teaming), or be partially self-directed [10]. Many parameters, both endogenous and exogenous may go into a utilitarian quantification of these preferences [11]. As a more complex preference it may take into account other agent's utilitarian quantification of their preferences and so it may become a social preference (see [5] for social preferences). Parameters for such preferences in complex agents are hard to enumerate and they change over time with the agent's experience. Only in the most circumscribed situations we can apply normative decision theory to autonomy determination [11]. Although we believe designing parameters is a task that has to be done in each domain, we will show that there are useful domain-independent influences among agent autonomies. Agents may have to revise their autonomies in light of other agent's autonomy and might be provoked to establish new autonomy decisions.

Previously, we attempted to decompose autonomy in terms of Belief-Desire-Intention (BDI) paradigm [10]. This approach is useful in that it allows agents to reason over BDI and to relate autonomy to action selection. We will elaborate on autonomy potentials perceived by the agent. Our formulation will be via a BDI logic a la [24].

For a given goal, an agent's autonomy determination and use can be divided into three stages. The first stage is *potential determination* in which the agent considers its beliefs, desires, and intentions and produces a potential to elect as executor of goal, self only, other agent only, shared with other agents, or partially self. The first stage is where our BDI formulation is most useful. The second stage is *utility analysis* in which the agent weights potential alternatives and decides on one. The second stage is where our decision theoretic formulation will be used. The third stage is *enactment* where an agent will complete its action selection taking into account autonomy selection. Agents in the real world will spend different amounts of time and cognitive resources on each stage. What determines the agent's allocation of time and resources is an interesting area of study beyond the scope of this paper. Our formulation on stages 1 and 2 are useful for our exposition. However, any implementation of our stages in a computer program can take these in spirit and there is no need to take these literally.

We perceive the most significant contribution of this work to be elaboration of steps involved in explicit autonomy determination and adjustment. Explicit reasoning about autonomy is needed in many knowledge-rich and dynamic environments where automated groups of agents have to reason about collaborative and collective work as in team formation. We have identified autonomy as a central feature of team formation [3]. Near the end of this paper we will discuss a real-world case study of a emergency response system to flashfloods that outlines uses of autonomy reasoning. There are many other areas that can benefit from this work. One of those areas is e-commerce and cross-enterprise agent negotiation. [15].

The paper is organized as follows. In the next section we present the first stage of autonomy determination in terms of Belief, Desire, Intention BDI paradigm of multiagency. This preliminary stage is to determine potential autonomies. We then explore the second stage where the agent explores arbitration among its potential autonomies. This is followed by a brief outline of a case study in flood management that illustrates reasoning about autonomies. We conclude with remarks about relationships between the stages.

II. BDI FORMULATION OF AUTONMY

Here we will explore a BDI model of autonomy. Let's begin by agent abilities. We first borrow the idea of a single agent first-hand ability and a group's first-hand joint ability from [24] in the following definition.

Definition 2.1 Can (see [24] for details):
$(Can^0\ i\ \psi) \equiv \exists \alpha.\ (Bel\ i\ (Agt\ \alpha\ i) \wedge (Achvs\ \alpha\ \psi)) \wedge (Agt\ \alpha\ i) \wedge (Achvs\ \alpha\ \psi)$ (P. 150)), that is, agent i has a first order ability to achieve the state ψ iff there is an action α that i knows i can perform and α is guaranteed to achieve ψ. This leads to a definition for higher order ability, viz.
$(Can^k\ i\ \psi) \equiv (Can^{k-1}\ i\ (Can^0\ i\ \psi))\ k > 0$ (P. 152)
and finally a joint ability among agents. The following states that a groupf agents have a first ability to achieve ψ iff there is an action α that the group g knows g can perform and α is guaranteed to achieve ψ.
$(J\text{-}Can^0\ g\ \psi) \equiv \exists \alpha\ (M\text{-}Bel\ g\ (Agts\ \alpha\ g) \wedge (Achvs\ \alpha\ \psi)) \wedge (Agts\ \alpha\ g) \wedge (Achvs\ \alpha\ \psi)$ (P. 153)

Here α is a rigid designator. This means that an agent has the ability to achieve some state ψ if it knows of an action that it can perform which it knows will achieve the state of affairs. "Can" stands for the ability of a single agent, whereas "J-Can" stands for the joint ability of a group of agents. As in the above requirement for the action designator *belief* needs to be explicated with respect to artificial agents. When agents believe in J-Can, the agents individually have a belief that they collectively have the ability to possibly achieve the goal. It does not imply that they, in fact, have that ability.

An agent that has the ability to with regards to state ψ may also experience *liberty* to do anything to achieve it. We define *free* as permission from its social group for doing actions to achieve the state. Obviously, there are many variations to who will permit what action to whom. But we will state only two with group consensus. Definition 2.2a captures the notion of total freedom over any action whereas definition 2.2b captures freedom over some action.

