Social Conceptions of Knowledge and Action:

DAI Foundations and Open Systems Semantics

Les Gasser

Distributed Artificial Intelligence Group
Computer Science Department
University of Southern California
Los Angeles, CA 90089-0782

Paper prepared for

Artificial Intelligence Journal

Special Issue on Foundations of Artificial Intelligence

January 1991

USC Distributed AI Group Research Note 54

Abstract

This article discusses foundations for Distributed Artificial Intelligence (DAI), with a particular critical analysis of Hewitt's *Open Information Systems Semantics* (OISS). The article sets out to do five things:

- It presents a brief overview of current DAI research including motivations and concepts, and discuss some of the basic problems in DAI.
- It introduces several principles that underly a fundamentally multi-agent (i.e., *social*) conception of action and knowledge for DAI research. These principles are introduced to provide definitions, to delimit the discussion of OISS and as background against which to assess its contributions.
- It analyzes the main points of OISS in relation to these principles.
- It shows how attention to these principles can strengthen OISS approach to foundations for DAI.
- It traces some of the implications of this synthesis for theorizing and system-building in AI.

The OISS approach productively challenges some conceptions of knowledge, reasoning, and action in classical AI research. However, it sometimes ignores the sophistication and richness of contemporary DAI research. Several of the key concepts of OISS are not clearly enough defined or operationalized, and the article points out several ways to strengthen the OISS approach.

All real systems are distributed.
- F. Hayes-Roth, in [Davis80]

1 Introduction

Artificial Intelligence research is fundamentally concerned with the intelligent behavior of machines. In attempting to create machines with some degree of intelligent behavior, AI researchers model, theorize about, predict, and emulate the activities of people. Because people are quite apparently social actors, and also because knowledgeable machines will increasingly embedded in organizations comprising people and other machines, AI research should be concerned with the social dimensions of action and knowledge as a fundamental category of analysis. But current AI research is largely *a-social*, and because of this, it has been inadequate in dealing with much human behavior and many aspects of intelligence.

In most contemporary AI research and practice, the unit of analysis and of development is a computational process with a single locus of control, focus of attention, and base of knowledge—a process organization inherited from von Neumann computer architectures and from psychology. While it is becoming easier to implement such a process as a concurrent system using an underlying distributed processing layer or a parallel language (such as concurrent prolog or lisp) the basic mechanisms of reasoning and problem solving generally remain bound to a single, monolithic conception of knowledge and action. Recently, however, there has been a revival of interest in approaches to analyzing and developing intelligent "communities" which comprise *collections of interacting, coordinated knowledge-based processes.* The body of research that deals with this problem-level concurrency in AI systems has come to be known as *distributed artificial intelligence* (DAI). Researchers in DAI are concerned with understanding and modeling action and knowledge in collaborative enterprises.

DAI research provides a very rich ground for re-examing some of the premises and formalisms upon which notions such as representation and reasoning, or knowledge and action, are classically located. This article analyzes Open Information Systems Semantics (OISS) from the standpoint of contemporary research in DAI, and inherently *social* conceptions of knowledge and action (actually, *interaction*). For a statement of OISS we draw mostly upon [Hewitt91], which is the most recent statement in a larger coherent body of research. Since a short discussion of elements of this larger body is presented in the appendix of that paper, we also occasionally draw on [Hewitt77, Kornfeld81, Hewitt84, Hewitt85, Hewitt86].

We will present this investigation in several parts. First, since OISS has been proposed as a new foundation for DAI, but its differences from existing DAI foundations and research are not apparent, we briefly examine contemporary research in "classical" DAI. Following that, we introduce some principles which are desiderata for an inherently social conception of knowledge and action, consistent with the premises of open DAI systems. Next, to examine the impact of OISS, we examine it in the light of these principles and current DAI research. Finally, we discuss several ways to strengthen and extend OISS.

2 The Current State of DAI Research

There are many reasons for wanting to distribute intelligence or cope with multi-agent systems. In some domains, (e.g., distributed sensing, medical diagnosis, air-traffic control), knowledge or activity is inherently spatially distributed. The distribution can arise because of geographic distribution coupled with processing or data bandwidth limitations, because of the natural functional distribution in a problem, because of a desire to distribute control (e.g., for fail-soft degradation), or for modular knowledge acquisition. Other reasons for distribution include adaptability, reduced cost, ease of development and management, increased efficiency or speed, history, needs for isolation or autonomy, naturalness, increased reliability, resource limitations, and specialization. Opportunity is a second reason for studying DAI systems. Hardware and software mechanisms for distributing and controlling the interaction of multiple processes have begun to reach maturity, in both shared-memory and distributed-memory multicomputer ensembles. Third, there is interest in integrating existing AI systems to gain power and to leverage capability, which necessarily means coping with problems of discrepancies in representation and design. Fourth, problems are sometimes simply too large or complex to solve by single processes, for reasons of semantic representation as well as computational power; distributed approaches may provide solutions. Finally, it is an empirical observation that most[1] human activity involves more than one person. As researchers have tried to understand and model human problem solving and intelligent behavior, they have begun to take this observation more seriously as a foundation for theories (See, e.g., [Chandrasekn81, Wesson81]).

Research in DAI promises to have wide-ranging impacts in basic AI research (problem representations, epistemology, joint concept formation, collaborative reasoning and problem solving), cognitive science (e.g., mental models, social cognition), distributed systems (reasoning about knowledge and actions in distributed systems, architectural and language support for DAI), the engineering of AI systems ("cooperating expert systems," distributed sensing and data fusion, cooperating robots, collaborative design problem solving, etc.), and human-computer interaction (task allocation, intelligent interfaces, dialogue coherence, speech acts). As Nilsson has pointed out [Davis80], DAI research is attractive for fundamental reasons: to coordinate their actions, intelligent agents need to represent and reason about the knowledge, actions, and plans of other agents. DAI research can help to improve techniques for representing and using knowledge about beliefs, action, plans, goals, etc., as well as helping us to discover the extent to which, when analyzed *from the outside in*—from the social to the individual—these concepts are useful or necessary.

