A COMPUTATIONAL MODEL FOR ORGANIZATIONS OF COOPERATING INTELLIGENT AGENTS [1]

Alan H. Bond[2]
Concurrent Computation Program,
California Institute of Technology
Pasadena, California 91125.

Abstract

A computational model is developed, which captures properties of relationship and organization in sets of distributed intelligent agents. The model is inspired by sociological theory, and uses centrally a concept of commitment.

Symbolic interactionist concepts in sociology are reviewed from the perspective of Distributed Artificial Intelligence. This includes the work of Anselm Strauss and Elihu Gerson.

A formulation of the concept of commitment is proposed. The concept is propagated into notions of agent and organization.

A formulation and computational model of the basic notion of commitment is given, which represents commitments as mutually agreed constraints on action, belief and world state.

The issues of agent integrity, plans and resources, and their representation, are then discussed, and computational mechanisms defined.

An organization is defined as a set of agents with consistent mutual commitments. A mechanism is defined to check and enforce the consistency of these commitments.

[1]Presented at COIS90 - Conference on Office Information Systems, Sponsored by ACM-SIGOIS and IEEE Computer Society TC-OA, In cooperation with IFIP 8.4 MIT, Cambridge, Massachusetts, April 25-27, 1990

[2]email: bond@cs.ucla.edu

1 Introduction

In its present stage of development, DAI [BG88] is seeking new concepts which will allow DAI problems to be better expressed and better solved. To this end, we examine the concepts of symbolic interactionism with a view to finding inspiration and insights into the description of organizations of cooperating agents. Goals of our research include the development of concepts, a formal model, and an implementation of a computer system which would allow organizations of intelligent agents to be programmed. Such an implementation would:
1. provide the ability to create model organizations for contemplation and study.
2. allow test cases and applications to be programmed as organizations of intelligent agents.
3. provide a support tool for human organizations. Each computer agent would solve problems, maintain goals, etc., as support for a human agent.
We are also interested in organizations which reorganize in solving problems, using global organizational principles. This is related to load balancing and parallel computer systems.

2 Symbolic interactionism

George Herbert Mead [Mea34] argued that it is social interaction that gives rise to the minds and selves of agents, and that human behavior consists of perceptions and actions which are continually constructed by the self. Joint actions are collective forms of action constituted by fitting together of the lines of behavior of the separate participants.

2.1 Anselm Strauss and negotiated order

Although the concept of negotiated order is already in Mead, and was used by some prior researchers, it was Strauss who developed explicitly the notion of negotiated order from a study of the organization of two psychiatric hospitals. Strauss [SFSW85] summarizes the main points derived from this study:
(i) Social order is negotiated order. In the organizations studied, apparently there could be no organizational relationships without accompanying negotiations.
(ii) Negotiations were patterned; they were contingent on specific structural conditions, viz., who negotiated with whom, when and what about, etc.; they had different characters ; they had different consequences for agents and organizations.
(iii) The products of negotiation, viz., contracts, understandings, rules, and so on, all had temporal limits, and were eventually reviewed, reevaluated, revised, revoked or renewed.
(iv) Negotiated order and the bases of concerted action needed to be continually reconstituted; new ones were made daily.
(v) The negotiated order at any particular time could be conceived as the sum total of the organization's rules and policies, together with whatever agreements, understandings, pacts, contracts and other working arrangements currently obtained. These included formal and informal agreements at every level of the organization.
(vi) Any changes impinging on the negotiated order, including any new event, and disruption or any new agent, called for renegotiation and reappraisal.

(vii) The reconstitution of order could be conceived as a complex relationship between a daily negotiation process and a periodic appraisal process, and between the more stable elements of organizational order and more fleeting working arrangements.

The term "negotiation" in Strauss has a very broad and inclusive meaning, to include any type of reaching of agreement such as bargaining, compromising, making arrangements, getting tacit understandings, exchanging, engaging in collusion and so on. These terms seem to the author to all involve:
(i) interaction or communication
(ii) where an agreement is reached
(iii) and as a result each agent may be changed, and also may have to adjust and internally reintegrate to the agreement.
If all three are present, we can say that negotiation occurs. [Since the word 'negotiation' is now so broad and imprecise, and since there is only one activity which structures the social organization and the selves of agents, we could just refer to it as *the (social) process*.]