Definition 2.2a Total Freedom:
An agent i has total freedom with respect to state ψ in its social group g iff everyone in its social group g permits it to perform every action that can achieve ψ.
$(TFree\ i\ \ g\ \psi) \equiv \forall j \in g - \{i\}\ \forall \alpha\ (Agt\ \alpha\ i) \wedge$
$(Achvs\ \alpha\ \psi) \Rightarrow (Per\ j\ i\ \alpha)$

The predicate "Per" stands for the deontic notion of Permission. We intend to capture social obligations with "Per" of one agent giving another agent permission for an action. (See [22] for typical usage of permissions)

Definition 2.2b Limited Freedom:
An agent i has limited freedom with respect to state ψ in its social group g iff everyone in its social group g permits it to perform an action that can achieve ψ.
$(LFree\ i\ \ g\ \psi) \equiv \exists \alpha\ \forall j \in g\text{-}\ \{i\}\ (Agt\ \alpha\ i) \wedge$
$(Achvs\ \alpha\ \psi) \wedge (Per\ j\ i\ \alpha)$

It is obvious that limited freedom is a subset of total freedom.

Corollary 2.1:
If an agent's social group grants an agent total freedom over a state ψ, then it also gives it limited freedom.

We are now in a position to define potentials for self-directed autonomy. Potentials are determined at this stage prior to determining the actual autonomies, which will be done in the second stage of autonomy determination. The self-directed potential autonomy preference is that the agent elects itself to be the executor.

Definition 2.3 An agent i is potentially self-autonomous with respect to ψ in its social group g iff it can do ψ and has limited freedom from its social group.
(Pself-autonomous i g ψ) ≡ (Can i ψ) ∧
(LFree i g ψ)

Definition 2.3 uses limited freedom to define potential self-autonomy. If used total freedom, we would still produce potential for self-autonomy and nothing more. This is because we are focusing on relative (interagent) autonomies. An agent with more choices might feel more autonomous. Furthermore, an agent who ranks its choices, might seek to increase its highly ranked choices and decrease its low marked choices. Such an agent may seek company of agents that influence its number and quality of its choices. This is a social phenomena and has been studies in social networks theory. We feel that an analysis of choices and perceived autonomies based on choices is a basis for developing an absolute measure of personal autonomies. But we will leave that to another paper and here we will continue to focus on the relative sense of autonomy.

In order to talk about delegation, we need to define granting freedom. If an agent gives another agent permission to perform at least one its actions that achieves a state ψ, it has granted that agent some freedom. We will not distinguish between levels of freedom from total freedom and limited freedom.

Definition 2.4 Granting Freedom:
An agent i grants freedom to agent with respect to state ψ iff whenever agent i believes there is an action by agent j that can achieve state ψ, agent i permits agent j to perform that action.
(GFree i j ψ) ≡ ∃ α (Bel i (Agt α j) ∧
(Achvs α ψ)) ⇒ (Per i j α))

Lfree and Gfree are not reflexive in that agents do not have or grant self-freedom. This type of self-freedom is beyond the scope of this paper. LFree and GFree are not symmetric, which means the agents do not necessarily reciprocate limited freedom and ranting freedom. LFree and GFree are not transitive in that if agent 1 has an attitude toward agent 2 and agent 2 has a similar attitude toward agent 3, it does not follow that agent 1 has a similar attitude toward agent 3.

An agent i may prefer to delegate ψ to another agent j. Agent i would have to believes that agents j can do ψ and furthermore desire that agent j desire it and allow that agent some freedom.

Definition 2.5 An agent i is potentially delegation-autonomous toward agent j with respect to ψ iff it believes that agent j can do ψ, it desires that agent j adopt the desire to achieve ψ, and grants it some freedom to consider achieving ψ.
(Pdel-autonomous i j ψ) ≡ (Bel i (Can j ψ)) ∧
(Des i (Des j ψ)) ∧ (GFree i j ψ)

An agent i may prefer to share its autonomy with another agent j. In that case, agents i would have to believe that together with agent j, they can bring about state ψ, grant some freedom to agent j over that state, and desire that agent j desire that state. This leads to a definition for potential for sharing autonomies from agents i towards agent j.

Definition 2.6. An agent i is potentially shared-autonomous with agent j with respect to ψ iff it believes that they (g will stand for the group of i and j) can jointly do ψ, i desires that j desire ψ, and that i grants freedom to j.
(Pshared-autonomous i j ψ) ≡ (J-Can^0 g ψ) ∧
(Des i (Des j ψ)) ∧ (GFree i j ψ)

An agent i may potentially have a partial self-directed autonomy. Agent i will have less than a first hand ability to bring it about the state but will have earned some freedom from someone in its social group.