2.1 Basic Research Problems in DAI Literature

Since OISS proposes to create new foundations for DAI, what are the existing foundations, and how adequate are they? In characterizing the recent state of DAI research, Bond and Gasser developed six basic problems that current DAI systems had begun to address [Bond88a]. These six problems are inherent to the design and implementation of any system of coordinated problem solvers, whether in open or closed worlds.

Here we give a brief exposition of these problems, each of which appears in some form in all DAI application domains. Greater detail and more citations can be found in [Bond88b, Gasser90]. The

[1] Or rather, all—cf. "taking the role of the other" in [Mead34].

problems include:

- *How to formulate, describe, decompose, and allocate problems and synthesize results among a group of intelligent agents.* Most approaches to these issues rely on designers. Work on the Contract Net and DVMT systems has provided mechanisms for flexible decomposition and allocation [Davis83, Durfee87a]. Many bases for decomposition have been suggested, including abstraction levels, functional, data, or control dependencies, and interaction density. Little work has been done on automated problem formulation and description, but see [Hinke90].

- *How to enable agents to communicate and interact: what communication languages or protocols to use, and what and when to communicate.* The major approaches here include formalized interaction and negotiation protocols such as the Contract Net Protocol and Partial Global Plans (PGPs) [Durfee87a, Smith80], Lenat's scheme based on common agent structures, and planned communications based on reasoning about the knowledge states of agents [Cohen79, Rosenschein82].

- *How to ensure that agents act coherently in making decisions or taking action, accommodating the nonlocal effects of local decisions and avoiding harmful interactions.* This has been treated as the major problem of DAI research. Primary approaches include establishing organization, improving local awareness and skill [Durfee87b], multi-agent planning [Georgeff87], abstraction, and resource-directed coherence [Kornfeld81].

- *How to enable individual agents to represent and reason about the actions, plans, and knowledge of other agents in order to coordinate with them; how to reason about the state of their coordinated process (e.g., initiation and termination).* Principal approaches include the use of utility theory and game theory to represent rational choice [Rosenschein85], symbolic models of agents' capabilities and roles [Gasser87], belief models [Cohen87], and graph models of organizational relationships [Durfee87a, Wesson81]. Approaches to system behavioral modeling and analysis have been presented in [Hudlicka87].

- *How to recognize and reconcile disparate viewpoints and conflicting intentions among a collection of agents trying to coordinate their actions.* Main approaches include assumption surfacing using ATMS techniques [Mason89], parallel falsification and microtheories [Kornfeld81, Hewitt86], partial global planning [Durfee87a], knowledgeable mediation, standardization, and various approaches to negotiation [Durfee89, Sathi89, Sycara89]. Star [Star89b] provides another promising characterization based on malleable "boundary objects" with dual semantics.

- *How to engineer and construct practical DAI systems; how to design technology platforms and development methodologies for DAI.* Numerous technology platforms have been built, including testbeds such as the DVMT [Durfee87b] and MACE [Gasser87]; integrative systems such as ABE [Erman88], reflective, concurrent object-based languages such as Mering-IV [Ferber88], and blackboard or distributed blackboard systems such as GBB [Corkill86], BB1 [Hayes-Roth85], CAGE/POLIGON [Nii89], etc.

Solutions to these problems are intertwined. For example, different procedures for communication and interaction have implications for coordination and coherent behavior. Different problem and task decompositions may yield different interaction or agent-modeling requirements. Coherent, coordinated behavior depends on how knowledge disparities are resolved, which agents resolve them, and so on.

Solving these problems in a fundamental way is quite complicated. Some of the basic DAI problems (e.g., some problems of disparate representations) can be designed away in a carefully engineered DAI system, by analyzing key questions such as what kinds of communication protocols to use, which conflicts are to be settled by people outside the system and which are to be handled autonomously by the system itself, how different conflicts interact, and how their settlement will be coordinated. In any case, we reveal the current foundations of DAI by examining how researchers have stated and solved these problems.

While these six problems appear widely in the literature, most extant theoretical and experimental solutions to them go only part way toward a basic scientific account of multi-agent collaboration, because they have not grappled with several other more basic issues. Virtually all current approaches to these problems are based on common interagent semantics with at most one or two meta or contextual levels, correspondence theories of representation and belief, global measures of coherence, and the individual agent as the unit of analysis and interaction. Most DAI experiments and theories depend upon closed-system assumptions such as common communication protocols, a shared global means of assessing coherent behavior, some ultimate commensurability of knowledge, or some boundary to a system. While the six problems presented above still provide much fruitful ground for study, solving them still would not provide an adequate foundation for DAI. To make substantial theoretical progress, we must begin to lay firmly social foundations for DAI research. As a framework in which to analyze the OISS proposal and its decentralized foundations, let us discuss some desiderata of a social framework for DAI.

3 Social Conceptions of Knowledge and Action for AI

DAI systems, as they involve multiple agents, are *social* in character; there are properties of DAI systems which will not be derivable or representable solely on the basis of properties of their component agents. We need to begin to think through and articulate the bases of knowledge and action for DAI in the light of their social character. Here, we suggest and briefly discuss several principles that ought to underly the scientific and conceptual foundations for DAI systems from a social perspective. Since theories that support the construction of DAI systems ought to follow these principles, we will use the principles as a framework for analyzing OISS claims.

Principle 1: *AI research must set its foundations in ways that treat the existence and interaction of multiple actors as a fundamental category.*

Since we observe and actually are building multi-agent systems, we should investigate how to conceive aspects of representation and reasoning as fundamentally grounded in multi-agent systems. This leads directly to a serious research question for AI, namely:

How can we usefully conceptualize representation, reasoning, problem-solving, and action, when we begin with multiple participants?

