Multi-agent Interactions: A Vocabulary of Engagement*

Patricia Goldweic and Kristian J. Hammond
Artificial Intelligence Laboratory
Computer Science Department
The University of Chicago
Chicago, IL 60637
goldweic@cs.uchicago.edu

Abstract

Our project concerns the definition of a content theory of action appropriate for agents that act in a multi-agent environment and its implementation in a multi-agent system. Such a theory has to explain what agents know and how they use this knowledge; it has to identify what resources are available to the agents when they must decide on an action; it has to allow agents to reason and engage in concrete activity in their domain. More important for our research, and in contrast with numerous works in Distributed Artificial Intelligence, such a vocabulary must provide a basis for agents to decide and learn *when, how* or *with whom* they should cooperate. In this paper we suggest a vocabulary of interactions for intelligent agents. Our vocabulary attempts to do justice to the situated character of action with respect to the disparate but related dimensions of physicality, sociality and experience.

Introduction

Realistic multi-agent environments are characterized by uncertainty, distribution of skills and knowledge, and some degree of unpredictability. Yet many of these environments, like workplaces, can become relatively stable over time, enabling routine patterns of interactions to emerge. Many AI researchers have chosen to develop reactive agent architectures to deal with unpredictability and uncertainty. However, it is striking that almost no multi-agent systems have been built that take advantage of the relative stability of the agents' interactions to learn and improve the agents' behavior over time (Bond and Gasser, 1988; Gasser and Huhns, 1989). Our research attempts to fill this gap.

Our approach is to define a content theory of action appropriate for agents that act in a multi-agent environment and implement it in a multi-agent system. Such a theory has to explain what agents know and how they use this knowledge; it has to identify what resources are available to the agents when they must decide on an action; it has to allow agents to reason and engage in concrete activity in their domain. More important for our research, and in contrast with numerous works in Distributed AI (DAI), such a vocabulary must provide a basis for agents to decide and learn *when, how* or *with whom* they should cooperate.

In this paper we present a vocabulary of interactions for intelligent agents. This vocabulary attempts to do justice to the situated character of action with respect to the disparate but related dimensions of physicality, sociality and experience.

What Supports Action?

To intelligently act in a multi-agent environment, an agent needs to have access to a varied set of resources that serve to influence its actions. We distinguish three main categories of resources.

The first category includes the agent's knowledge about how to act in its domain (e.g. what goals to pursue, how to pursue them, etc.). This knowledge has been frequently called *know-how* or *planning* knowledge.

The second category concerns the social, historical and environmental context of the task at hand.

A third category relates to habitual practices. As opposed to *know-how*, these resources are not idiosyncratic[1]. They include knowledge about standard or recurring interaction patterns (explicit or implicit), social conventions, and dispositions of agents to interact in set ways. Agents who are asocial use resources in the first two categories to decide on action. However, common practice must

*This work was supported in part by AFOSR grant number AFOSR-91-0112, DARPA contract number F30602-91-C-0028 monitored by Rome Laboratories, DARPA contract number N00014-91-J-4092 monitored by the Office of Naval Research, Office of Naval Research grant number N00014-91-J-1185

[1] In fact they represent the *culture* of the agents' community.

orient the agents' behavior if they are to intelligibly act in a community.

But how do agents *use* all these resources? What role do these resources play in the evolution of patterns of interactions? In the rest of this paper we provide example interactions of our project domain that serve to motivate our vocabulary and to suggest answers to these questions. We believe that such a vocabulary, with which agents can reason and engage in concrete situated activity in their domain, is a *must* in any realistic theory of action for intelligent interacting agents.

Interactions and AI

Psychology and MA systems

Recent work in social psychology applied to multiagent systems (Castelfranchi and Miceli, 1991; Castelfranchi, 1990) has been formally addressing some important points overlooked by the DAI community, mainly the notions of *dependence* and *power* among agents[2]. In fact, the research we describe in this paper is compatible with these authors' notions of autonomy, and with their claim that dependence among agents is an important basis for decision making. We view these concepts as part of the set of resources available to the agents when deciding what to do next.

Interactions in CBR

Recent work in Cased Based Reasoning has been concerned with the indexing of stories that involve multiple agents (Schank and others, 1990). This work has been aimed at defining a universal vocabulary useful for indexing such stories from different perspectives. By covering a large space of possible situations (involving one or more agents), these researchers attempt to contribute to explaining human *reminding* phenomena. Although our vocabularies partly overlap, our main goal is, however, very distinct: instead of developing a *content theory of indexing*, we are trying to develop a *content theory of action* that will support an agent's decision making when actually *participating* in activity. Thus, what we are actually after is a vocabulary that deals *centrally* with issues of *planning* and situated *action*; we expect our indexing vocabulary to derive from it and not vice versa.