Definition 2.7. Agent i who is part of a social group g is potentially partially self-directed with respect to ψ iff it has a less than first-hand ability about doing it and there is someone in its social group g who has granted it some freedom over state ψ.
(Ppartial-self-autonomous i g ψ) ≡
(Can^k i ψ) ∧ ∃ j ∈ g – {i} (GFree j i ψ) where $k > 0$

Whereas potential autonomies are not transitive, we will show in the next section that there is influence among actual autonomies, which results in derived autonomies.

Proposition 2.1 Potential shared autonomy and potential delegation autonomy are not transitive.

Proof. *LFree* and *Gfree* are not transitive.

We are now in a position to consider potential group autonomies. We will start by a joint autonomy between two agents that have potential shared autonomy toward one another.

Definition 2.8. A group of agents a_1, a_2, …, a_n who are reciprocally potentially shared-autonomous towards one another with respect to ψ are considered to have a potentially **joint shared autonomous**.

(Pjoint-shared-autonomous g ψ) $\equiv \forall a_i, a_j \in$ g (Pshared-autonomous a_i a_j ψ) $\wedge$ (Pshared-autonomous a_j a_i ψ)

Agents with joint autonomy could join group activity. For instance, they could enter a team if they further have a cooperative stance and come to have a joint intention and awareness [3]. For teams of agents, in addition to shared autonomy, we require agents to have a joint cooperative attitude [21]. We posit that agents must adopt the principle of social rationality in order to be cooperative [13]. Under this principle, agents who are part of a group will prefer actions with the property that the joint benefit to the group is larger than its joint loss. Further teaming discussion is beyond the scope of this paper.

A weaker form of potential autonomy is when agents are in a closed chain of potential shared autonomy links without direct reciprocity. E.g., if agent i is potentially shared autonomous towards agent j, agent j is potentially shared autonomous towards agent k, and agent k is potentially shared autonomous towards agent i, then agents i, j, and k are associated by the potential sharing of autonomies.

Definition 2.9. A group of three or more agents a_1, a_2, …, a_n are potentially **shared autonomy friend** with respect to ψ iff every agent has another agent who has a potential shared autonomy toward it with respect to ψ.

(Pfriend-shared-autonomous g ψ) $\equiv \forall a_i \in g \; \exists a_j \in g$ (Pshared-autonomous a_j a_i ψ)

Similar to shared autonomies, agents in a strong delegation relationships might want to delegate their task to an agent outside their group. A group of agents who are trying to delegate the task to any other group member share a joint attitude for delegation. The following is a definition for this notion.

Definition 2.10. A group of agents a_1, a_2, …, a_n who are reciprocally potentially delegation-autonomous towards one another with respect to ψ are considered to have a potentially **joint delegation autonomous**.

(Pjoint-delegation-autonomous g ψ) $\equiv \forall a_i, a_j \in$ g (Pdelegation-autonomous a_i a_j ψ) $\wedge$ (Pdelegation-autonomous a_j a_i ψ)

Similar to shared autonomies, agents in delegation autonomies will have a friendship of potential delegations if everyone in the group is oriented in delegating the task to someone else.

Definition 2.11 A group of three or more agents a_1, a_2, …, a_n are potentially **delegation autonomy friend** with respect to ψ iff every agent has another agent who has a potential delegation autonomy toward it with respect to ψ.

(Pfriend-del-autonomous g ψ) $\equiv \forall a_i \in g \; \exists a_j \in g - \{a_i\}$ (Pdel-autonomous a_j a_i ψ)

Thus far we have discussed potential autonomies determined in the first stage of autonomy consideration. After stage 1, the agent will consider utilities that will be discussed in section 3 and then convert from potential to an actual autonomy.

III. DECISION THEORETIC FORMULATION OF AUTONOMY

We now turn to the second stage of autonomy determination where potential autonomies are further considered. At this stage, we will explore a utilitarian model of autonomy. Let X_i be parameter(s) that affect an agent i's autonomy. Intuitively, an agent's level of autonomy may be different when it is alone and when it is in the company of others or in a different environment with varying levels of resources. This change can be measured as a second hand effect in the agent's change in performance or it can be measured first-hand in its experience of relative quantity or quality of its choices. We have looked at some of these measures elsewhere [4]. With such computations, an agent may have access to a continuum of autonomy options. Say, an agent may be inclined to work with other agents or environments at dynamically varying levels. An agent may prefer to be in the company of agents or environments that enhance the quality and quantity of its choices. However, such a discussion and relativistic analysis is beyond the scope of this paper. Here, we will offer an account of how an agent arrives at a determination of autonomy preference. We will assume the agent is able to compute utilities for a simple set of choices. UI_i is the utility function of agent i for self-directedness. UD_{ij} is the utility function of agent i for delegation toward agent j. US_{ij} is the utility function of agent i sharing its autonomy with agent j. UP_i is the utility function of agent i for partial self-directedness.

