In proceedings of the AISB 2002 symposium on Adaptive Agents and MultiAgent Systems, AAMAS-02, London, UK

Architectural Principles for Social Influence among Benevolent Agents

Henry Hexmoor
Computer Science & Computer Engineering Department
University of Arkansas, Fayetteville, AR 72701
hexmoor@uark.edu

Abstract

Benevolent agents with a high level of autonomy can be designed to be aware of their social influences in order to make sure the overall missions are successful. Additionally, the system of social influences can be engineered to meet the designer's objectives of predictable agent behavior. In this paper we outline a few agent architecture principles for such agents.

1 Introduction

Often the objective in designing multiagent systems of *benevolent agents*[1] is a system where interaction among agents is most congruent and beneficial to the overall mission. This leaves us with a choice in design between (a) detailed protocols and policies for interaction that leaves agents with little autonomy, and (b) agents with high autonomy that are driven by their social relations to determine their own rates and types of interaction relative to their shared mission. We choose the latter, which requires agents to continually monitor and to adjust their relationships for the overall mission success. An important paradigm in agent-based systems is to consider intentional notions of Belief, Desire, and Intention (BDI agents) (Wooldridge 2000). BDI agents possess update and revision functions for each intentional component. Beliefs are adapted to the agent's current state of mind and changes in the environment. Desires or specific goals are adapted to the agent's beliefs and attuned with the changing environment. Intentions over specific actions are adapted to agent's beliefs and desires. These adaptations create adaptivity for the agent at a behavior level called action selection. Coordination with other agents provides yet another form of adaptivity. Consider a scenario where two agents A1 and A2 are building a structure of blocks as shown in Figure 1. Both A1 and A2 have access to piles of blocks and know about the structural stability of blocks. An agent can only move one block at a time. In order to preserve stability, both agents must simultaneously place blocks 3 and 4. We can specify this problem with temporal constructs of parallel actions and ordering of agent actions or by allowing one agent to dynamically use its social relationship to take a lead role and cue the other. Specifically when it comes to block positions 3 and 4, whoever is the lead agent will set the rate of placement and the follower agent will monitor the lead agent for coordination. A1 and A2 opportunistically become lead depending on which one gets the block first and they are both equally willing and likely to play lead or follower.

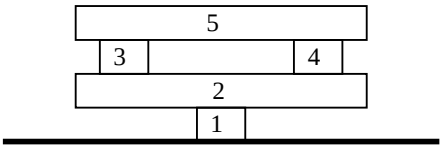

Figure 1 Blocks world

In this paper, we describe architectural principles that account for these dynamic adjustments due to social influences in the context of collaborative work. We discuss intuitively plausible types of responsibility and how agents can maintain it. In the remainder of this paper we will discuss social influences, agent relationships, and how agents can maintain the notion of the mission success. The paper ends with concluding remarks.

2 Social Influence and Relationships

As shown in Figure 2, social actions promote social influences. Since social factors are inter-related, initial social influences due to social actions might produce secondary and indirect influences. Actions of agents are

[1] Agents who are not benevolent have a more complex relationship to social influence and responsibility and this is beyond our current scope. Benevolent agents are characterized by their propensity to offer help.

guided by social values and norms, which affect the relationships among social influences. Therefore, social actions of one agent will influence the other agent in the context of prevailing norms and values as well as by the strength of the social action.

Whereas *physical actions* predominantly produce physical change, and *speech actions* predominantly produce epistemic change, *social actions* predominantly produce influence. Social action might be an action by one agent toward another, a mutual action of multiple agents, a bilateral action by multiple agents, or a group action. For brevity and illustration we limit our attention to the actions that commonly cause influences over the following notions: Help, Permission, Delegation, Service, and Resources. Actions about Help are performed to aid others in their task and specific actions might be to provide, to withhold, to request, or to reject help. Common actions over Permission are to request, to grant, to deny, and to withdraw. Actions over delegation might be to issue, to accept, or to reject. Service is an act but unlike help that is pro-active, it is passive. Common actions over service are to request, to give, or to deny. Common action over resources might be to request, to offer, to reject, to take away, or to prevent.