Interactions in DAI

A theory of social interactions for artificial agents must be able to account for the moment-to-moment accomplishments of the individual agents. Numerous works in DAI have neglected the development of theories of action and thus concentrated on building centralized systems or systems in which the decisions on what to do next are left to the system's designers (Durfee and Lesser, 1987; Rosenschein, 1988; Georgeff, 1984; Georgeff, 1983). Other work has provided descriptive theories of cooperation which, although insightful, are not readily applicable to agent design (Cohen and Levesque, 1987; Werner, 1989; Lochbaum *et al.*, 1990). For a theory of interactions to be applicable to agent design it *must* support the moment-to-moment decisions and actions of the individual interactants. And it can only do so by explaining how agents can decide upon and engage in meaningful action: what agents know and how they use this knowledge, how the current context determines and shapes action, how standard practice[3] influences the agents' behavior. Without addressing these points, a descriptive theory is only useful to rationalize interactions *a posteriori* of their occurrence, but will never be able to explain the *a priori* process by which agents become socially motivated and act towards common objectives.

Additionally, a theory of interactions, as part of a theory of action, must be able to account for the evolution of multi-agent interactions over time. Most work in AI has completely ignored the fact that interactions do not occur in a vacuum, but that they are historically situated. In our project we attempt to provide a theory that can explain how multi-agent interactions can evolve over time by first acknowledging the fact that agents are situated in the context of their own experiences. Our vocabulary provides a step towards the development of such theory.

The Project

The project's domain is a simulated world in which agents engage in maintenance tasks: they clean floors and windows, move furniture between rooms, and deliver mail within the confines of a unique building. Agents meet when they perform their tasks, either because they happen to be working in the same room or corridors, or because they explicitly decide to interact.

Agents that habitually interact tend to stabilize their relationships over time. Sometimes, the distribution of skills and tasks among the agents, and the dynamics of the activity itself, are such that agents develop cooperative routines to better pursue their goals. For example, two agents that often work in the same rooms cleaning floors and carpets respectively, may soon find out that pushing

[2]It is interesting to note that the example we provide in this paper shows what can occur in an extended interaction when a dependence relationship (the fact that one agent cannot readily push a heavy object by itself) exists.

[3]Although we are concerned with the design of artificial societies (that will nonetheless eventually have to coexist with the human society), we believe there is in fact a need to identify what *practice* in those societies is or can be all about.

heavy furniture is easily done cooperatively, and that a shared wastebasket will be better off in a place where both agents have easy access.

On the other hand, when interactions between agents tend to disrupt individual accomplishments (i.e., agents getting in each other's way), agents are better off if they explicitly coordinate their tasks, or even if they avoid each other. In these cases, they can either individually reorganize their tasks, or explicitly negotiate to avoid the inconveniences.

Perhaps most importantly, agents do not have to continually engage in new interactions from scratch; they can take advantage of their past social experiences to decide on action that is beneficial in the long run. For example, if two agents have skills that complement each other's, we would expect them to mutually cooperate in a way that the agent with the strongest skill helps the weakest party. In these cases, agents can make tacit or explicit deals to more effectively pursue their goals. The rationale is that long term stable relationships are better than beneficial one-time interactions.

An Example: Moving Furniture

In this section we provide three examples of consecutive interactions that motivate our vocabulary. The following are the questions we will be interested in answering:

- What specific resources are available to the agents when they intelligently engage in their activities?
- How do those resources interplay to suggest appropriate actions?
- Are there any acceptable or standard behaviors that restrict or help construct the agents' activities?

A first encounter

Consider the following interaction between two parties: Tom and Bob, members of the maintenance team. Assume that both of them have moved heavy objects collaboratively in the past, although not with each other.

> Tom and Bob both want to push a heavy couch. Tom is the first to approach it. He does not see Bob, who just entered the room. Bob realizes that Tom is, with some difficulty, trying to push that same heavy couch and thus offers him help. Tom accepts and they push the couch together to the right place. In doing so, they decide to hold the two opposed ends of the couch and coordinate to push them simultaneously. They communicate briefly in order to coordinate their joint activity.