To elaborate the utilities, each is a function of a goal (G), parameters Xi that affect its autonomy, and utilities of other agents that the agent i cares to consider. U_k is the maximum utility value for agent k from the perspective if agent i. I.e., $UI_i = UI_i(G, X_i, U_1, U_2, \ldots, U_k)$

The difference between UI and UP is responsibility. With UI, the agent assumes a high level of responsibility whereas with UP, the agent assumes a low level of responsibility. The next four definitions are used for arbitration of autonomies to be exclusively one of UI, UD, US, or UP.

Definition 3.1. An agent i's autonomy is solely self-directed iff:

- $UI_i > UD_{ij}$ where j = all agents agent i considers for delegation, and
- $UI_i > US_{ij}$ where j = all agents agent i considers for sharing, and
- $UI_i > UP_i$

Definition 3.2. An agent i's autonomy is solely delegation-directed toward agent j iff:

$UD_{ij} > UI_i$ where j = all agents agent i considers for delegation, and
$UD_{ij} >= UD_{ik}$ where k = all agents agent i considers for delegation, (i.e., agent j is the most appropriate agent for delegation), and
$UD_{ij} > US_{ik}$ where k = all agents agent i considers for sharing, and
$UD_{ij} > UP_i$.

Definition 3.3. An agent i's autonomy is solely oriented toward sharing it with agent j iff:

- $US_{ij} > UI_i$ and
- $US_{ij} > UD_{ik}$ where k = all agents agent i considers for delegation, and
- $US_{ij} > US_{ik}$ where k = all agents agent i considers for sharing, (i.e., agent j is the most appropriate agent for sharing), and
- $US_{ij} > UP_i$.

Definition 3.4. An agent i's autonomy is solely partially self-directed iff:

- $UP_i > UI_i$ and
- $UP_i > UD_{ij}$ where j = all agents agent i considers for delegation, and
- $UP_i > US_{ij}$ where j = all agents agent i considers for sharing.

Subsequent to each agent's decision about its own autonomy, in light of other agents autonomy decisions, an agent may change its original decision. The next section discusses impact of a second agent's decision on the first agent's revision.

A. Revision of Autonomies Between Agents

With only two agents and 4 autonomy types, there are $2^4 = 16$ cases. Here, we only point out the most plausible likelihood for revision and not be exhaustive. The agent's actual revised autonomy is subject to the situated parameters available to it. Our analysis points out natural tendencies that exist for revision. This analysis is useful in predicting agent behavior. Let's make the following assumptions:

- Both agents already consider a common goal.
- Executing a goal once is enough.
- We will not consider goal adoption or persuasion of any sort by the agents.
- We consider the first agent has decided on its autonomy, learns of the second agent's autonomy decision, and considers revision of its original decision.
- Agents are self-interested. We will not consider competition, altruism, or malevolence.
- When there is a autonomy orientation between two agents, both are aware of it.
- Sharing responsibility toward goals is a separate issue and not considered here.
- Agents have identical autonomy parameters.
- Agents have identical capabilities.
- Agents are at peer level for our considerations.
- Roles and authorities will be omitted from our consideration.

In the following list of cases, the autonomies are formed in a temporal sequence. Agent 2's autonomy is known after agent 1's and the statements are about what agent 1 will do for revision.
UI_1, UI_2 will result in no change.
UI_1, UD_{21}will result in no change unless there is some sort of social transaction.
UI_1, US_{21} may provoke a revision to share. This may also involve politeness or other social relationships.
UI_1, UP_2 will result in no change.
UD_{12}, UI_2 Agent 1 will consider the goal taken care of by agent 2. This may provoke a revision to a partial self-directedness.
UD_{12}, UD_{21} A deadlock exists. This may provoke a revision to avoid the deadlock such as UP1.
UD_{12}, US_{21} may provoke a revision to share or a partial self-directedness.
UD_{12}, UP_2 may provoke a revision to a partial self-directedness.
US_{12}, UI_2 may provoke a revision to a partial self-directedness.
US_{12}, UD_{21} may provoke a revision to a partial self-directedness.
US_{12}, US_{21} will result in no change and can be the basis of a coalition or teaming.
US_{12}, UP_2 may provoke a revision to a partial self-directedness.
UP_1, UI_2 will result in no change.
UP_1, UD_{21} may provoke a revision to increase autonomy self-directedness.
UP_1, US_{21} may provoke a revision to increase autonomy self-directedness.
UP_1, UP_2 will result in no change.

Among three agents, once the first and second agents have determined their own autonomies, the first agent might have an indirect autonomy toward the third agent. The next section discusses the types of indirect autonomies.

B. Derived Autonomies

Let's alter the assumptions in section 3.1 by assuming that there are 3 agents. Furthermore, we assume agent 1, having seen agent's 2's autonomy with respect to agent 3, is likely to have an indirect autonomy. We'll call this a derived autonomy toward agent 3.

In the following list of cases, the autonomies of the first two agents are determined and the statements are about what agent 1 will experience toward agent 3 as a derived autonomy. Here are the cases:

- UI_1, UI_2 will not result in no change.
- UI_1, UD_{23} may provoke a derived autonomy for agent 1 to share with agent 3.

