

## Chapter 3

# MODELING SOCIAL RELATIONSHIP

### *An agent architecture for voluntary mutual control*

Alan H. Bond

*California Institute of Technology*

bond@cs.caltech.edu

**Abstract** We describe an approach to social action and social relationship among socially intelligent agents [1], based on mutual planning and mutual control of action. We describe social behaviors, and the creation and maintenance of social relationships, obtained with an implementation of a biologically inspired parallel and modular agent architecture. We define voluntary action and social situatedness, and we discuss how mutual planning and mutual control of action emerge from this architecture.

## 1. The problem of modeling social relationship

Since, in the future, many people will routinely work with computers for many hours each day, we would like to understand how working with computers could become more natural. Since humans are social beings, one approach is to understand what it might mean for a computer agent and a human to have a social relationship.

We will investigate this question using a biologically and psychologically inspired agent architecture that we have developed. We will discuss the more general problem of agent-agent social relationships, so that the agent architecture is used both as a model of a computer agent and as a model of a human user.

What might constitute social behavior in a social relationship? Theoretically, social behavior should include: (i) the ability to act in compliance with a set of social commitments [2], (ii) the ability to negotiate commitments with a social group (where we combine, for the purpose of the current discussion, the different levels of the immediate social group, a particular society, and humanity as a whole), (iii) the ability to enact social roles within the group, (iv) the ability

to develop joint plans and to carry out coordinated action, and (v) the ability to form persistent relationships and shared memories with other individuals.

Presumably, a socially intelligent agent would recognize you to be a person, and assign a unique identity to you. It would remember you and develop detailed knowledge of your interaction history, what your preferences are, what your goals are, and what you know. This detailed knowledge would be reflected in your interactions and actions. It would understand and comply with prevailing social norms and beliefs. You would be able to negotiate shared commitments with the agent which would constrain present action, future planning and interpretation of past events. You would be able to develop joint plans with the agent, which would take into account your shared knowledge and commitments. You would be able to act socially, carrying out coordinated joint plans together with the agent.

We would also expect that joint action together with the agent would proceed in a flexible harmonious way with shared control. No single agent would always be in control, in fact, action would be in some sense voluntary for all participants at all times.

To develop concepts and computational mechanisms for all of these aspects of social relationship among agents is a substantial project. In this paper, we will confine ourselves to a discussion of joint planning and action as components of social behavior among agents. We will define what voluntary action might be for interacting agents, and how shared control may be organized. We will conclude that in coordinated social action, agents voluntarily maintain a regime of mutual control, and we will show how our agent architecture provides these aspects of social relationship.

## 2. Our agent architecture

In this section we describe of an agent architecture that we have designed and implemented [3] [4] and which is inspired by the primate brain. The overall behavioral desiderata were for an agent architecture for real-time control of an agent in a 3D spatial environment, where we were interested in providing from the start for joint, coordinated, social behavior of a set of interacting agents.

**Data types, processing modules and connections.** Our architecture is a set of processing modules which run in parallel and intercommunicate. We diagram two interacting agents in the figure. This is a totally distributed architecture with no global control or global data. Each module is specialized to process only data of certain datatypes specific to that module. Modules are connected by a fixed set of connections and each module is only connected to a small number of other modules. A module receives data of given types from modules it is connected to, and it typically creates or computes data of other types. It may or may not also store data of these types in its local store. Processing by a module