A social perspective on the nature of intelligent behavior is not a new idea. For example, Mead stated:

> "We are not, in social psychology, building up the behavior of the social group in terms of the behavior of the separate individuals composing it; rather we are starting with a given social whole of complex group activity, into which we analyze (as elements) the behavior of each of the separate individuals composing it. We attempt, that is, to explain the conduct of the individual in terms of the organized conduct of the social group, rather than to account for the organized conduct of the social group in terms of the conduct of the separate individuals belonging to it. For social psychology, the whole (society) is prior to the part (the individual), not the part to the whole; and the part is explained in terms of the whole, not the whole in terms of the parts." – [Mead34, p. 7].

The traditional set of analytical categories and implementation techniques used in AI does not include fundamentally social elements; the focus is on the individual actor as the locus of reasoning and knowledge and the individual proposition as the object of truth and knowing. For example, a number of researchers are studying *commitment*, a basic concept in OISS, as a foundation of many concepts in AI and DAI, including intentions and goals, negotiation, and knowledge [Bond90, Cohen87, Cohen90, Fikes82, Winograd86]. The research literature most often portrays commitment as a kind of rational choice made by an individual actor. Along these lines, Cohen and Levesque [Cohen87, Cohen90] have developed a notion of commitment based on what they call a *relativized persistent goal*. Some agent A relativizes its goal g to a predicate q, so that A gives up g only when A believes that either something has satisfied g, or nothing can satisfy g, or $\neg$q. To use one of Cohen and Levesque's examples, when rain is falling, A may reason that it will be committed to the goal of getting an umbrella unless it believes that 1) it has obtained an umbrella, or 2) it cannot get an umbrella, or 3) the rain has stopped and it no longer needs an umbrella. They state that "Persistence involves an agent's *internal* commitment over time to her choices....This is not a *social* commitment; It remains to be seen *if the latter can be built out of the former*" [Cohen87, p. 410 (final italics mine)].

Symbolic interactionist sociologists, and authors of recent investigations in the sociology of science, have begun to provide some conceptually fruitful, though not presently computational, approaches for understanding knowledge and action in social terms. (See [Clarke90] for an illuminating review, and [Social Studies89, Collins90] for discussions directly related to AI.) In contrast to Cohen and Levesque, for example, Becker [Becker60] and especially Gerson [Gerson76] develop commitment as the overall organization of an agent's participation in many settings simultaneously. For example, imagine that a Los Angeles industrialist takes off in an airplane from Narita airport, bound for California, after formulating preliminary business deals in Tokyo and telephoning her associates in Los Angeles. While flying, she is participating in many settings simultaneously: the activity in the plane, the ongoing business negotiations in Tokyo and in Los Angeles (where people are planning for her arrival and making business judgements while considering her views, even in her absence)[2]. Her simultaneous involvement in interlocking courses of action in all of these situations provides the commitment to her arrival in California. Both she and others balance and trade off her involvement in joint courses of action in many different situations. Moreover, whether she makes a choice or not, she is committed to landing in LA because the plane is not in her control. Her commitments in any of these settings *amount to* the interaction of many activities of many agents

[2] Of course we leave out many, many others—her family, etc.—including some she may not be aware of.

in many other settings. Since this multi-setting participation occurs *simultaneously* in many places, it can't be located simply to where she physically "is." In other words, the notion of commitment is distributed because the agent of commitment—"she"—is a distributed entity.

This approach rests on a somewhat untraditional idea of what an agent is: in Gerson's formulation, an agent A is a reflexive collection of processes involved in many situations. To varying degrees, the agent–that is, some component process of the agent—can take on the viewpoint of any participant in those situations. Commitment of A (i.e., continued participation of A) in a course of action in any particular setting is a product of the interactions among its simultaneous participations in many other settings—*whether* A *explicitly "knows" fully about the other settings beforehand or not.* Thus, if A has goals, they can't be effectively "relativized" because the relativizing conditions that would make A's goals contingent can't necessarily be known beforehand (a version of the qualification problem [McCarthy77]). Moreover, since continued participation is distributed and simultaneous, it isn't based on localized, individual choices and goals.

In the umbrella case, from the social perspective, an infinite variety of circumstances may arise under which A's participation in other settings could change A's participation in the umbrella-getting course of action; at any time, some other agent could act in a way such that A is no longer a participant in that course (In simple cases, A could get hit by a car or unplugged); this presents problems for Cohen and Levesque's notion of commitment. Commitment from the social perspective is grounded in the actions of many agents' activities *taken together*—it is not a matter of individual choice. It is A's actions in relation to those of others (and vice-versa) that maintain A's participation in a course of action (e.g., by providing resources, etc.—see below). Commitment in this sense is the outcome of a web of activity, or in OISS terms, it is "Systems Commitment."

Moreover, this social notion of commitment doesn't rely upon a more-primitive mental concept such as "belief" or "goal" (this is how it unifies the individual and the social). In fact, this notion of commitment cannot be grounded on individual belief or choice, because it is not located "within" the individual. Because of this, it extends in varying degrees to objects as well as people as active participants in settings, and to mutiple levels of analysis. For example, for the industrialist to call Los Angeles from a coin-operated telephone, a both she and a telephone system must together enter into a course of action that involves consuming coins, providing dial tones, and so on. *They are mutually committed to doing those things in that way to make a phone call*, regardless of whether she or the telephone has any mental state such as a state of belief, or any shared view of the situation. (A self-dialing modem can make phone call. Does a telephone have a viewpoint to share?). The industrialist's other commitments (e.g. in the loan-funding process) are simultaneously mediated by the actions of the telephone—and of course of the whole telephone network and organization behind it: waiting time, missed connections, etc. (Cf. [McCarroll90]).