2.2 Gerson's 'quality of life' model and commitments

The notion of commitment was first developed explicitly in sociology by Becker [Bec60]. Becker's description is that individuals participate in several organizations or settings and hence to regard their behavior in any one setting as consistent lines of activity we must introduce the notion of commitments (or 'side-bets', as he images it) consequent on the individual's participation in other settings. These side-bets constrain the individual's actions and also can be used explicitly in negotiation in the setting. Gerson [Ger76] produced his 'quality of life' social model as a synthesis of several ideas, and with as basis the Meadian tenet that self and society are aspects of the same social process. Gerson's 'quality of life' model has the following elements, expressed by quotations from his paper:

1. "The self consists of a series of negotiated working arrangements for the conduct of affairs in a variety of situations"
2. "..selves are constructed as part of a cooperative process which takes place in specific working contexts"
3. "participation in any situation is, therefore, simultaneously *constraining* in that people make contributions to it, [are] bound by its limitations, and yet *enriching* [my italics], in that participation provides resources and opportunities otherwise unavailable"
4. "Participation .. provides some distribution of constraints and resources"
5. "The pattern of situations in which an individual participates sets the degree to which resources are transferable among settings"
6. "I shall call the overall organization of commitments associated with any delimitable social object the *sovereignty* of that object"
7. " ..the character and nature of this joint participation defines the nature of the *flows of commitment* among settings which constitute their relationship to one another"

8. " 'what is negotiated', therefore, are patterns of commitment organization, as expressed in flows of resources and constraints upon them"
9. "commitments and their organization must be expressed in terms of resources and constraints upon their use".
10. The resources considered include
    (a) Money
    (b) Time, commitments are called *schedules*
    (c) Skill, commitments are called *technologies*
    (d) Sentiments, commitments are called *solidarities* [my plural]
11. "..individuals and society (generate) each other via a continuing process of negotiation"
12. "..patterns of commitment (are) measured as the joint allocation of money, time, skill and sentiment by individuals and settings".
13. "..the ancient problem of balancing off the good of the individual and the good of society-as-a-whole is subsumed under a different problem, that of creating and managing a pattern of negotiations, which is viable .."

3 The concept of commitment

The word 'commitment'. Although the word 'commitment' in English (and American), means (i) to give in charge or entrust, or (ii) to pledge a certain course of action, it originates from the Latin cum - with, mittere - to send, and indeed committere - to bring, join or combine two or more objects into one whole, to connect or unite. This to the author suggests the underlying meaning of a constraint or binding, which further is united and integrated into the world producing a change of world view. This incidentally is consistent with the standard explanation of a specialized meaning of 'commit', viz., 'to perpetrate', via the notion of collusion.

Commitments as actions and beliefs. The constraints involved can be of various types, notably to follow a certain course of action, or to hold certain beliefs. A commitment concerns either acting in a certain way, conditional upon circumstances, or it can be a commitment to hold a certain belief. *Prospective* commitments concern the future, *retrospective* commitments concern the past. A retrospective commitment is a set of beliefs about past events. Commitments of course may be *conditional* on the state of the world, including the time. A commitment concerning another agent's action or belief will be called an *expectation*. An agent may have the expectation of some *unspecified* agent providing resources.

Commitments are social. However, a commitment is more than a constraint, it involves a trust, intention or belief. It is thus a relationship with other agents that the constraint be held to. It is also one thing to make a commitment, in the gloom of one's study, and another to make other agents aware of it. As in the denouement of Dr. Strangelove, the whole point of creating the Doomsday machine was to tell everyone about it!

Commitments as (social) goals. Further, a commitment to future action is really to a goal. Thus it is to attempt and solve the goal by performing articulation work, i.e., problem solving activity. This activity presumably is not completely unrestrained but within agreed social limits. In some cases, a goal can persist and hence a commitment could continually evoke certain actions to maintain the satisfaction of the goal.

Commitments involve resources. As we can see from Gerson's thinking, commitments for him always involve resources, they are constraints on resources. However, for him, resources involve not only basic resources and time but knowledge, sentiment and almost anything else that the agent needs. Hence, any action commits resources. A current action commits and may consume resources. A future committed action commits to the use of resources at a future time.

Commitments involve agent integrity. It seems to the author that an essential point is that agents can and do trust each other. A commitment is not just a prediction that a certain constraint, action or belief, will be adhered to, but it is an undertaking of modification of world view and of doing whatever it reasonably takes, to adhere to it. As such it makes contact with the concept of agent as reasonable, predictable, reliable and socially aware. It implies that agents have some well-formedness in their beliefs and actions. In other words, for the notion of commitment to be applicable, the agents must satisfy some general integrity and sociability properties.