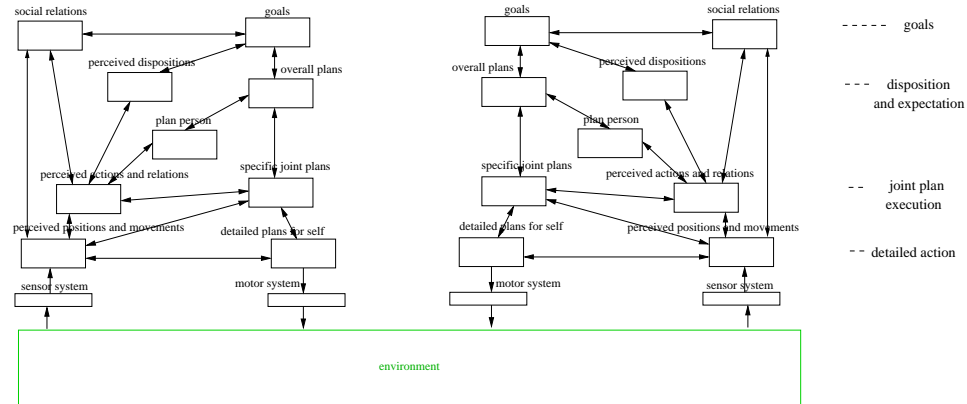


Figure 3.1. Our agent architecture

is described by a set of left-to-right rules which are executed in parallel. The results are then selected competitively depending on the data type. Typically, only the one strongest rule instance is allowed to “express itself”, by sending its constructed data items to other modules and/or to be stored locally. In some cases however all the computed data is allowed through.

**Perception-action hierarchy.** The agent modules are organized as a *perception-action hierarchy*. This is an abstraction hierarchy, so that modules higher in the hierarchy process data of more abstract data types. We use a fixed number of levels of abstraction.

There are plans at different levels of abstraction, so a higher level planning module has a more abstract plan. The goal module has rules causing it to prioritize the set of goals that it has received, and to select the strongest one which is sent to the highest level plan module.

**Dynamics.** We devised a control system that tries all alternatives at each level until a *viable* plan and action are found. We defined a viable state as one that is driven by the current goal and is compatible with the currently perceived situation at all levels. This is achieved by selecting the strongest rule instance, sending it to the module below and waiting for a *confirmation* data item indicating that this datum caused activity in the module below. If a confirmation is not received within a given number of cycles then the rule instance is decremented for a given amount of time, allowing the next strongest rule instance to be selected, and so on.

A viable behavioral state corresponds to a coherent distributed process, with a selected dominant rule instance in each module, confirmed dynamically by confirmation signals from other modules.

### 3. Social plans and joint action

We generalized the standard artificial intelligence representation of plan to one suitable for action by more than one collaborating agent. A *social plan* is a set of joint steps, with temporal and causal ordering constraints, each step specifying an action for every agent collaborating in the social plan, including the subject agent. The way an agent executes a plan is to attempt each joint step in turn. During a joint step it verifies that every collaborating agent is performing its corresponding action and then to attempt to execute its own corresponding individual action. We made most of the levels of the planning hierarchy work with social plans, the next to lowest works with a “selfplan” which specifies action only for the subject agent, and the lowest works with concrete motor actions. However, the action of these two lowest levels still depended on information received from the perception hierarchy.

**Initial model and a social behavior.** To make things more explicit, we’ll now describe a simple joint behavior which is a prototype of many joint behaviors, namely the maintenance of affiliative relations in a group of agents by pairwise joint affiliative actions, usually called grooming.

The social relations module contained a long term memory of knowledge of affiliative relations among agents. This was knowledge of who is friendly with who and how friendly. This module kept track of affiliative actions and generated goals to affiliate with friends that had not been affiliated with lately. Each agent had stored social plans for grooming and for being groomed. Usually a subordinate agent with groom and dominant one will be groomed. We organized each social plan into four phases, as shown in the figure: orient, approach, prelude and groom, which could be evoked depending on the current state of the activities of the agents. Each phase corresponded to different rules being evoked.

Attention was controlled by the planning modules selecting the agents to participate with and communicating this choice to the higher levels of perception. These higher levels derived high level perceptual information only for those agents being attended to.

### 4. Autonomy, situatedness and voluntary action

**Autonomy.** The concept of autonomy concerns the control relationship between the agent and other agents, including the user. As illustrated in our example, agents are autonomous, in the sense that they do not receive control imperatives and react to them, but instead each agent receives messages, and perceives its environment, and makes decisions based on its own goals, and that is the only form of control for agents.