Many other concepts which are basic to AI researchers and AI programs, and typically (in AI) associated with individual actors or problem-solvers, are, in sociological terms, reifications, constructed through joint courses of action and *made* stable by webs of commitment [Becker60, Gerson76], or "alliances" [Latour87, Latour88] among the actors using them. Some examples include concepts such as *problems* [Fujimura87], *knowledge* [Becker86, Clarke90, Lave88], *facts* about the the world [Latour79], and even technical objects [Hughes83]. From this perspective, stable alliances or systems of commitment even produce the demarcation and ongoing existence of individual agents as units of knowledge and interaction. In the case of people, for example, alliances among cells, chemical processes, and the environment at the lowest levels and among social actors at the more macro levels (e.g., organizations such as hospitals) yield stable and ongoing individuals[3]. In a computational intelligent agent, such a web includes (at least) the structure of the computing system and all that keeps it running "properly," including the program, the evolving content of its data structures (e.g., a set of represented propositional beliefs) the language processor, the hardware, and the resources and activity (electricity, maintenance, and so on) that keep it active over time (cf. [Kling82]). This is as true of a connectionist system as of a symbolic one. Perhaps the nature of this idea of alliances, and the conception of both agents and knowledge as stable systems of alliance, are easier to see if we examine what it takes to remove an agent's influence in a situation (e.g. by disabling it or discrediting its knowledge). What alliances must be broken? Actually it can be fairly simple - unplug the machine; or change the operational semantics of its program e.g., by changing the operating system, language processor, or hardware [Sethi89]; or change the behavior of another agent upon which it critically relies; or change the definition of a set of possible worlds which establishes the semantics of a proposition in its belief set.

Treating problems, knowledge, and facts as webs of commitment is a fundamentally non-local, distributed conception. Like conventional AI conceptions, such distributed conceptions account for change in knowledge and world states. They have the additional advantage of accounting for the stability and robustness of facts or agents or procedures in the face of challenges posed by alternative viewpoints or discrediting activities (sometimes known as "brittleness"), and for what OISS calls the intederminate nature of systems.

Certain existing approaches to overcoming brittleness are theoretically problematic. For example, TMS/ATMS systems and belief networks, which do locate belief in a network of supporting evidence, rely on unwinding of assumptions and the posing of incommensurate alternative worlds or contexts—but they cannot account for how to resolve inconsistency at the assumption levels; these are battles that agents resolve outside the system. They rely on the option of keeping alternatives separate, until some unifying viewpoint or discriminating facts appear from some external source. They also don't allow for nth-order flexibility or robustness–e.g. in the choice of world representations, proof theories, etc., and they are subject to deductive indeterminacy, as the OISS proposal points out.

Principle 2: *DAI theory and practice must address the basic tension between the local, situated, and pragmatic character of knowledge and action, and the ways in which knowledge and action necessarily implicate multiple contexts.*

The notion that the meaning of a message is the response it generates in the system that receives it was introduced by Mead (see, e.g., [Mead34, Chapter 11]) and was later used independently in the context of computing by Hewitt [Hewitt77]. Using this conceptualization, a message that provokes no response has no meaning, and each message with impact has a *specific* meaning, played out as a set of specific response behaviors. In an asynchronous and open distributed system, no message can be guaranteed to lead to the same set of behaviors twice. Thus knowledge in an open system always means something local and situated. (See also [Agre88, Lave88, Suchman87]). As to the implications for action, actors take actions (including reasoning and planning actions) at specific times and places with specific (but of course possibly selective, incomplete, faulty, etc.) knowledge

[3] The issue is the nature of the individual as the locus of interaction and knowledge. Bentley [Bentley54] and Dewey [Dewey96] lay out the problems well; [Buss87] and [Wimsatt80] discuss evolutionary changes in biological units of selection from cells to higher-order aggregates and [MacFarlane78] discusses transformations of units of knowledge, action, and ownership from the social to the individual in English history.

brought to bear. In a sense *action is a particular commitment to doing things a particular way*—a way conditional upon the actor's particular knowledge in and of the situation of the action (cf. [Moore77]).

It seems, however, that some sort of generalization across situations is what makes knowledge useful and what ultimately makes knowledge knowledge. General knowledge makes possible *action-at-a-distance*[4]: reasoning about and taking control over activity located at some other place in space or time such as the future, another network site, or over actions taken by another agent—in other words, acting in a distributed fashion. There is, then, a basic tension between a local, "situated" conception of knowledge and action, and the non-local conception of action-at-a-distance. It appears that the ability to generalize across situations and the utility of doing so makes knowledge inherently *non*-local. The knowledge is derived from and can apply in many situations.

Still, any general knowledge, to be useful, has to *be applied* in a local setting, hence *made local* again. Generalization leads to transportability across contexts, and thus helps in achieving action-at-a-distance, but does not obviate the need for reintegration into a local context of use. A production rule with variables exemplifies transportability. Variables make the rule applicable in any setting where they can be bound. Such a rule is useless with variables unbound[5]; binding variables specializes the rule into a specific *rule instantiation*, i.e. makes it local and specific again. Moreover, the localization process itself (e.g., the binding of variables) is another purely local and situated process.

Principle 3: *Representation and reasoning approaches used in DAI must 1) assume that multiple representations are recursively possible at any level of analysis or action, 2) assume that actors will employ multiple representations individually and collectively, and 3) provide mechanisms for reasoning among multiple representations.*

In order to understand fully the implications of the OISS analysis on the limits of deduction, we need to understand the character of what we usually view as "shared" knowledge. This is important, for example, in understanding the nature of contradiction, a concept crucial to several OISS arguments. Shared knowledge, as I think we normally conceive it[6], is impossible; nonetheless, we have ways of pragmatically aligning our activities and acting *as though* we share knowledge (see, e.g., [Suchman87]). The difference becomes an issue precisely when conflict arises, and appeals to shared knowledge are inadequate both to explain the nature of and to resolve conflicts. Approaches to conflict that rely on logical formulations necessarily require a common semantics even to decide that conflict exists. Conflict means inconsistency and inconsistency is impossible without a common model. In an open DAI system without a-priori assumptions of globality, we need another definition of conflict. The choices we have come down to conflict in action and more specifically conflict in the consumption of resources, not just conflict in representations.