Commitment and natural language pragmatics. A discussion is given by Winograd and Flores [WF87], who lean on Searle [Sea69]: Speech acts may be illocutionary acts which have five possible *illocutionary points*:
(i) assertives - commit the speaker to something's being the case
(ii) directives - attempt to get the hearer to do something
(iii) commissives - commit the speaker to some future course of action
(iv) expressives - express a psychological state of affairs, e.g., apologizing
(v) declarations - establish correspondence between a proposition and reality, e.g., pronouncing a couple married.
Speech acts also have a *propositional content* and an *illocutionary force*. "The centrality of commitment in speech act theory has been brought out particularly clearly by Habermas [Hab79]. Habermas argues that every language act has consequences for the participants, leading to other immediate actions and to commitments for future action. In making a statement, a speaker is doing something like making a promise - making a commitment to act in appropriate ways in the future" [WF87] p. 59.

4 A DAI notion of commitment

We now present a formal DAI notion of commitment, which we hope captures most of the sociological concept of commitment. We shall ramify the concept into a view of DAI agents and organizations of agents.

Social agents. The main idea is to program agents in terms of their commitments to each other. Thus in the description of an agent, we have
(i) actions it will perform and resources needed
(ii) beliefs it will hold

(iii) expectations of actions of others and resources supplied by them
(iv) the supply of resources to others by their holding expectations
An agent has explicit representations of the resources it is generating and their benefit to others, and the resources it needs and their supply and support by others. We can formally represent a commitment as a logical formula, which can be a goal, belief or action. Associated with a formal commitment will be the specified resources for the commitment.

The main points.

1. *Actions* can be on data or the state, or can be communication acts.
2. *Beliefs* can be about the world, about the self or about other agents.
3. *Resources* can be *basic resources*, such as storage, processor time, communication bandwidth, etc., or *knowledge*, or *results* or *commitments*.
4. A *commitment* can be an action, a belief or a goal. Each must have a specification of associated resources. Any may be conditional on the state.
5. Each action involves all the resources that it needs. Any action commits resources. A use of a resource may also consume that resource.
6. The basic activity is the *execution* of commitments. If all resources are met, then a commitment can and probably should be executed.
7. To execute an action is to perform the action, to execute a belief is to assert it, i.e., believe it, to execute a goal is to attempt it. Notably a belief can concern another agent's action.
8. A current action commits and consumes resources. A future committed action commits to the use of resources at a future time.
9. One type of action, and therefore type of commitment, is to *make* a commitment. This is usually a future commitment, but it can be the establishment of a retrospective belief.
10. One type of belief, and therefore commitment, is that a specified agent has made a specified commitment; this is an *expectation*.
11. We can thus define a concept of *agent* as
    (a) a locus of control
    (b) which makes and executes commitments.
12. The AI notion of *plan* can be generalized to that of a set of related commitments.
13. An agent has *knowledge* of methods (or theorems) which match to situations, including those involving goals, and which evoke plans. We can imagine a simple control structure similar to Planner or Prolog, but this could be changed.
14. An agent should probably communicate its commitments truthfully to affected agents, or at least respond truthfully to requests for information concerning its commitments.
15. An agent receives not only data, results, knowledge and other resources from other agents, but also commitments.

16. An agent has *integrity constraints* on its commitments; they must be consistent with each other.

4.1 Organizations of cooperating agents

We can now define a software, DAI, concept of organization. The concept of organization has been variously defined and has variously eluded definitive conceptual representation. We shall use a general abstract concept, in which
(i) there is no predefined structural type, but rather any pattern of relationships, which dynamically changes.
(ii) the relationships are any sets of agreements on activity and resource use.
An *organization* will thus be defined as a set of agents with mutual commitments. These commitments should be integrated and consistent:
(a) Their beliefs about the state of the world should be compatible:
if a1 believes statement x1 in time interval i1, under world conditions w1
and a2 believes statement x2 in time interval i2, under world conditions w2
then, if i1 and i2 overlap, and w1 and w2 overlap in this interval overlap, then x1 and x2 should be consistent with each other under these overlapped conditions in the overlapped interval.
(b) if a1 has a commitment c1 of the form commitment(a2,c2), then a2 should have a commitment c2' which subsumes c2, or else the set of a2's commitments should be consistent with a1's expectation of a2.

5 Formulating the basic notion of commitment

In order to work with the notion of commitment, we shall distinguish among three related but different problems: (i) *expression* of commitments, how commitments are to be represented, (ii) *generation* of commitments, and (iii) *negotiation* of commitments. In this paper, we shall be concerned almost exclusively with the first of these problems. In our judgement, many DAI mechanisms and systems can be represented by commitments which are generated by simple globally agreed mechanisms. The important and difficult issue of negotiation merits separate discussion [Syc87].

