

Detecting, Expressing, and Harmonizing Autonomy in Communication Between Social Agents

Henry Hexmoor

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

Heather Holmback and Lisbeth Duncan

Mathematics and Computing Technology
PhantomWorks, The Boeing Company
M/S 7L-43, P.O. Box 3707, Seattle, WA 98124-2207

Abstract

We discuss the importance of expression and understanding of autonomy among social agents. We present steps in design of systems that facilitate automatic detection and expression of interagent autonomy.

1. Introduction

Multiagent systems of the future will consist of humans and synthetic agents. The synthetic agents in these systems will need to be safe, predictable, and adaptive; hence, modeling the interaction of agents is an active area of research at this time. One popular approach to studying agent interaction is to model the norms, conventions, and collective actions of a group as they bear on a prototypical member of the group [Boman, 1997]. Far less attention has been focused on unique aspects of individual agents and the individual differences in using the norms and conventions. A systematic study of these individual differences that might lead to deviations from expected behavior is needed in order to preserve system qualities. A key factor in how a prototypical agent would respond personally in a given situation is its level of *autonomy*. We define “situated autonomy” as an agent’s stance toward a goal at a particular moment when facing a particular situation such that the stance will be used to generate the most appropriate action [Hexmoor, 2000]. In other words, situated autonomy concerns the status of the agent’s goal adoption and the level of independence assumed by the agent in achieving a goal. We have argued that a combination of the type and strength of an agent’s beliefs and motivations lead the agent to act in one of the following ways: (a) the agent chooses itself to be the executor of the goal; (b) the agent delegates the goal entirely to others; (c) the agent shares its responsibility with other agents; or (d) the agent has a relatively small and undetermined level of responsibility toward the goal.

Our interest in the long run is to develop a model that unifies action and language and specifies how human and synthetic agents can most safely and effectively interact in a group. An integral part of such a model is communication that incorporates natural language. Human language facilitates communication of personal or desired autonomies and encodes the conventions and rules of group interaction. Human discourse contains many expressions of directives and responses that propose extra-linguistic goals and specify expected levels of autonomy among discourse participants. These aspects of communication are embodied in both the illocutionary force of the utterance (i.e., the speech act that the speaker intends to convey with his utterance) as well as the perlocutionary effects (i.e., the effects that the speaker’s utterance or message has on the hearer). Perlocutionary effects relevant to situated autonomy include factors such as the perceived urgency and importance of the directive as well as the resulting trust and commitment of the hearer. An agent may generate a response that explicitly expresses a perceived level of urgency, importance, and trust, and implicitly conveys the agent’s status of goal adoption and the level of autonomy assumed. If synthetic agents are to be safe, predictable, and adaptive, we need to endow them with abilities similar to those employed by humans to detect, express, and harmonize their autonomies.

Our envisioned agents will be robust in that they can perceive self-autonomies, autonomies of others, and graceful accommodations in their autonomies. If we take an agent’s autonomy as its stance toward a future desirer or intender agent of a goal, *robust autonomy* may mean a fail-safe autonomy policy; one that recognizes patterns, considers alternatives, and has alternate agents in mind when things go awry. Our goal is to provide a first order account of autonomy assessment via dialogue. We will leave dynamic alterations of the agent’s autonomy assessment to future work.

While issues of autonomy can come into play in many types of communicative functions, we focus on

communication between human and synthetic agents involving directives (i.e. commands and requests) and responses to those directives. Responses can include acknowledgment, acceptance, refusal, clarification, and negotiation of the directive. Our goal is to develop a framework and system in which synthetic agents can: 1) detect the appropriate illocutionary force and perlocutionary effects in human communication in order to operate with the appropriate level of situated autonomy; 2) express relevant factors about their autonomy and commitment in their communications with humans; and 3) use spoken communication appropriately to harmonize their autonomy with the needs and constraints of the situation.

“The third wheel” and many other proverbial references point out the need for harmony in autonomy of agents. In collaborative tasks, agents need to compliment one another’s autonomy in order to minimize redundancies and crossing paths. The point of stating intentions before carrying out tasks is in-situ coordination. Stating impending or possible desire or intention is an approach to coordinate autonomies. “I might consider doing X”, “I will make sure X”, “I could do X” are examples.

To be more precise, let’s consider an agent’s autonomy can self-directed (Ii), other directed as in delegation (Dij), shared with other agents as in teaming (Sij), or be partially self-directed (Pi) [Hexmoor, 2000]. With only two agents and 4 autonomy types, there are $2^4 = 16$ cases. With more than two agents, there are many more cases. Interesting cases are the ones where a group agents have mutual shared autonomy that lead to team formation.