Different actors necessarily have different sets of commitments, by virtue of their different histories, the different resources they use, different settings they participate in, and so on. Multiple perspectives are a fundamental feature of any multi-agent system, simply by virtue of differing

[4] In general, "distance" here refers to some axis of distribution. [Bond88a] discusses numerous axes of distribution in this sense, including space, time, and semantics.

[5] Except of course when used itself as an object of discourse; It then becomes a localized and concrete representation employed in a higher-order (meta) process).

[6] I.e., as several agents knowing the *same fact* interpreted the *same way*–what would this mean, and how would the agents ever be able to verify it? Cf. Principle 5 below.

commitment histories and local circumstances. The interesting phenomenon, then, would be any apparent *commonality of perspectives* or mutually aligned, mutually supportive commitments—how would they get and stay that way [Gasser86]?

If multiple perspectives are basic, disparities in perspectives are an issue. Elsewhere, we have posed this issue as a basic problem for DAI, because of its theoretical consequence and its ubiquity in DAI research [Bond88a, Gasser90]. Moreover, multiplicity of perspectives raises the issue of the impossibility of global conceptions. As Star points out in her study of the development of a localization theory of the brain by a community of scientists[7]:

> "The momentum of the theory, professional developments, turf battles between specialists and general practitioners, and the rise of specialty hospitals with their separate domains of expertise made the theory impossible to comprehend from any single point." [Star89a, p. 193].

In effect, what the scientists involved talked about as "a theory" was in fact multiple theories by virtue of the multiple perspectives brought to the activity of expressing and understanding it.

Principle 4: *DAI theory and practice must account for resource-limited activity.*

All resources are limited, and real agents act in finite circumstances. Resources used by a collection of agents can be arranged and allocated in numerous ways, but the resources used to allocate resources are also limited, and in the end agents do take particular actions. "Optimal" resource allocations are in general not possible, for at least four reasons: 1) computing an optimal allocation might require infinite resources, 2) allocation actions must be taken opportunistically in a dynamic world, 3) there is no limit to how completely a an allocation decision situation can be specified, and 4) agents might not agree on criteria for optimality. Moreover, no agent supplies all of its own resources. Resource allocations are the *product* of interactions of many agents, and at the same time resources serve as a key *channel* of interaction among agents—as one agent uses up a resource, others' options are restricted. Thus a complete DAI theory must integrate a treatment of limited resources with a treatment of joint actions of multiple agents.

Principle 5: *DAI theory and practice must provide accounts of and mechanisms for handling the three key problems of joint qualification, representation incommensurability and failure indeterminacy.*

The impossibility of fully specifying the assumptions behind a characterization of any situation, has been termed the *qualification problem* by McCarthy [McCarthy77]. Given this, DAI theories must account for how agents can come to have and to act upon *mutually compatible* sets of assumptions (e.g., common defaults) in the face of partial descriptions and no global semantics. That is, how can agents leave compatible aspects of a situation unquestioned or unsupported—what accounts for how they can "stay out of each others' way" when they do? No agent can fully describe its assumptions to another, yet they must mutually take some things for granted to act jointly without conflict (see, e.g., [Suchman87]). This can be called the *joint qualification problem*, and a full DAI theory must account for it.

[7] For an analysis of multiple perspectives over time, as well as over agents, see also Lakatos' study of the reconstruction of mathematical theorems in [Lakatos76]

In the face of the assumption incompleteness, no agent can fully specify the semantics of its representations. If this is so, how are two agents to determine if they have the basis for joint action, or if they are in conflict? This can be called the problem of *representational incommensurability.* [Bond88a] discussed three types of disparity among agents' knowledge: incompleteness, inconsistency, and incompatibility. The first two are conventionally defined, and incompatibility referred to agents' representing the same situation with different kinds of descriptions. With incompatibility disparity, consistency could not be assessed. Incommensurability, is a still deeper problem. Two agents in principle cannot have *identical* representations—any pair of similar representations can always be differentiated by more complete description. So on what basis can agents be sure that they either 1) have common (e.g. Tarskian or possible worlds) semantics (since the definition of a model or possible worlds would have to be global), or 2) common semantics based on our earlier theory of meaning as response of the system (because response can only be assessed from some particular perspective)?

Finally, when there has been some disparity at some level between two agents with different representations of the same situation, and this conflict leads to a failure of action, how are they to determine where the cause of the failure lies? For example, both Agre and Gasser have discussed the nature of agents' behavioral *routines* (Agre in the single-agent case [Agre85], and Gasser for organized activity [Gasser86]). Suppose an agent A has a theory $\mathcal{T}$ of the routine behavior of another agent B. For example, $\mathcal{T}$ might be:

(1) If I send B a task announcement then B will reply with a bid request

Since this is a theory of a routine, it is necessarily an idealization—no routine behavior is actually carried out in precisely the same way twice [Gasser86]. Now suppose that to reason about B's behavior in a particular situation s, A qualifies or specializes $\mathcal{T}$ with some additional observations $\mathcal{I}$. (Since $\mathcal{T}$ is in idealization, $\mathcal{I}$ is necessary to make it fit s.) For example, since today is Monday and communication is via email, $\mathcal{I}$ might be:

(2) If today is Monday and bids are to be sent via email then (1)

$\mathcal{T}$ with $\mathcal{I}$ will lead to some prediction q about B's behavior:

Today is Monday
Bids are to be sent via email
I sent a task announcement to B
Thus:
q: B will reply with a bid

But suppose that A's observation q' of B's behavior is inconsistent with q?:

q': B does not reply with a bid

Does the problem lie with $\mathcal{T}$ or with $\mathcal{I}$?[8] (Perhaps B doesn't send bids to every task announcement, or perhaps it doesn't read email on Monday). The problem may even lie with the way $\mathcal{T}$ or $\mathcal{I}$ are interpreted in the situation (e.g., A got a message but was it a bid?). (Since interagent interaction

[8] This exposition makes use of Laymon's argument on the difficulties of using experiments for drawing conclusions about the truth of scientific theories [Laymon85].

is involved, the joint qualification and representation incommensurability problems also enter into the interprteive question.) The unfortunate problem seems to be that unless we already know that both the interpretive scheme and A's theory of B's routine are correct, A can't tell how to make them so because it can't deduce what failed—at least not using its own knowledge. This problem can be called the "failure indeterminacy" problem. It is the problem faced in any scientific experiment or court of law: Since the acceptability of a scientific theory depends on the experiment, and the experiment depends upon the apparatus, and the nature of the apparatus depends upon the theory, where is the source of experimental conviction? In court, is the defendant guilty or is the prosecution's theory wrong? Of course in either domain, like good distributed reasoners, we rely on many experiments and agreement among many participants, not just one—but this raises the joint qualification and representation incommensurability problems again [Collins85].