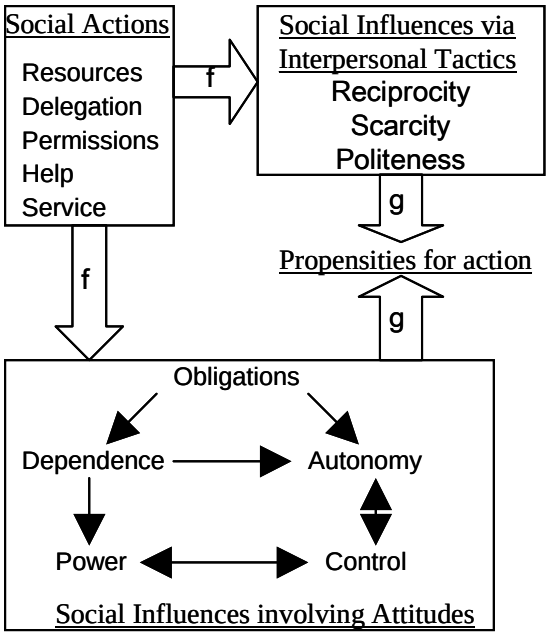

Figure 2 Social actions and influences; values and norms are not shown in the figure

The kind of influence we have in mind is close to normative (utilitarian) influence in social psychology in that agents reason about gains and losses from interpersonal interactions (Kassin 2001, Brehm, et al, 2002). Influences might involve overt interpersonal tactics such as reciprocity, scarcity, and politeness. Overt influences are immediate and deliberate as in interactions described in FaintPop (Ohguro, et al 2001). Reciprocity is when one agent returns an act by another or in effect *pays* for an act. An agent might reason about the reciprocity norm and perform a social act (e.g., help) based on an expected propensity in repayment. Scarcity is the norm that short supply produces a demand. An agent might use that norm and deny or hide services or resources. This is commonly used in theories of persuasion (Larson, 1998). Politeness as an interpersonal tactic is to get another agent to yield to another agent.

Influences might also be indirect and of indefinite duration. One type of indirect influence is via changes of attitudes. This is shown as the box in the lower part of Figure 2. These are perceived changes in social relationships that affect an agent's ties. The Figure shows our focus on Autonomy, Dependence, Obligations, Control, and Power and salient relationships we see among them. Later in this section we will discuss interdependencies among attitudes. Let's refer to the set of influences as I. Let's refer to the set of social actions as A. We define function f that maps the agent's current beliefs B, a set of currently active values V, a set N of currently active norms, and a set of social actions A to a set of influences I:

$$f: B \times V \times N \times A \rightarrow I.$$

An example is delegation of homework by a teacher to a student. Following shared norms governing relationships among teachers and students, assigning homework produce the influence for the student to adopt the obligation to carry out the homework.

Agents use influences that result from social actions they experience in their action selection. In addition to social influences, action selection accounts for means end analysis and rationality principles that are governed by the agent's endogenous sources. How action selection is affected by social influences is a complex issue that is beyond our current scope and is denoted as function g in Figure 2. Agents can project such a propensity for action in deciding to perform a social action. The reasoning might also include a chain effect where one agent produces an influence in another, which in turn produce an effect in another and so on. An agent can intend such a proliferation of influences and intentionally start such a chain reaction. This in fact is commonplace in a team setting.

We appeal to the reader's commonsensical beliefs, norms, and values to highlight a few relations between social actions and resulting social influences.[2]

[2] With more time and space we will state these more formally and concisely. We are seeking conditions under which these relationships hold and our enumeration here is only to point

1. When there is social action of request for help, the requesting agent is willing to lower its autonomy over a goal it cannot accomplish alone and would like to share or delegate the goal to someone else. An agent who finds itself in a position to respond to help is reciprocally willing to share or be delegated the goal that the requester cannot accomplish.
2. When someone has desirable resources that is somewhat scarce and announces availability of resource as a social action, there is a potential for dependence if other agents wish to access the resource. If some arrangement is made for provision of resources in exchange for something, the agents who want the resource and are willing to exchange, in addition to dependence, may experience a lowered autonomy over their goals that requires the resource.
3. An agent, who has a service that is commonly desired and is in short supply, has a high autonomy with respect to how it will make its services available. Reciprocally, an agent who need s scarce service will experience a low level of autonomy and high level of dependence toward agent who can provide that service.
4. An agent who delegates a task to another agent and when both agents are consenting to this delegation, the delegating agent will experience a dependence about the task and the delegee will experience an obligation to carry out the task toward the delegator.
5. When an agent increases or decreases permissions it gives to another agent, it is proportionately increasing or decreasing autonomy in other agents.