- UI_1, US_{23} may provoke agent 1 to have a derived shared autonomy with agent 3.
- UI_1, UP_2 will result in no change.
- UD_{12}, UI_2 will result in a derived autonomy for agent 1 for delegation to agent 3.
- UD_{12}, UD_{23} will result in a derived autonomy for agent 1 for delegation to agent 3.
- UD_{12}, US_{23} will result in a derived autonomy for agent 1 for delegation to agent 3.
- UD_{12}, UP_2 may provoke a derived autonomy for agent 1 for delegation to agent 3.
- US_{12}, UI_2 may provoke a derived autonomy for agent 1 for sharing with agent 3.
- US_{12}, UD_{23} will result in a derived autonomy for agent 1 for sharing with agent 3.
- US_{12}, US_{23} will result in a derived autonomy for agent 1 for sharing with agent 3.
- US_{12}, UP_2 may provoke a derived autonomy for agent 1 for sharing with agent 3.
- UP_1, UI_2 will result in no change.
- UP_1, UD_{23} will result in no change.
- UP_1, US_{23} will result in no change.
- UP_1, UP_2 will result in no change.

Let's first state a definition and then state a few of our intuitions in more precise terms of propositions.

Definition 3.5. If an agent i's shared or delegated autonomy toward another agent j is invariant to agent j's sharing or delegation autonomy toward a yet another agent k, then agent i's shared or delegation autonomy is said to have an *independence property* with respect to agent j and agent k.

The property of independence for shared autonomy guarantees that an agent's orientation toward shared or delegated autonomy is not affected by the second agent's subsequent sharing or delegation autonomy toward agent k. I.e., agent i treats agents j and k the same with respect to its autonomy. In the propositions 3.1 and 3.2 we state the idea of derived shared autonomy as a new, indirect autonomy between agents.

Proposition 3.1. If an agent i wishes to orient its autonomy toward sharing with agent j and agent j wishes to orient its autonomy toward sharing with agent k or to delegate to agent k, in absence of a delegation autonomy between agent i and agent k (in either direction), and as long as agent i's shared autonomy has the independence property, agent i would have a derived (i.e, new) autonomy to orient toward sharing with agent k.

Proof. We assume sharing with agent j yields the most utility for i, and by the independence property, agent i's utility is the same with respect to its autonomy orientation toward agents j and k. Then if it has already decided to have a sharing autonomy toward agent j, it can exchange j for k and have a sharing autonomy toward agent k.

Proposition 3.1 states that by sharing autonomies is transitive under independence property. Proposition 3.1 does not have the power to cancel delegation orientations. Furthermore, we make no assumptions about whether the third agent is aware of the first agent's shared autonomy orientation toward it.

We extend the notion of derived autonomy from 3 agents scenario to several agents.

Proposition 3.2. Let an agent 1 wish to share its autonomy with agent 2 with independence property and that agent 2 wishes for sharing or delegation autonomy with agent 3, agent 3 wish for sharing or delegation with agent 4, and so on until agent n. Assume absence of delegation autonomy between agent 1 and all other agents (in either direction). Then we assert that agents 1 would have a derived autonomy for sharing with agents 2, 3, …, n.

Proof. Using induction on Proposition 3.1.

The independence property is not transitive as stated in the following corollary. In propositions 3.1 and 3.2 it was not necessary for autonomies of 2^{nd} and subsequent agents to have this property. Furthermore, the derived autonomy does not possess this property.

Corollary 3.1: The independence property is not transmitted to derived autonomies.

Proof. When an agent 1 has an autonomy with independence property toward another agent 2, its utilities are unaffected by the second agent's autonomy orientation. However, when a derived autonomy is established between 1 toward another agent 3, there is no guarantee that agent 1's utilities will be unaffected by agent 3's autonomy choice since it was not taken into consideration by agent 1.

Corollary 3.2. If an agent wishes to share its autonomy with agent 2 with independence property and that agent wishes for sharing autonomy with agent 3, agent 3 wishes for sharing with agent 4, and so on until agent n, agents 1 would wish for share autonomy with every other agent 2, …, n.

Proof. This is a special case of proposition 3.2

In the propositions 3.1 and 3.2 we state the idea of derived delegation autonomy as a new, indirect autonomy between agents. These are very similar to the shared autonomy propositions.

Proposition 3.3. If an agent i wishes to orient its autonomy toward delegation to agent j and agent j wishes to orient its autonomy toward sharing with agent k or to delegate to agent k, in absence of a sharing autonomy between agent i and agent k (in either direction), and as long as agent i's delegation autonomy has the independence property, agent i would wish to orient its autonomy toward delegation to agent k.

Proof. Similar to Proof of proposition 3.1.