5.1 Formulation of commitments

We shall represent commitments in the main by constraints. We believe that a wide and interesting range of systems can be described using constraints on states, actions and beliefs. This could be extended to constraints on events and constraints on commitments themselves. First, we need to define a ground or *uncommitted agent*, which has as yet made no commitments. For this, we shall take a socially uncommitted agent to also lack plans, and to behave opportunistically. Thus, a socially uncommitted agent will have: (i) a script - a set of conditional actions, (ii) a set of beliefs, and (iii) behavior which follows an opportunistic regime.

Our well-defined deterministic interpretations are straightforward. They are as follows:
1. **Commitment to act.** Add a new (conditional) action to script.

2. **Commitment to believe.** Add a new belief.
3. **Constraint on action.**
(i) Before every action, test if forbidden,
(ii) and in every plan, test if forbidden
(iii) and in every action which constructs a planned action, check that forbidden action is not planned.
4. **Constraint on state.**
(i) Before changing state, test if forbidden
(ii) and in every plan, test if states are forbidden
(iii) and in every action which constructs a planned action, check that forbidden state is not planned.

5.2 Commitment is a social relationship with other agents

Here we introduce some strong constraints on commitments. The computational model would ensure that they are adhered to. They are part of the semantics of the representation system we are developing.
1. Commitment is maintained by negotiation with another agent, or set of agents, and it is labeled with this agent set.
2. Agents have knowledge of all commitments made that involve them, including knowledge of which other agents are involved.
3. Any attempt to change a commitment must be negotiated with the agents involved. A commitment cannot be changed unilaterally.

5.3 Computational model for the basic notion of commitment

Modeling the basic constraint notion. Since integration occurs by separate integration mechanisms, to be discussed in section 6, we just need to model and implement the basic constraint representation, and enforcement and solution mechanisms. To deal with constraints, we need at least to *enforce* them. We shall ensure that all relevant constraints for an agent are known to the agent, and, in addition, the system will enforce constraints in its interpreter. Merely enforcing constraints is not very satisfactory, and what is needed in addition are constraint *solvers*. Given constraints on action, world state and belief, the system should provide solvers which find sets of solutions for which actions, which world states and which beliefs the agent may have. Since constraints are on states, beliefs and actions, we can implement enforcement and solution using a *metainterpreter*.

Modeling the social relationship.
1. The form of a commitment includes information on the agents involved.
commitment(commitment type, actual commitment formula,{agent set})
where commitment type is action, state or belief.
2. The action of committing to a commitment is performed by the system. commit(commitment) commits the agent. Making the commit action passes this information to other agents in the agent set, and makes them aware of this commitment.
3. the action of removing a commitment. uncommit(commitment) must be received by all agents in the agent set of the commitment, and agreed by them.

6 Integrity, plans and resources

6.1 Agent integrity

As we argued, from common sense rather than sociological theory, in section 3, agents should have integrity. We can formally interpret this requirement as:
1. A consistent set of rules
2. Plans justified in terms of resources
3. Beliefs consistent
It is clear that the making arbitrary commitments by an agent can result in non-optimal if not pathological behavior of the agent. Thus, a commitment to avoid certain actions will, with only the basic apparatus of our interpreter, result in "almost" doing such actions, and then hopefully finding other nonforbidden actions, to proceed. Similarly for commitments to avoiding, and for maintaining, certain states and beliefs. Thus, on changing commitment, an agent should be *reintegrated*.

6.2 Integration mechanisms

We have, so far, identified three classes of integration mechanism, viz., *continuous replanning*, *knowledge revision* and *resource justification*. If the agents use plans for action, the search for actions and states that satisfy all their commitments can be done during planning, and not during real-time action. And further, using plans as parts of other plans means that the search for the subplan is only done once. Thus plans themselves provide some integrity. Then, the agent can continuously replan as commitments change, thereby reintegrating the new set of commitments with its plans. The agents' knowledge should be continuously revised as commitments change. This can be done in at least two ways. First, the performance parameters for the knowledge base should be altered to alter the precedence of suggestions, actions, etc. Second, beliefs themselves may be altered, to reintegrate with changed beliefs committed to.

6.3 Plans

A plan will be a set of commitments which is treated as a unit of behavior. Thus a plan can be evoked, in which case the entire set of commitments which constitutes the plan comes into effect. When the plan terminates, these commitments are removed. Not all actions are socially agreed actions, but they are constrained by whatever commitments are currently held by the agent. Plans are thus joint plans to the extent that they are constrained by commitments to other agents. Note that a change of belief is not necessarily a change of commitment. Only if it comes into conflict with a committed belief and requires a change in the committed belief, will it evoke renegotiation. An entire plan can be agreed to by a set of agents. This means that changes of commitment may not have to be negotiated since they are preapproved. This of course is the same as other agents making commitments as to how they will negotiate in the future, so it involves no new primitive idea. Thus, if a set of agents makes a plan, this may involve a set of changes of constraint which have been agreed upon prior to the evocation of the plan. In particular, at the termination of the plan, the set of commitments evoked as the plan is removed, without the necessity to negotiate this change in commitment. The commitments to allow these changes are made at the time that

the joint plan is evoked, not when the plan is created. The action to evoke the plan is also a commitment which is negotiated.