For simplicity, let’s consider only two agents I and j. In the cases (Ii, Ij), (Pi, Pj), (Sij, Sji), autonomies are harmonious. In the cases (Dij, Dji), (Dij, Sji), autonomies are in clear conflict. In the cases (Ii, Sji), (Pi, Sji), (Ii, Dji), (Pi, Dji), if we consider delegation or sharing autonomies to precede before the second agent’s autonomy determination, there is no disharmony. Otherwise, the agents have a disharmony in their autonomy. Disharmony can be resolved by dialogue as in utility-based negotiation.

In the remainder of this paper we will outline some salient principle, sketch an outline of a methodology, and give status of our implementations.

2. Our Approach

We make four observations here that characterize our approach. *First, detection and expression of implicit autonomies requires a shared understanding of the social roles and personal relationships of the participating agents.* For example, a complete stranger usually will not issue a directive to another agent. Furthermore, how seriously a directive is taken by one agent depends on its perception of the authority of the agent issuing the directive. Of course, the social relationship between agents is not static; the changing, or dynamic, relationship during the conversation affects the harmony of autonomies. If two agents have a positive relationship, each agent will change its autonomy to complement the other agent.

Second, the form of the directive holds clues for autonomy. One aspect of form is specificity. A very detailed directive already contains many of the needed decisions about the execution of the goal and leaves little flexibility; that is, it expects little autonomy on the part of the addressee. An indirect command (e.g. “Could you do X?”) generally gives more autonomy to the addressee than a direct command (e.g. “Do X.”). Likewise, the form of the response to a directive also conveys much information about the autonomy of the speaker. For example, if an agent responds to a directive by acknowledging it but not immediately agreeing to it, this indicates that he assigns it a low level of urgency relative to his current situation.

Third, the content of a directive and the responses to it contribute to the autonomy. There can be explicit semantic content (e.g. “Do X now.” vs. “Do X when you have a chance.”) that spells out the degree of autonomy the speaker is giving the addressee with respect to carrying out the directive. The content can also be less explicit and more dependent on pragmatic factors. For example, the amount of autonomy allowed by “Do X soon.” is highly dependent on the context in which it is uttered (e.g. the current tasks of the addressee, the shared information about the nature of X, the time constraints in doing X, etc.).

Fourth, an agent’s internal mechanism for autonomy determination affects the detection, expression, and harmony of autonomies. In other words, the communication between agents is not free from individual biases. Characteristics of individual agents are part of the situational context in which the communication is interpreted. There are many endogenous sources of autonomy. For example, transitory moods, drives, and overall temperament are sources of autonomy for humans. For instance, the hunger drive may lead to a distressed mood, and in this

situation, an agent with an aggressive temperament will tend to adopt a high level of self-autonomy to seek food. Similarly for synthetic agents, there are possible endogenous sources of autonomy, such as the need to replace a battery or other energy source before being able to continue any further interaction with other agents. In such situations, agents will be unlikely to harmonize their autonomy with others.

In sum, the form, content, and the pragmatics of communication (i.e., the interpretation of the form and content of an utterance relative to information in the static and dynamic context) all contribute to situated autonomy. A synthetic agent will need to analyze and interpret a human agent's linguistic input relative to the context in order to detect information relevant to forming its stance toward the goal. Likewise, a synthetic agent will have to use the appropriate linguistic form and content relative to the context to generate a response to the human that communicates its stance toward the goal.

In our approach, human-agent communication has many similarities to human-human communication. This is natural since in human-agent communication there is a human in the loop using natural language along with many of its implicit assumptions. However, human-agent communication differs from human-human communication and is more similar to agent-agent communication in the following ways: 1) there are more limited topics and communicative functions involved; 2) there is a lack of digressions; and (3) there is a greatly reduced emphasis on politeness for social needs. We plan to use our work on natural language understanding and generation [Duncan, et al., 1999] as well as our framework of *conversation policies* that govern agent-agent communication [Greaves, et al., 1999].

A conversation policy is a model of the structure and content of a type of agent conversation for a certain communicative function. Conversation policies can combine to form larger policies and can have varying degrees of specificity. An instantiated conversation policy includes information about the context in which the communication is situated (e.g. timing constraints, roles and relationships of the conversational participants, reliability of the communication channel). A set of conversation policies for the function of one agent attempting to get another to carry out a certain task would specify the timing and sequencing of the directives and responses as well as any required negotiation, repair of miscommunication, reporting of the results of carrying out the task, beginning and ending of the conversation, etc. In agent-agent communication, these policies are intended to govern the transmission of messages in some Agent

Communication Language (ACL). We will use the same type of policies as a structure for human-agent communication in our proposed work, but the policies will govern natural language expressions. The conversation policies needed for human-agent communication will be embellishments and extensions of those governing the communication between simple synthetic agents, but will be similar to conversation policies needed for more intelligent, more autonomous agents. Our approach captures the similarities between agent-agent and human-agent communication, while allowing for differences in the actual expression of the input and output of the agent system (i.e., an ACL or natural language).