Proposition 3.4. Let an agent 1 wish to have a delegation autonomy toward agent 2 with independence property and that agent wish for sharing or delegation autonomy with agent 3, agent 3 wish for sharing or delegation with agent 4, and so on until agent n. Assume absence of sharing autonomy between agent 1 and all other agents (in either direction). Then we assert that agents 1 would wish for delegation autonomy towards agents 2, 3, …, n.

Proof. Using induction on Proposition 3.3.

It is obvious that if two agents mutually hold a shared autonomy toward one another they would want to enter a coalition or a team. This can also easily extend to several agents. The coalition or a team would work best if all agents mutually shared their autonomy. To aid in formalization of coalition/team formation, we define shared autonomy friends agents in definition 3.6. This definition is used in [3]. This definition is similar to potential shared autonomies we presented in definition 2.8 but it differs in that here the agent is working with actual autonomy orientations.

Definition 3.6: A group of agents a_1, a_2, …, a_n are **jointly shared autonomous** iff they have reciprocal shared autonomy toward one another.

When shared autonomy is not reciprocal, they do not enjoy a joint autonomy. However, they might still be closely related. We define shared autonomy friend in definition 3.7, which recapitulates definition 2.9 but with actual autonomy orientations.

Definition 3.7: Three or more agents are **shared autonomy friend** iff they are part of at least one cycle of original shared autonomy links among them. We say “original” to contrast with derived links that used proposition 3.1.

Using definition 3.7, for two agents to be shared autonomy friend, both agents must orient its autonomy toward sharing with the other, I.e, they must reciprocate shared autonomy. With three or more agents, the reciprocity is not needed.

Similar to shared autonomies, agents in a strong delegation relationship might act as a group. We called it potential joint delegation autonomous in definition 2.10. Here we revisit this notion with actual autonomy orientations define joint delegation autonomous in definition 3.8.

Definition 3.8. A group of agents a_1, a_2, …, a_n are **jointly delegation autonomous** iff the have reciprocal delegation autonomy toward one another.

When delegation autonomy is not reciprocal, the autonomies might be non-uniform but they might still be closely related. We define delegation autonomy friend in definition 3.9, which recapitulates definition 2.11 but with actual autonomy orientations.

Definition 3.9: Three or more agents are **delegation autonomy friend** iff they are part of at least one cycle of original delegation autonomy links among them.

In the next section we will discuss a few properties of autonomies in a group that are used to construct a group autonomy attitude.

C. Toward Group Autonomy

Autonomy compatible agents might be less than reciprocal in their autonomy. To measure this asymmetry, we define shared harmony.

Definition 3.10: Shared harmony among n shared autonomy friend agents is the ratio (number of original shared autonomy orientations)/ (n * (n-1)).

We can easily see that in a shared autonomy friendship, the shard harmony will have a minimum value as stated in the following corollary.

Corollary 3.3: Shared harmony among shared autonomy friend agents is greater or equal 1/(n-1).

Proof. Since in a shared autonomy friendship, every agent will have at least one agent who is oriented to share its autonomy with the agent, there will be n such links and it follows from the definition of shared harmony that the ratio will be at least as large as 1/(n-1).

In order to construct a group autonomy attitude like delegation, we need to have define a property of strength in the group’s autonomy.

Definition 3.11: Shared (or delegation) autonomy friendship is **strong** iff the cycle of original shared (or delegation) autonomy links among them have the independence property.

When a group of agents have a strong friendship, their utilities are not affected by one another’s subsequent choice. This allows us to define attitude for the group as a whole. Before proceedings with defining group autonomy attitude, we make two state properties of strong friednship in the following four corollaries.

Corollary 3.4: With the independence property, agents in a strongly shared autonomy friendship have reciprocal shared autonomy. I.e., any two agents in the group have shared autonomy.

Proof. Using proposition 3.1.

Corollary 3.5: The graph of shared autonomy in a strongly shared autonomy friendship is strongly connected and complete, i.e., for every two agents there is a directed shared autonomy link between them.

In common sense terms, agents in a strongly shared autonomy friendship can treat one another as potential collaborators.

Corollary 3.6: With the independence property, agents in a strongly delegation autonomy friendship have reciprocal delegation autonomy. I.e., any two agents in the group have delegation autonomy.

Proof. Using proposition 3.1.

Corollary 3.7: The graph of delegation autonomy in a strong delegation autonomy friendship is strongly connected and complete, i.e., for every two agents there is a directed delegation autonomy link between them.

In common sense terms, agents in a strong delegation autonomy friendship can treat one another as potential subordinate. Being in such a group may appear useless. But we will see that this can be basis of group delegation autonomy.

Let's consider a structure where a number of agents are attempting delegation and one agent is found that assumes self-autonomy and thereby the delegation has been successfully passed on to an agent. We will call this a delegation chain in the following definition.

Definition 3.12: If agent 1 holds delegation autonomy toward agent 2, afterwards (temporally succeeding) agent 2 holds delegation autonomy toward agent 3, and so on until agent n, where agent n then holds a self-directed autonomy, we say that agents 1..n are in a **delegation chain**.