The agents' resources

What are the resources the agents use to reason and engage in the previous interactions? First, they **know how** to push couches. This knowledge provides them a basis for purposeful action. Second, they know about **common social practices**, in particular about how to cooperate and coordinate to perform concrete activities. This knowledge gives them expectations on what the other party will do or say during the interaction, and restricts their available choices of action. In addition, the agents **perceive** their shared physical environment: they see and hear each other and can see and touch the objects in the room. In this particular interaction, the physical context constitutes a very important part of the common ground of interaction. In fact, the agents' ability to act opportunistically, their appropriate know-how and the fact that they inhabit a shared physical environment, are the basis for successful coordination between the agents themselves and with their physical world.

An agent's perspective on an interaction

The first story exemplifies a common situation in which two agents share a goal: pushing a couch. However simple, little has been done in AI to explain why two parties such as those in the situation above, would realistically come to collaborate in their enterprise. We will attempt to do exactly that.

Agents need to have access to a vocabulary that characterizes different aspects of an interaction. This vocabulary represents part of the common sense knowledge an agent needs in order to engage in meaningful action and in order to learn from it. In what follows, we describe how one participant understands this interaction in terms of such a vocabulary[4].

First, Tom recognizes that Bob is the **initiator** of the interaction. Although apparently trivial, this fact will be important for an *a posteriori* evaluation of the interaction.

Tom did not expect such interaction because he was focusing his **attention** on his activity when Bob interrupted him. Moreover, he recognizes that when Bob initiated the interaction, he had not perceived Bob's presence. This tells Tom that he could not have anticipated that such an interaction was going to take place.

Next, Tom understands that the **rationale** for such an interaction involves:

1. the fact that he cannot readily push the couch by himself, and
2. the fact that Bob has an overlapping goal of moving the same couch, which Bob has communicated to him.

[4]We understand other interpretations are possible. The one we provide seems plausible and does not seem to require complicated reasoning machinery.

Tom also recognizes that if he accepted Bob's offer, he could abandon his current plan for pushing the couch, and more easily achieve his goal with Bob's help. This constitutes Tom's **individual perspective** on the interaction.

From a **social perspective**, Tom recognizes that the interaction would promote a stable relationship between him and Bob. He also recognizes that the social context of his activity would change from a situation of disengagement to one of joint engagement in the couch-pushing task.

After pushing the couch with Bob, Tom is able to evaluate this particular interaction from different perspectives. Because he and Bob were able to successfully move the couch to an agreed upon location, he recognizes that the interaction has done justice to its rationale.

Additionally, Tom will tend to reciprocate Bob's cooperative behavior in the future if the situation provides for doing so. We believe it is important to analyze *how* Tom can reach such decision. Although everybody would agree that reciprocation plays a role in social interactions, it is harder to articulate how an agent can make use of this 'principle' or when to use it. A possible explanation would be that Tom explicitly considers Bob's goals and beliefs and then decides that that is the appropriate thing for him to do. Another explanation, which is the one we prefer, suggests that since the agents are participating in a certain common social situation, they orient to each other in ways that are also part of the common practice. In this particular case, the fact that Tom was taken from a situation in which he was socially disengaged to one in which he is jointly engaged with Bob in an activity he desired, is enough for him to be willing to reciprocate in the future[5].

What can Tom learn from this interaction? He could learn the following:

- Whenever he and Bob have the overlapping goal of pushing a couch, a plan to do it collaboratively benefits both of them and should therefore be suggested if a future situation provides for doing so.
- Whenever Tom is faced with the task of pushing a heavy object, a reasonable plan (or piece of know-how) he could use, involves asking Bob for help.

A second encounter: how experience shapes the activity

Consider the following story:

> Tom and Bob are cleaning a room. Tom is cleaning the floors and Bob the windows. Tom needs to push the heavy couch to clean underneath. He decides to ask Bob for help. Bob accepts and they proceed to move the couch as before.

How does Tom's analysis change with respect to the previous case? Tom anticipates his failure to easily push the couch and he gets reminded of the previous interactive experience. He is now able to use this piece of knowledge as another means to achieve his goals (have the couch be moved by both agents). Since Bob had not shown any interest in moving the couch, he decides to seek Bob's help (thus actively initiating this interaction). He understands that this interaction, building upon a previously established relationship, clearly counts as a favor to him, and thus he will attempt to reciprocate Bob's behavior in the future.

Bob's perspective is richer to analyze. When he remembers the previous interactive experience with Tom, he realizes that he had been able to take advantage of a positive interaction among the two agents (both were pursuing the same goal and had a chance to achieve it jointly). He also reasons that his own goal of moving the couch recurs every time he wants to clean the adjacent window. So in order to take advantage of such an opportunity once again, it would be convenient if his goal of moving the couch were to arise exactly when Tom's does. Clearly, Bob does not have control over Tom's goal generation processes, but he can reasonably behave in one of the following two ways. He could try to make an explicit deal with Tom to fix the schedule of pushing. Alternatively, he could reschedule his own activities so that his couch-pushing goal arises at a convenient time (Hammond and Converse, 1991). Since his desire to push that couch arises as a consequence of his desire to clean the adjacent window, then he could choose to clean that particular window now instead of the one he was currently working on.