Principle 6: *Overall, DAI theory and practice must account for how aggregates of agents can achieve joint courses of action that are robust and continuable (ongoing) despite indeterminate foulups, inconsistency, etc. which may occur recursively at any level of the system.*

OISS raises the issue of self-reliance for DAI systems: how can agents preserve local autonomy (i.e. become robust to failure and challenge) while still drawing from and providing resources to the larger community? The first 5 principles above point to numerous possible sources of failure, discrepancy, and potentially indeterminate states of knowledge in which any agent in a multiagent system can find itself. Principle 6 takes note of the fact that robust DAI systems that handle all of these contingencies *do exist*: many human social organizations as well as deeply embedded information systems (e.g., [Gasser86, Kling82, McCarroll90]. Any complete DAI theory must account for how this is possible and what the limits are; a complete set of mechanisms for DAI ought to provide us the capability to construct such systems within the limits.

4 DAI Foundations and Open Systems Semantics

With these principles in mind, then, let us move ahead to consider OISS as a proposed foundation for DAI. The OISS viewpoint has two primary components. One is an investigation of the *deductive indecision problem*, and the other is a characterization of open systems, the the nature of problem solving in them. A style of reasoning that elsewhere has been called *due process* [Hewitt86], built from concepts such as trials-of-strength, commitments, and negotiations, glues these two together. We shall first discuss the nature of the overall OISS argument. Then we shall investigate how effectively the OISS proposal addresses the 6 principles presented in Section 3, contrasting the utility of the OISS approach with that of existing DAI research. The thread of the OISS argument is as follows:

1. DAI research is concerned with work in large-scale open systems, but DAI does not yet have a clearly articulated vocabulary or common conceptual machinery. Open Information Systems Semantics (OISS) can provide a useful and coherent set of concepts, some tractable research issues, a methodology, and a comparative vocabulary for DAI.
2. DAI systems trade off the costs and benefits of *self-reliance*—the ability to take effective local action and to become robust against indeterminacy and conflict, with *interdependence*—contributing to the performance of the overall aggregate and drawing from it.

3. “Deductive microtheories” are the primary competing foundation for DAI. Logical semantics are sufficient for reasoning in closed systems, and hence can be used as a foundation for reasoning within deductive microtheories. Problem solving in open systems involves interacting proposals founded in different microtheories. Different microtheories are generated and modified asynchronously, and involve differing commitments among their participants. Thus, logical and representational conflict is endemic to open systems. Logic is insufficient for reasoning in the presence of conflict and meta-conflict (i.e., conflict over the boundaries of decisionmaking—e.g., circumscription axioms in [Hewitt91]) and therefore for conflict resolution. Thus, **Conclusion 1:** Because conflict is endemic, and logic is insufficient for processing under conflict, deductive microtheories are insufficient as a foundation for large-scale DAI in open systems (though they may be useful components).

4. Alternative and more powerful foundations can be built upon the notions of *trials of strength* and *systems commitments*. Commitments are commitments because they are relatively stable or *robust* in the face of challenge or conflict.

5. Constructing and exchanging “representations” is a basic activity; representation is not possible without communication. The “meaning” of a representation is defined to be the ways in which it modifies systems commitments.

6. *Negotiations* (and other trials of strength) are the tools by which conflict is processed. Negotiations can occur recursively at many levels of analysis, have many potential outcomes, are inherently creative, and generate further commitments.

7. **Conclusion 2:** Founding DAI in OISS is a different proposition from founding DAI in classical AI terms. OISS is inherently more “social,” “grounded in large-scale information systems” rather than individual agents, and provides a different account of representation processes.

Up to **Conclusion 1** the OISS argument is relatively strong, but within some narrow limits (which incidentally are left underspecified). It is not entirely true that DAI has failed to crystallize a common conceptual vocabulary, including a set of problems, methods, and terms. Section 2 presented a collection of these, gathered from a thorough examination of the DAI literature. Another very detailed proposal for a core set of DAI problems that coheres closely with those above can be found in [Decker89]. With some exceptions the more basic principles presented in Section 3 above have not in general been fully articulated or addressed in extant DAI research.

Some deeper questions are the extent to which DAI has been addressing the right set of problems at the right level of analysis, and how OISS may focus us on a different set of problems that is either more fundamental or that allows us to make better headway by changing our perspective. The implication of the OISS perspective seems to be that DAI has not chosen the appropriate set of problems. Deductive indeterminacy is clearly an issue that DAI research has certainly not openly considered until now, though other disciplines have addressed variants (see below). It is not properly subsumed in the 6 DAI problems of Section 2. The self-reliance—interdependence problem is a clearer and more encompassing notion than “global coherence with local control.” But the only other problems posed (e.g., understanding negotiation, commitment, representation, etc.) are also precisely the set of concepts proposed as solutions, and the way they are to be woven together in a mutual foundation is unclear. There are several other key problems that must be addressed for a complete account of open DAI systems, including some of those discussed in several of the principles above. OISS actually does provide ways of thinking about them, but they are not clearly articulated as problems.

The observation that DAI is inherently concerned with work in large scale open systems is only partially true; DAI certainly *should* be concerned with that question, but most contemporary researchers have had their hands full grappling with the (apparently) far simpler problems of coordination and performance of collections of agents under certain closure assumptions (see Section 2). An interesting open question, then, is what is the extent to which providing new foundations such as those of OISS will simplify the problems of knowledge and action in closed systems as well, and possibly go some distance toward eliminating the categories "open" and "closed[9]."