Further, agents may act continuously, and their behavior is not constrained to be synchronized with the user or other agents.

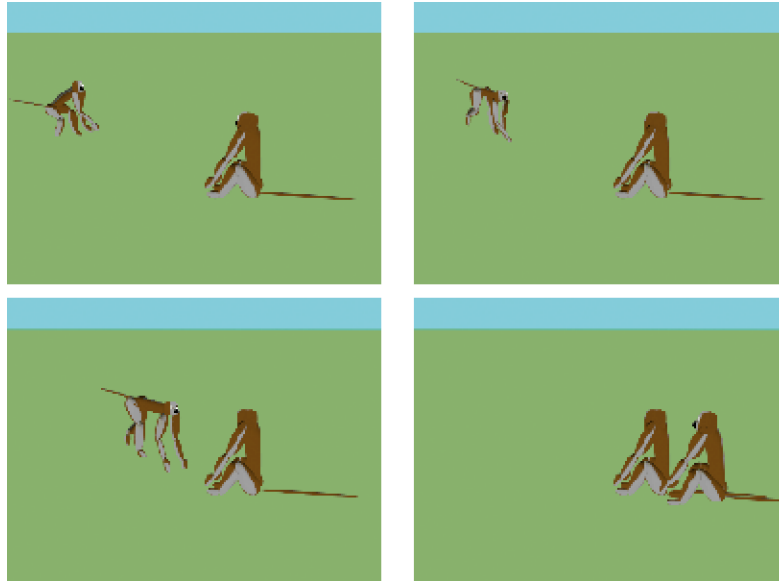


Figure 3.2. Four phases of grooming

**Constraint by commitments.** A social agent is also constrained by any commitments it has made to other agents. In addition, we may have initially programmed it to be constrained by the general social commitments of the social group.

**Voluntary control.** The joint action is “voluntary” in the sense that each agent is controlled only by its own goals, plans and knowledge, and makes its own choices. These choices will be consistent with any commitments, and we are thus assuming that usually some choice exists after all such constraints are taken into account.

**Situatedness of action.** However the action of each agent is conditional upon what it perceives. If the external environment changes, the agent will change its behavior. This action is *situated* in the agent’s external environment, to the extent that its decisions are dependent on or determined by this environment.

Thus, an agent is to some extent controlled by its environment. Environmental changes *cause* the agent to make different choices. If it rains, the agent will put its raincoat on, and if I stop the rain, the agent will take its raincoat off.

This assumes that the agent (i) does not make random and arbitrary actions, (ii) does not have a supersmart process which models everything and itself, in other words (iii) it is rational in the sense of using some not too complex reasoning or computational process to make its choices.

## 5. Mutual planning and control

Our agent architecture is flexibly both goal-directed and environmentally situated. It is also quite appropriate for social interaction, since the other agents are perceived at each level and can directly influence the action of the subject agent. It allows agents to enter into stable mutually controlled behaviors where each is perceived to be carrying out the requirements of the social plan of the other. Further, this mutually controlled activity is hierarchically organized, in the sense that control actions fall into a hierarchy of abstraction, from easily altered details to major changes in policy.

We implemented two kinds of social behavior, one was affiliation in which agents maintained occasional face-to-face interactions which boosted affiliation measures, and the other was social spacing in which agents attempted to maintain socially appropriate spatial relationships characterized by proximity, displacement and mutual observability. The set of agents formed a simple society which maintained its social relations by social action.

During an affiliation sequence, each of two interacting agents elaborates its selected social plan conditionally upon its perception of the other. In this way, both agents will scan possible choices until a course of action is found which is viable for both agents.

This constitutes mutual control. Note that the perception of the world by distal sensors is quite shared, however perception by tactile, proprioceptive, and visceral sensing is progressively more private and less shared. Each agent perceives both agents, which has some common and some private perception as input, and each agent executes its part of the joint action.