6.4 Resources

When a new action, condition or belief is added to the agent's script, we need to allocate resources, or set goals to find resources. More generally:
1. An agent has a set of plans, with justification (search graph) of resources needed.
2. On change of commitment, an agent must revise all plans, and again justify resources needed.
For every action, there is a set of resources required to carry out that action. These are part of the belief system of the agent. A conditional action is only well-formed if it contains provision for obtaining the resources needed for the action. "Obtaining" is here meant to include simply checking that resources are in place, and also actions for obtaining resources and goals for obtaining resources. This structure of beliefs, actions and goals is the justification of the resources needed for the action or commitment. The same idea applies to plans.

Representing resources and their use. In our model, each action has an associated set of resources needed, and hence a set of associated goals to obtain these resources. In decision making, we can then articulate a goal tree to a point of certainty, and choose an optimal path.

6.5 Computational model for plans and resources

The model has now to be extended to handle plans as sets of commitments and resource checking.
1. A set of agents may agree on a plan, and this preapproves a set of commitments, which hold while the plan is active. Thus a plan is a set of commitments which can be evoked as a unit, and revoked as a unit. The system should support this construct.
2. On changes of commitment, plans should be replanned to ensure some agent integrity. The system should notify the agents involved that replanning is recommended.
3. The system should enforce well-formedness of commitments, and plans, in terms of resources. Resources would have to be declared as such. An accounting of resources required and how these resources are to be provided, would be carried out by the system. At the least, goals to acquire resources must be in place.
4. It would be possible to provide standard planning mechanisms to provide planning with resource management, for example, a decision tree search with check out for resource provision and optimization for resource use.

7 Organizations

Organizations of cooperating agents. An *organization* will be defined as a set of agents with mutual commitments. In addition, there may be global commitments, which apply to the entire set of agents in the organization. The set of commitments, held by the agents in the organization, should be integrated and consistent.

Computational model for organizations. The computational model simply has to check the consistency of the mutual commitments of the agents in the organization, and to enforce global commitments. The check for consistency is as outlined in section 3, and it should be performed after each change of commitment in the system. A solver should also be provided which at least constructs a description of the commitment disparity for the agents, but should also provide a set of solutions for the set of mutual commitments that is consistent.

Acknowledgements
I am very grateful to my colleague and friend, Professor Les Gasser of University of Southern California, for continuing discussions and collaboration on DAI issues, and for introducing me to the concept of commitment, and for pointing out Eli Gerson's paper. This research was supported by the Caltech Concurrent Computation Program, directed by Professor Geoffrey Fox, to whom I am also very grateful.

References

[Bec60] Howard S. Becker. Notes on the concept of commitment. *American Journal of Sociology*, 66:32–40, 1960.

[BG88] Alan H. Bond and Les Gasser. *Readings in Distributed Artificial Intelligence.* Morgan Kaufmann Publishers, San Mateo, CA, 1988.

[Ger76] Elihu M. Gerson. On 'quality of life'. *American Sociological Review*, 41:793–806, 1976.

[Hab79] Jurgen Habermas. *Communication and the Evolution of Society.* Beacon Press, Boston, 1979.

[Mea34] George Herbert Mead. *Mind, Self, and Society.* University of Chicago Press, Chicago, 1934.

[Sea69] John R. Searle. *Speech Acts: An Essay in the Philosophy of Language.* Cambridge University Press, 1969.

[SFSW85] Anselm L. Strauss, Shizuko Fagerhaugh, Barbara Suczek, and Carolyn Wiener. *The Social Organization of Medical Work.* University of Chicago Press, Chicago, IL, 1985.

[Syc87] Ekaterini P. Sycara. *Resolving Adversarial Conflicts: An Approach Integrating Case-Based and Analytical Methods.* PhD thesis, School of Information and Computer Science, Georgia Institute of Technology, 1987. Also Technical Report GIT-ICS-87-26, Georgia Institute of Technology.

[WF87] Terry Winograd and Ivan Flores. *Understanding Computers and Cognition.* Addison Wesley, Reading, Massachusetts, 1987.