Bob chooses the second option because it is socially preferred and because it is not too costly or disruptive for him to switch to working on another window. In general, it is more acceptable that the approached party attempt to cooperate if asked to do so (especially when an ongoing stable relationship among two agents already exists) unless the situation would disrupt his current activities.

A third encounter: developing a more enduring approach

Imagine that the situation we just described repeats itself. How would the interaction between the two agents change? We postulate the following scenario.

Bob is faced with a richer experiential context

[5] This analysis greatly differs from an alternative one based on an explicit consideration of goals and beliefs and expected cost/benefit of a potential interaction. Note that according to such an alternative, Tom might reason that, since the two agents shared the goal of the activity, there wouldn't be anything to reciprocate for.

than ever before. He is now able to recognize other valuable pieces of information. First, he remembers that he rescheduled his activities (previous story) to accommodate for the same sort of situation. This would conceivably make Bob intend to go for a more enduring approach. Moreover, he now recognizes that his desire to push that couch is not only a recurrent goal of his but of Tom's too. These two pieces of knowledge combine to suggest going for an explicit deal this time. Such a deal would include, for example, fixing the schedule for pushing and also for replacing the couch when the activities of the two agents do not depend on the couch's location anymore.

Note that the deal takes long term advantage of a positive inter-agent interaction and reduces the agent's future cognitive requirements (no need to reorganize activities anymore). Moreover, the deal also promotes a relationship by maintaining a cooperative social context while having the agent act within the socially acceptable bounds: because the situation involves a recurrent goal of both agents, it is socially acceptable to try to cut a deal that is convenient to both [6].

The three examples we have described exemplify how a particular multi-agent routine can develop among parties over time that does justice both to individual agency and to a historical process of interactions.

A Closer Look at the Vocabulary

The shared facts

Our vocabulary includes a set of items representing some resources commonly available to the interaction's participants:

- the agents involved
- the interaction's *initiator*
- the physical setting of the task
- the shared past experiences
- knowledge about common social practices

The idiosyncratic resources

Additional vocabulary items represent the following idiosyncratic resources:

- the interaction's *rationale*. An agent may interact with others due to different reasons. Some of these reasons can be traced back to the planning domain's vocabulary for plan failures as applied to any of the participants of the interaction: lack of physical ability, lack of knowledge or skill, or lack of resource. However, other rationales rise out of social considerations. For example, an agent may anticipate inter-agent conflict and seek to avoid it. Interaction can also occur simply because "it's always done that way."
- whether the interaction *satisfied its rationale*.
- whether the interaction was *expected* or not;
- whether the interaction was *desired* or not;
- how the interaction *relates to* the agent's current, suspended or future goals and activities. This vocabulary concerns functional relationships among goals and activities of the same or different agents and extends work in (Wilensky, 1978; Hammond, 1990). Some examples are:
  - The interaction may complete the task in which the agent is engaged, or some part of the task.
  - The interaction may require that the agent suspend work on his task.
  - The interaction may steal a resource needed for the agent's task, or be counter to one of the agent's goals.
- the agent's *perspective on* how the interaction relates to the other agents' activities;

The social perspective

If agents are to decide *when*, *how* or with *whom* they should cooperate, our vocabulary must be able to describe how a particular interaction can affect an inter-agent relationship. This knowledge partly constitutes an agent's social awareness. The fact that agents are socially aware partly explains why they decide to establish long term relationships with others instead of engaging in one-time interactions: relationships may sometimes help agents better (or more cheaply) pursue their goals.

This vocabulary deals with concepts such as whether an interaction *promotes*, *conflicts with* or *stabilizes* an ongoing relationship. Central to this is the notion that an agent holds a certain *attitude* towards the other participants which socially situates himself in the interaction. Thus, our developing vocabulary includes items that represent the following:

- *initial attitude* toward the other participants of the interaction;
- *changes in attitude* towards them as a result of the interaction.

Implementation and Future Work

Our program currently implements the first example we described and some of its variations. The architecture of the agents in our system is based on work on *opportunistic memory* (Hammond, 1989; Hammond *et al.*, 1989) and *agency* (Hammond *et al.*, 1990). We have mostly been concerned with the issues of plan representation (individual or multi-agent) and the *situated use* of those plans. We are currently working towards an implementation of the three stories described in this paper.