Next we highlight a few interdependencies among social influences that involve attitudes. Once an initial social action produces an influence on an attitude, indirect influences are propagated via relationships among attitudes. We feel engineering agents with specific connection among social attitudes is a powerful method for tuning agents indirectly.

1. When an agent has power over another agent it has indirect control over the agent. Similarly, when an agent has control over another agent, it has indirect power over the agent.
2. Autonomy and control are complimentary. I.e., Control and autonomy add up to a total amount and one lacks in some it has excess in another.
3. Obligations induce dependence.
4. Dependence diminishes autonomy.
5. Dependence induces power.

Next, we suggest that quantities of influence are in part based on utilities of norms and values. Additionally, magnitudes of social influences experienced depend on utilities derived from intensity of the social action. M(I) will denote magnitude of an influence I. In general, there might be differences between intended influence by the agent performing the social action and the receiving agent but for simplicity we assume these are the same amounts.

3 Behavior Guarantees

The values and norms can be designed to encode the mission of the collaborative effort in the agents as we stated in the introduction of this paper. For instance, in our blocks world example, a value can be stated as "preserve structural stability". Agents A1 and A2 might engage in a dialogue when they are seeking the next block to put on, when one finds a block and tells the other "I found one", the second agent who does not see a suitable block might say "go ahead". The permission by the second agent bolsters the agent's autonomy in attempting to place the block. When they have to place blocks simultaneously, once both agents have found suitable blocks for placement, one of them might take lead and say "let's go". Taking lead is a function of experiencing a relatively higher autonomy. Elsewhere, we have discussed relative autonomy (Brainov and Hexmoor 2001, Hexmoor 2002a, 2002b, and 2001). In this example, this relative autonomy might be due to relative skills in block placement or a norm such as "whoever finds a block and waits for the second agent takes lead." At any rate, taking lead implies control. As we explained earlier, the follower observes the actions of the lead and matches its actions for coordinated effort. If the follower can't keep up, it might say "wait" indicating a request for help and setting up an indirect adjustment in autonomy and control.

So far we have discussed how social relationships are interdependent and can be used to coordinate and produce a coherent set of actions. Social actions of one agent might affect another who might repeat by a social action that produces an influence on another and so on. There might also be shared or joint social influences. For example if the structure of blocks become unstable, both agents might independently experience an obligation to protect the structure. But when they become aware of one another's obligation, they arrive at a joint responsibility to protect the structure. This joint influence will produce a joint autonomy, which might lead to a joint action such as steadying the table under the structure. This is the basis of group influence and group action.[3]

We said that social influences rely on prevailing norms and values. An agent might have several levels of such values and norms at any one time. Obligations to uphold the ultimate group intent are derived from corresponding values and norms at the global level. These obligations

out the salient relationships. Also, not all relationships in the Figure 2 are explained.

[3] See http://csce.uark.edu/~hexmoor/AAAI-02/AAAI-02-cfp.htm for a workshop on this topic.

may compete against an agent's social influences at lower levels. Let's differentiate values and norms for an agent into n levels with level 1 being the highest (i.e., ultimate) and levels n being the lowest. Values and norms will be labeled with their level as values V_i and norms N_i. Let's redefine function f with f^i that maps an agent's set of beliefs B, a set of values V_i, a set of norms N_i, and a set of social actions A to a set of influences I^i:

$$f^i: B \times V_i \times N_i \times A \rightarrow I^i.$$

As designers of agent systems we can design mechanisms for encoding the desired ontological level to match the agent's responsibility level. This design-time responsibility encoding is a method that can be used to assure predictable agent behavior. If we design obligation categories (i.e., responsibilities within ontological levels for the agent), an agent might be directed to adopt specific obligations about certain tasks to perform on behalf of a chosen agent or the human user in case the agent interacts with a human. This will affect the agent's autonomy and control with respect to the agent (or the user). For example if the project is safety-critical, overall project goals (and corresponding values and norms) are given a higher ontological status in the agent's makeup.