We make no assumptions about whether the terminal agent in a delegation chain is obeying a social code by forming its self-directed autonomy determination subsequent to delegating agent's autonomy determination. I.e., we do not assume that the agent is taking orders from the other agents. Therefore, other than delegation autonomy friendship as discussed earlier, the delegating agents do not form any sort of delegation affinity.

Now we are almost ready to suggest a derived group autonomy attitude for delegation. Let's consider a shared autonomy friendship group of agents. Intuitively, a group of agents in a shared autonomy friendship have the right attitude to be a team. What jeopardizes the team is if at least one of the agents gets involved in a delegation chain with agents outside the group friendship. For example, imagine 3 agents in a shared autonomy friendship, I.e, US_{12}, US_{23}, US_{31}. If either agent, say agent 1, finds a fourth agent toward whom it has delegation autonomy and the fourth agent has a self-directed autonomy (UD_{14}, UI_4), then the delegation chain (agents 1 and 4), spoil the potential for collaboration among agents 1, 2, and 3. Let's consider the 3 agents to have a strongly shared autonomy friendship, I.e., their shared autonomy has the independence property. Now agents 2 and 3 drive shared autonomy toward agent 4 whereas agent 1 has delegation autonomy toward agent 4. Since they do not agree on their autonomy orientation toward agent 4, they cannot act as a group toward agent 4. If the agents 1, 2, and 3 instead had delegation autonomy with independence property, they would have a group delegation. The following proposition states the derived group delegation autonomy.

Proposition 3.5: If a group of agents in a strong delegation autonomy friendship, also has a delegation chain where one of their agents in the group in involved in a delegation autonomy chain toward an agent outside the group and that agent subsequently has a self-directed autonomy, then we say that the group has a derived **group delegation autonomy** toward the outsider agent.

Proof. By using proposition 3.3, each agent derives delegation autonomy toward the outsider agent and since there is consensus among agents to have delegation autonomy toward the outsider agent, the group shares delegation autonomy.

Group delegation autonomy is a group attitude and is shared by each agent in the group. Additionally, as long as the right conditions exist, it will persist even when agents enter or leave the group.

Another group attitude is derived when a group of agents in strong-shared autonomy friendship have a member that is oriented to share its autonomy toward an agent outside the group and that agent forms a self-autonomy attitude. The group will develop a derived partial autonomy since the outside agent is assumed to take care of the task. We state this in the following proposition.

Proposition 3.5: If a group of agents in a strong shared autonomy friendship, has a member who holds an attitude to share autonomy with an agent outside the group, and that agent subsequently has a self-directed autonomy, then we say that the group has a derived **group partial autonomy**.

Proof. By using proposition 3.2, each agent derives shared autonomy toward the outsider agent and since there is consensus among agents to have shared autonomy toward the outsider agent, the group shares its shared autonomy. However, since the outside agent has a self-autonomy, each group member revises its autonomy to partial autonomy.

It appears to us that in addition to group members sharing group attitudes the attitudes can exist beyond a single agent. This suggests construction of a virtual agent (VA) that possesses such group attitudes and can interact with its counterparts. Such agents are different than middle agents of multiagency primarily since their life-time is limited by the group attitude. VA can be very useful in (a) fault-tolerance in case one of the agents disabled, and (b) forming larger units of agents. Further exploration of virtual agents that extend group attitudes is left to future work.

IV. A CASE STUDY

In this section we will briefly discuss a case study that motivate the use of notions developed in this paper. We will focus on techniques for flood management. In one such system designed with rivers, dams, and flashfloods, we imagine four types of agents: problem detection agents, water-control agents, management agents, and news/log agents. Problem detection agents will detect actual and predicted flashfloods based on sensory data and will alert agents in control and management. Water-control agents will open and close water-flow levels in dams. Management agents are decision makers to coordinate actions of all other agents. New/log agents will record patterns and prepare reports and alert the news agencies. Each of these types of agents will have a human user interface that allow a human operator to gain information about agent activities and to override or alter agent actions if necessary. In interacting with the human operator, agents need to reason about sharing their autonomy with the human operator and to adjust their level of autonomy as the human operator provides input.

Problem detection agents must reach a degree of confidence to issue a problem. Therefore, they will need to reason about their joint shared autonomy. Water control agents must reason about delegation and shared autonomy as they are in adjoining or upstream from one another. They might often have a problem with a delegation chain as one agent might have the crucial role of starting the flood relief action. Management agents will often be reasoning about delegation autonomies and possibly weak forms of delegation friendships. They will a position to delegate as a group to control agents so it will reason with group delegation autonomy. News/log agents are delegated to log and prepare information by the management agent(s).

V. CONCLUSION

Knowledge workers in organizations with software agents and human knowledge workers must rely on agents' ability to reason about their own autonomy as well as autonomy of others.