[6] Clearly, not every inter-agent relationship will evolve in the same way. The particular evolution (if any) will be based on the particular dynamics of the agents' interaction.

We believe that a theory of interactions is useful only if it can be used to *produce* actual multi-agent behavior. Thus, while we continue improving our vocabulary of engagement we expect to experiment heavily with our multi-agent system and show how evolving patterns of interactions serve to stabilize the agents' interactions over time. We expect our system not only to provide us with feedback useful to constrain and orient the development of our theories, but to help us test their plausibility as well.

Acknowledgements

We would like to thank Mitchell Marks, Paul Crawford, Daniel Fu and an anonymous reviewer for their useful comments on earlier drafts of this paper.

References

P. Agre. *The Dynamic Structure of Everyday Life.* PhD thesis, M.I.T., 1988.

P. Agre. What are plans for? In *New Architectures for Autonomous Agents: Task-level Decomposition and Emergent Functionality.* MIT Press, Cambridge, Massachusetts, 1990.

A. Bond and L. Gasser. An analysis of problems and research in DAI. In A. Bond and L. Gasser, editors, *Distributed Artificial Intelligence*, chapter 1. Morgan Kauffman, 1988.

C. Castelfranchi and M. Miceli. Dependence relations among autonomous agents. In *Proceedings of the Third European Workshop on Modeling Autonomous Agents in a Multi-agent World*, 1991.

C. Castelfranchi. Social power: a point missed in multi-agent, DAI and HCI. In Y. Demazeau and M. J., editors, *Decentralized AI.* Elsevier Science Publishers B.V. (North Holland), 1990.

E. Durfee and V. Lesser. Using partial global plans to coordinate distributed problem solvers. In *Proceedings of the 1987 International Joint Conference on Artificial Intelligence*, 1987.

L. Gasser and M. Huhns. Themes in Distributed Artificial Intelligence research. In L. Gasser and M. Huhns, editors, *Distributed Artificial Intelligence, Volume II.* Morgan Kauffman, 1989.

M. Georgeff. Communication and interaction in multi-agent planning. In *The Proceedings of the Third National Conference on Artificial Intelligence*, 1983.

M. Georgeff. A theory of action for multi-agent planning. In *The Proceedings of the Fourth National Conference on Artificial Intelligence*, 1984.

K. Hammond and T. Converse. Stabilizing environments to facilitate planning and activity: An engineering argument. In *Proceedings of the Ninth National Conference on Artificial Intelligence.* Morgan Kaufmann, July 1991.

K. Hammond, T. Converse, and M. Marks. Learning from opportunity. In *Proceedings of the Sixth International Workshop on Machine Learning*, Ithaca, New York, June 1989. Morgan Kaufmann.

K. Hammond, T. Converse, and C. Martin. Integrating planning and acting in a case-based framework. In *The Proceedings of the 1990 National Conference of Artificial Intelligence*, August 1990.

K. Hammond. *Case-based Planning: An Integrated Theory of Planning, Learning and Memory.* PhD thesis, Yale University, 1986. Technical Report 488.

K. Hammond. Opportunistic memory. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence.* IJCAI, 1989.

K. J. Hammond. Explaining and repairing plans that fail. *Artificial Intelligence*, 1990.

W. Hanks. *Referential Practice.* The University of Chicago Press, 1990.

H. Levesque, P. Cohen, and J. H. T. Nunes. On acting together. In *The Proceedings of the Eighth National Conference on Artificial Intelligence*, 1990.

K. E. Lochbaum, B. Grosz, and C. Sidner. Models of plans to support communication: An initial report. In *The Proceedings of the Eighth National Conference on Artificial Intelligence*, 1990.

S. P. Robertson, W. Zachary, and J. B. Black. *Cognition, Computing and Cooperation.* Ablex Publishing Corporation, 1990.

S. Rosenschein. Synchronization of multi-agent plans. In A. Bond and L. Gasser, editors, *Distributed Artificial Intelligence.* Morgan Kauffman, 1988.

R. Schank et al. A content theory of memory indexing. Technical Report 2, Institute for the Learning Sciences, Northwestern University, Evanston, IL, March 1990.

L. Suchman. *Plans and Situated Actions: The Problem of Human-Machine Communication.* Cambridge University Press, 1987.

E. Werner. Cooperating agents: A unified theory of communication and social structure. In L. Gasser and M. Huhns, editors, *Distributed Artificial Intelligence, Volume II.* Morgan Kauffman, 1989.

R. Wilensky. *Understanding Goal-Based Stories.* PhD thesis, Yale University, 1978. Technical Report 140.