It is true that there has not been enough methodological clarity, debate or variety in DAI[10], as has been pointed out in both [Bond88b] and [Gasser89a]. But for many of the standard DAI problems, existing representation and experimentation methods have provided fruitful progress[11]. It is not entirely clear what the *methodology* of OISS is, or whether the OISS methodological focus is analytical or constructive. To what extent will it help us explain the behavior of existing DAI systems? How can a constructive methodology be built upon the explanatory theory? As an analytical theory, we are provided with a set of concepts but little guidance for how to go about finding instances, studying, comparing, or operationalizing them. Useful research methods for studying OISS questions analytically have been clearly articulated in sociology, upon which OISS has has drawn for its concepts (e.g., [Strauss87]), but these or other such are not integrated into the current OISS approach as methods.

To the extent OISS provides a mathematical or computational analysis, the Actor model for concurrent systems [Agha86] is the chosen descriptive calculus, but at the moment, the connection is incomplete. There are three partially clear links from features of OISS to the descriptive machinery of the Actor model. *Actor configurations* are ways of providing local abstractions or closures, but are not clearly connected with OISS foundation concepts such as commitments or trials of strength. *Serializers* are one way of settling a trial of strength by arbitrating the handling of simultaneously-arriving messages, and they do capture fundamental indeterminacy of open systems. *Replacement behaviors* give Actors both local autonomy and participation in joint enterprises, and thus help to address self-reliance issues. The relationship between the Actor model and other concepts such as negotiations, cooperation, commitments, etc. are not clear, and thus the formal descriptive power of OISS is currently limited.

In previous work, such features of open systems as arms-length relationships and asynchrony have been treated as the sources of difficult problems to be overcome. Now, from an OISS perspective, these are also provide benefits for components of DAI systems. The notion of "self-reliance/interdependence" is used to capture the advantages and disadvantages of becoming more autonomous while somehow staying integrated with a larger community of agents. But the unit of analysis over which this self-reliance occurs is not clear—what is the self that is self-reliant? Is it a particular node in a system? If so, how are the boundaries of this node defined, by reference to a fundamentally distributed conception of knowledge and action? We can contrast the OISS notion

[9] Recent ferment in sociology, history, and philosophy of science is moving in precisely this direction. See, e.g., [Collins85, Garfinkel81, Gerson77, Star89a, Teil90].

[10] Or in AI in general; see, e.g., [Hall85].

[11] One notable exception is the issue of reflexive modeling and reasoning about DAI system behavior for development purposes and as a foundation for organization self-design. See [Gasser90, Hudlicka87].

of self-reliance to Gerson's concept of *sovereignty* which is:

> "...the overall organization of commitments associated with any delimitable social object...the net balances of resources and constraints available to a person, organization, or other demarcatable group across the full range of settings in which he (or she, or it) participates." [Gerson76, p. 798].

Sovereignty can be seen as the kinds and degrees of constraint an object faces, over all the situations in which it participates simultaneously, and resulting from its interactions in those settings. As Gerson points out, the locus of sovereignty is any particular social object it is convenient to use for analysis, and it also "removes the distinction between 'individual' and 'society' considered as abstract entities apart from their activities and each other" [Gerson76, p. 798]. An object has its particular type of sovereignty by relationship to those other entities and situations in which it participates; it never stands alone (cf. the discussion of commitment in Section 3 above). The self-reliance/interdependence framework maintains the distinction between the individual and the larger system in which it participates.

The question of the limitations of deductive microtheories for open systems reasoning is not a new one, though I have not seen it formulated in circumscriptive terms before. Göedel's second incompleteness theorem is based on a variant of it [Quine81], as is Garfinkel's famous description experiment[12] [Garfinkel67]. The importance of the OISS account is that it draws our attention to a basic limitation of a tool drawn upon by DAI theorists, and because it stresses the need for other computational approaches.

OISS also presents a proposal for alternative foundations for DAI, based on the new lexicon of trials-of-strength, systems commitments, representations, negotiation, cooperation, etc. Our problem is to investigate how clearly and how completely the OISS proposal addresses each of the principles for DAI foundations. One difficulty of doing this is the vagueness of some definitions. The nature and scope of concepts such as "trial of strength," "commitment," "systems commitment," or "negotiation," are matters of inference from examples, not definition. Without greater background it is sometimes difficult to see which features of an example are relevant to the concept under elucidation. For example, does "in place" mean something like "continuable" or "ongoing" (i.e., not deadlocked or otherwise become impossible)? Or does it mean something like "robust" (able to face many different challenges and withstand them)? Part of the problem may be that some concepts have not yet reached conventionalized status (e.g., "negotiation," a term that has been used in literally dozens of different ways in the DAI literature).

With the specter of misconceptions of definition looming over us, let us examine how OISS addresses the 6 principles of Section 3, how it extends current wisdom in the DAI literature, and how it is deficient.

[12] Garfinkel asked students to explain the meaning of a conversation by annotating it, and ultimately to give a set of instructions for unambiguously describing the meaning. Students took this as a request for more complete description, but finally realized that the task was impossible. Further description only muddied the issue because the descriptions themselves were potential sources of ambiguity. There had to be some other way of achieving conversational coherence besides shared *a priori* assumptions. This idea underlies Suchman's discussion of human-machine communication in [Suchman87].