In each phase of grooming, each agent's social plan detects which phase it is in, has a set of expected perceptions of what the other may do, and a corresponding set of actions which are instantiated from the perception of what is actually perceived to occur. If, during a given phase, an agent changes its action to another acceptable variant within the same phase, then the other agent will simply perceive this and generate the corresponding action. If, on the other hand, one agent changes its action to another whose perception is not consistent with the other agent's social plan, then the other agent's social plan will fail at that level. In this latter case, rules will no longer fire at that level, so the level above will not receive confirmatory data and will start to scan for a viable plan at the higher level. This may result in recovery of the joint action without the first agent changing, however it is more likely that the induced change in the second agent's behavior will cause a similar failure and replanning activity in the first agent.

In the case of grooming, during orientation and approach, the groomee agent can move and also change posture, and the groomer will simply adjust, unless the groomee moves clearly away from the groomer, in which case the approach

behavior will fail. When the groomer arrives at prelude distance, it expects the groomee to be not moving and to be looking at him, otherwise the prelude phase will not be activated. Then, if the groomee make a positive prelude response, the groomer can initiate the grooming phase.

Agents enter into, and terminate or modify, joint action voluntarily, each motivated by its own perceptions and goals.

## 6. Coparticipation and engagement

Our notion of social plan has some subtlety and indirectness, which is really necessitated by the distributed nature of agent interaction. There is no agreed shared plan as such, each participant has their own social plan, which includes expectations of the actions of coparticipants. Each participant attempts to find and to carry out their “best” social plan which satisfies their goals. In constrained situations, it may be that the best social plan of each participant is very similar to the best social plans of coparticipants. Thus social plans of individuals may be more or less *engaged*. Engagement concerns the agreement and coherence among the instantiations of the social plans of the participants.

A standard example is the prostitute and the client, which coparticipate and cooperate, each with his or her own goals and social plan. Thus, for social action, the prostitute needs to sufficiently match the client’s social plan and model of prostitute appearance and behavior, and the client needs to behave sufficiently like the prostitute’s idea of a client.

Adversarial coparticipation occurs with lawyers representing defendent and plaintiff. Since however there is always a residual conflict or disparity and residual shared benefits in all relationships, it is difficult to find cases of pure cooperation or even pure adversality.

The initiation (and termination) of joint action usually involves less engagement between the social plans of coparticipants. The grooming preludes observed in social monkeys are for example initially more unilateral. Initiation and termination usually involve protocols by which coparticipants navigate paths through a space of states of different degrees of engagement.

In this model, social interaction is never unilateral. First, some “other” is always an imagined coparticipant. Second, even in the case of hardwired evolved behaviors, the behavior is intended for, only works with, and only makes sense with, a coparticipant, even though, in this case, there is no explicit representation of the other. It is not clear for example what representation, if any, of the mother a baby may have. There is for example biological evidence of tuning of the babies sensory systems during pregnancy, and immediately after birth, to the mother’s odor and voice. Thus, the mother constructs an explicit coparticipant and the baby acts as if it has a coparticipant.

## 7. Summary

We argued for and demonstrated an approach to social relationship, appropriate for agent-agent and user-agent interaction: *In a social relationship, agents enter into mutually controlled action regimes, which they maintain voluntarily by mutual perception and by the elaboration of their individual social plans.*

## References

- [1] A. H. Bond and L. Gasser, "An Analysis of Problems and Research in Distributed Artificial Intelligence," in *Readings in Distributed Artificial Intelligence*, pp. 3–35, San Mateo, CA: Morgan Kaufmann Publishers, 1988.
- [2] A. H. Bond, "Commitment: A Computational Model for Organizations of Cooperating Intelligent Agents," in *Proceedings of the 1990 Conference on Office Information Systems*, pp. 21–30, April 1990.
- [3] A. H. Bond, "Describing Behavioral States using a System Model of the Primate Brain," *American Journal of Primatology*, vol. 49, pp. 315–388, 1999.
- [4] A. H. Bond, "Problem-solving behavior in a system model of the primate neocortex," *Neurocomputing*, vol. to appear, 2001.