While some aspects of situated autonomy will be the same in agent-agent and agent-human interaction, in the latter case the detection, expression, and harmonization of situated autonomy involves the interpretation and generation of natural language utterances relative to the context provided by a set of appropriate conversation policies governing the human-agent interaction. To address this, as part of this contract, we will extend our work in the following areas: identifying the important elements of situated autonomy; developing more complex conversation policies for human-agent communication; and expanding our natural language understanding and generation capabilities to facility spoken communication between agents.

3. Sketch of an Autonomy Assessment Theory via Dialogue

Here, we outline a methodology for detecting autonomies in the agent's top level loop.

1. Upon receipt of a new command, use NL techniques for initial assessment of desired autonomy at the lowest confidence level.
2. If no feedback is received in a reasonable amount of time, increase confidence.
3. If feedback is received in a short amount of time, update desired autonomy using NL techniques.
4. If feedback is ambiguous, interrogate and update autonomy:
 - 4.1 Ask about expected intentions and desires
 - 4.2 If 4.1 is ambiguous, ask about abilities of commanding agent and its expected abilities from it.

4. Outline of Two Experimental Setups

We are developing the following two programming environments for gathering data and conducting experimenting. The first program gathers human responses from suggestions made by a software agent while playing a game that requires the human subject to move a graphic object from one point to another based on verbal interactions with the software agent. This program involves a navigation task in which a human player uses a joystick to navigate a submarine in a minefield. The player's objective is to move the submarine from one point to another without hitting any mines. The submarine has movement delays and there is a time limit for moving from one point to another. The software agent has more knowledge of the domain and offers suggestions. The majority of its suggestions are helpful; however, some of its suggestions are not useful or will produce adverse results. The human player does not initially know the relationship between himself and the software agent and the relationship changes over time. We gather online feedback from the human player. We have not finished implementation of this system but it appears that people's personal biases toward the agent suggestion are difficult to factor out of the human-machine interaction.

The second program aims at how a software agent can detect, express, and harmonize its autonomy in conjunction with a human operator. Our experimental design involves a mobile robot that uses sonar and limited visual abilities to map an unfamiliar area that contains randomly located pieces of furniture. The robot interacts with a human agent using natural language understanding and generation in order to achieve or to modify its objectives in accordance with the changing dynamic context of the situation. The majority of the human's instructions are helpful, but some might have adverse effects on the robot's ability to complete its mapping task. The human can make 3 movement requests: turn right/left/back; change mapping grid-size. Human request for these commands come in 5 shades of soft request to hard request. Hard requests should be processed immediately. This is giving the robot very little autonomy. Whereas soft requests are to be processed after considering whether they improve the mapping task. This is giving the robot more autonomy. The robot does not have access to the hardness of the request (i.e., desired autonomy of the human operator) and guesses at it. By observing the pattern of subsequent commands, the robot modifies its guess at hardness of the request.

5. Conclusion

In summary, our work on how synthetic agents detect, express and harmonize levels of autonomy will affect the theory of interoperability of intelligent, autonomous agents. Our work contributes to an understanding of how agent reasoning and communication can be made safe, predictable, and adaptive. We hope to extend our work in order to contribute to the general theory of pragmatics in communication involving humans and synthetic agents.

References

- M. Boman (1997). Norms as Constraints on Real-Time Autonomous Agent Actions. In M. Boman and van de Velde (eds.), *Multi-agent Rationality*, Proceedings of the 8th European Workshop on Modeling Autonomous Agents in a Multi-agent World, **MAAMAW'97**, Lecture Notes in AI, vol. 1237, Springer Verlag, Berlin, pp. 36-44.
- L. Duncan, W. Brown, C. Esposito, H. Holmback, P. Xue (1999). Enhancing Virtual Maintenance Environments with Speech Understanding. In *The Boeing TechNet*, March 1999.
- M. Greaves, H. Holmback, and J. Bradshaw, (1999). What is Conversation Policy? In *Proceedings of the Autonomous Agents '99 Workshop on Specifying and Implementing Conversation Policies*, Seattle, WA, May 1999.
- H. Hexmoor, (2000). A Cognitive Model of Situated Autonomy, In *Proceedings of PRICAI-2000 Workshop on Teams with Adjustable Autonomy*, Australia.