Agents might experience simultaneous social actions that have influences at different value and norm levels. These influences will also have different magnitudes. At times there might be conflicts among these influences. The conflict might be within an agent or between agents. A simplistic conflict resolution for an agent is the rule "if an influence at a high level is conflicted with an influence at a low level and as long as the magnitude of the influence at the low level is not much larger than the one at the high level, choose the influence at the high level." I.e., higher influences suppress the lower influences unless the strength of influence at the low level is significantly larger than the influence at the high level. In the exception case, we can introduce case-by-case domain rules to make sure hierarchies are largely maintained but specific over-riders are possible.

When agents share ontological levels of values and norms, it is easy to see that they have a greater chance of harmony. Conflicts among such agents can also be resolved using our resolution rule. In this case, one agent might sacrifice its highest social influence for another's even higher social influence.

Overall missions can be guaranteed among agents who share the values pertaining to that mission if we specify certain social influence tolerances in agents. First, we can specify how much tolerance we allow for adverse social influences before reacting to them. Next, we can specify the threshold of deviation from other social influences for suppression of lower level social influences.[4] We can use this method to other levels of norms and values and produce similar guarantees at those levels. The notion of guarantee we introduce here differs from validation and verification. In validation and verification, programs are tested to obey certain properties (Engelfriet, et al 2002). This is not easily possible with multi-agent programs that have many more paths of execution due to the level of autonomy we provide agents. For instance, in nontrivial systems, chain effects of social influences are too complex to account for agent actions. We envision methods for setting up combinations of norms and values such that they contain unfolding chains and suppress undesirable results of influence chains.

Consider our blocks world example with three agents instead of two. As a delegation social action, all three agents might have found blocks that can fit in block positions 3 and 4 and out of politeness (and the corresponding social influence) one agent might say to another "go ahead" (this is a social action) and second agents might experience the same influence and say "go ahead" to the third agent and back to the first agent. A norm that can break this influence chain is "if politeness leads to inaction, the earliest agent to be polite will proceed".

4 Conclusions

We outlined how social actions generate social influences and showed a few salient interdependencies among social influences. We then discussed agents that can be designed to favor social influences that pertain to their highest level of norms and values and some exceptions. If agents shared norms and values, we can design agents that guarantee guarding against adverse social influences, suppression of social influences due to lower level norms and values, and undesirable chains of influence. This gives us a practical methodology for using social influence in implementing social responsibility among benevolent agents.

Acknowledgements

This work is supported by AFOSR grant F49620-00-1-0302.

[4] Agents who share mission level values and norms might not share norms and values at lower levels.

References

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

S.S. Brehm S.M. Kassin, S. Fein 2002. *Social Psychology*, Houghton Mifflin pub.

J. Engelfriet, C.M. Jonker, and J. Treur, (In press 2002). Compositional Verification of Multi-Agent Systems, In Temporal Multi-Epistemic Logic, *Journal of Logic, Language and Information*.

H. Hexmoor, (In Press, 2002a). In Search of Simple and Responsible Agents, In the Proceedings of *NASA GSFC/JPL Workshop on Radical Agents*, MD.

H. Hexmoor, (In Press, 2002b). From Inter-Agents to Groups, In *International Symposium in Artificial Intelligence, ISAI-01*, India.

H. Hexmoor, 2001. A Cognitive Model of Situated Autonomy, In Advances in Artificial Intelligence, Springer LNAI2112 -pages 325-334, Kowalczk, Wai Loke, Reed, and William (eds).

T. Ohguro. K. Kuwabara, T. Owada, and Y. Shirai, 2001. FaintPop: In touch with the social relationships, In International Workshop on Social Intelligence Design, The 15th Annual Conference of JSAI, Japan.

S. Kassin, 2001. *Psychology*, Third Edition, , Prentice-Hall.

C. U. Larson, 1998. *Persuasion: Reception and Responsibility*, 9th edition. Boston: Wadsworth.

M. Wooldridge, 2000. *Reasoning about Rational Agents*, MIT Press.