We have developed a formalism for representing agent autonomies as the potential to elect an executor of a goal either as self only, other agent only, shared with other agents, or partially self in terms of its beliefs, desires, and intentions. This BDI-style autonomy potential determination is the first stage of autonomy determination that leads the agent into a utilitarian analysis of the agent's actual autonomy choice. We have presented analysis of autonomy choice. The third stage is *enactment* where an agent will complete its action selection taking into account autonomy selection.

We discussed revision of autonomies in light of other agent's autonomy. We discuss the conditions for derived autonomies that not explicitly hold. A for m of derived autonomy is group autonomy shared by all agents of the group. Group attitudes are a novel concept and form a strong basis for developing theories of dynamic organizational structure and are useful for fault tolerance.

ACKNOWLEDGEMENTS

This work is supported by AFOSR grant F49620-00-1-0302. I appreciate Svet Brainov and Gordon Beavers for discussing this work with me.

REFERENCES

[1] 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, Seattle, WA.

[2] Barber K. S., Goel A., and Martin C.E., 2000. *Dynamic Adaptive Autonomy in Multi-Agent Systems*, In Journal of Experimental and Theoretical Artificial Intelligence, 12(2): 129-147.

[3] Beavers G., Hexmoor H., 2001. *Teams of Agents*, In Proceedings of the IEEE Systems, Man, and Cybernetics Conference, Arizona.

[4].Brainov S, Hexmoor H., 2001. *Quantifying Relative Autonomy in Multiagent Interaction*, In IJCAI-01 Workshop, Autonomy, Delegation, and Control.

[5] Brainov S., Sandholm T., 1999. *Power, Dependence and Stability in Multiagent Plans*. AAAI/IAAI 1999: 11-16.

[6] 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.

[7] Castelfranchi, C. 2000. *Founding Agent's Autonomy on Dependence Theory*, In proceedings of ECAI'01, pp. 353-357, Berlin.

[8] Dorais G., Bonasso R.P., Kortenkamp D., Pell P., and Schreckenghost D.. 1998. *Adjustable Autonomy for Human-Centered Autonomous Systems on Mars*, Presented at Mars Society Conference.

[9] Dworkin G., 1988. *The Theory and Practice of Autonomy*, Cambridge.

[10] Hexmoor H., 2000a. *A Cognitive Model of Situated Autonomy*, In Proceedings of PRICAI-2000 Workshop on Teams with Adjustable Autonomy, Australia.

[11] Hexmoor H., 2000b. *Case Studies of Autonomy, In proceedings of FLAIRS 2000*, J. Etherege and B. Manaris (eds), p. 246- 249, Orlando, FL.

[12] Hexmoor H., Kortenkamp D., 2000. *Autonomy Control Software*, An introductory article of the special issue of

Journal of Experimental and Theoretical Artificial Intelligence, Kluwer.

[13] Jennings N. R. and Campos J. R. 1997. *Towards a Social Level Characterisation of Socially Responsible Agents*, In IEE Proceedings on Software Engineering, 144 (1), 11-25.

[14] Jennings N.R.. *On Agent-Based Software Engineering*, In Artificial intelligence 117, 2000, pp. 277-296.

[15] Kumar A. and Zhao, J.L., 1999. Dynamic Routing and Operational Controls in Workflow Management Systems, *Management Science*, Volume 45, No. 2, February 1999, pp 253-272.

[16] Liebowitz J., 1999. *Knowledge Management Handbook*, CRC Press.

[17] Mele, A. 1995. *Autonomous Agents: From Self-Control to Autonomy*, Oxford University Press.

[18] Schneewind, J.B.1997. The Invention of Autonomy: A History of Modern Moral Philosophy, Cambridge Univ. Press.

[19] Steels L., *Corporate knowledge management*, Proceedings of ISMICK'93, Compiègne, France, pp. 930, 1993.

[20] Tiwana A., 2000. *The Essential Guide to Knowledge Management: E-Business and CRM Applications*, Prentice Hall.

[21] Tuomela R., 2000. *Cooperation: A Philosophical Study*, Philosophical Studies Series, Kluwer Academic Publishers.

[22] Wieringa R.J. and Meyer J.-J.Ch., 1993. *Applications of Deontic Logic in Computer Science: A concise overview*, In Deontic Logic in Computer Science, 17--40, John Wiley & Sons, Chichester, England.

[23] Varela F.J. 1979. Principles of Biological Autonomy, North Holland, New York.

[24] Wooldridge M., 2000. *Reasoning about Rational Agents*, The MIT Press.

[25] Yu E. and J. Mylopoulos, *Towards Modelling Strategic Actor Relationships for Information Systems Development -- With Examples from Business Process Reengineering*, In Proceedings of the 4th Workshop on Information Technologies and Systems, Vancouver, B.C., Canada, December 17-18, 1994. pp. 21-28.