4.1 Principle 1: Multiple Actors.

Some statements of fundamental AI problems have recognized that multiple actors with different viewpoints are an important part of AI (e.g., [Fikes72, McCarthy77]). Of course, DAI research by definition deals with multiple agents, but to date, DAI research has had only limited theories. What theories do exist take certain aspects of system closure for granted, as pointed out in Sections 2 and 3. Many OISS concepts have already been in widespread use in DAI systems. For example, Mason and Johnson have designed a Distributed ATMS system for nuclear seismic analysis [Mason89]. In this application it is essential that each node avoid compromising its local set of beliefs and assumptions by integrating faulty or malicious messages from other sensing nodes—that is, each node must maintain local autonomy and arms-length relationships while incorporating useful information generated by others. Mason and Jonson's approach is to let each node use non-local information for local focus-of-attention decisions, but never to propagate it. Similarly, the DVMT of Lesser and Corkill [Lesser83] includes mechanisms to experimentally vary a node's degree of local autonomy and how greatly it can be "distracted" by information from others; they term this "internal versus external control," and note that positive and negative distractions are sometimes hard to distinguish with a local perspective. Their definition of organization as a set of well-defined problem-solving roles and communication patterns implemented by restrictions on agent capability can be interpreted as a collection of "systems commitments"—but they are commitments by virtue of nodes' lack of sovereignty over their own roles, which is to say by virtue of the actions of designers and reflexive limits of representational theories.

OISS provides a strong foundation for DAI to the extent that it provides an account of knowledge and action from the social level to the individual (which it begins to do), recognizes the possibilities for fundamental disparities in agents' views (which it clearly does), and presents a theory of how agents act despite these disparities and without global knowledge (which it does partially).

In another light, OISS attacks in some sense the wrong problems of multi-agent systems. The important issue is not necessarily the inadequacy of closed-system microtheory techniques for OS problems, (about which there is likely to be little debate) but instead the nature of the processes of "closure" - when and how it is appropriate to make and rely upon closures, and what to do when they break down. This is my reading of one intent of circumscription and other foundations for nonmonotonic reasoning - to provide a promising but necessarily incomplete theory of how to make useful closures in a local reasoning process.

4.2 Principle 2: Tension Between Situated, Pragmatic Knowledge and Action-at-a-Distance.

In contemporary DAI research this principle is addressed by reference to the problem of "how to achieve global coherence with local control" [Decker87, Lesser87], which involves the first five of the six basic DAI problems discussed in Section 2 above. Typical analyses assume global views are possible, (e.g. by an observer or oracle, to measure global coherence), that disparities that impede global coherence occur only at one level of interaction, and that general knowledge can be applied in remote settings by communicating it. For example, representations and interaction protocols are generally assumed to be fixed within the system or theory, making performance *theory-relative* to the descriptive limits imposed by them (cf. [Smith86, Maes88]. The local utility approaches of J. Rosenschein and colleagues probably come closest to accounting for locality of knowledge because they do not depend on shared notions of utility, but they are, again, single-level analytical schemes. Agre and Chapman's "indexical" approach has promise, but it is not clear how to scale it up to aggregate interaction, and they still take the individual agent and its relationship with the world as the locus of knowledge and activity—see, e.g., [Chapman87].

OISS concepts useful for addressing Principle 2 include the self-reliance/interdependence tradeoff, the reliance on local processing of Representations and the notions of Systems Commitments. OISS proposes negotiation as a basic mechanism. In OISS, global coherence would be conceptualized as the situated outcome of a negotiation—as long as agents collectively reach agreement (and agree that they have), their actions are coherent. But because of indeterminacy and late arriving information, a preordained concept of global coherence doesn't make sense for OISS—it is necessarily a *post-hoc* notion.

Latour provides a partial and not computational answer to the problem of posed by Principle 2, that has not been fully assimilated by OISS, but that is coherent with much DAI work. The way to achieve action-at-a-distance is "...by *somehow* bringing home these (distant) events, places, and people..." [Latour87, p. 223]. How to do this? By turning the remote entities into "immutable mobiles" which are *mobile* (transportable across contexts), *stable* (so that they keep their useful qualities in new contexts) and *combinable* (so that they can be usefully entered into associations with other such things). That is, by either bringing back preserved, representative samples (e.g. collections of animals or plants) or by bringing back *representations* of distant terrain (e.g., maps, notes, descriptions) built in a systematic (combinable) language. As indicated above, much DAI research has investigated the problems of *building models of other agents*, and of using and exchanging these models as foundations for coordination [Durfee87a, Gasser89b, Lenat75, Sycara89, Rosenschein82, Rosenschein85]; from an action-at-a-distance perspective, models of other agents are the crucial immutable mobiles.

However, the stability (immutability) of any of these "mobiles," reflected in their continued representativeness, is always problematic. Transporting plants, animals, and other exemplars necessarily strips them of their context, and may render them uncombinable (e.g. if they die in a new habitat). Transporting representations raises problems of completeness (is the map detailed enough?), and of interpretation in a new context. Others in AI have begun to deal with the problem or re-interpretation in new contexts, and have suggested that it be considered in the context of the hermeneutic problem [Winograd86]. Latour's account doesn't deal fully with the mechanisms for keeping mobiles stable. OISS addresses the concern with stability of representations, in part, by delimiting its scope to open *information* systems, which are defined to be systems which manipulate digital information. The advantage of digital information is precisely its stability over time and space, and (ideally) its combinability with other digital information. It is not clear to me, however, that digital information is ihherently more or less combinable than any other information, except insofar as its combinability can be automated; Some studies have shown the inherent difficulty of combining digital information [Gasser86]. Moreover, the stability of interpretation over context is still problematic for OISS. Conceptually the problem can be handled by better integrating the ideas of webs of commitment developed by Becker and Gerson, but it still needs to be made computational.

It doesn't seem sensible or complete, then, to take the OISS view and say simply that representations are "information conveyed using digital communications." Instead, it seems more accurate to characterize representations as artifacts ("inscriptions"—[Latour79]) that can be passed around and reinterpreted. Latour's point is that inscriptions are useful precisely because they are transportable across spatial or semantic contexts and they are combinable. (cf. Star's discussion of boundary objects in [Star89b]). To link representation and communications, therefore, we can say that any knowledge intended to be used non-locally must be converted into a stable mobile (represented) and (re)interpreted in the local context where it is delivered. In the light of the need to keep the mobile stable, communication can be seen as the maintenance of a collection of commitments across contexts. Communication takes place via the webs of commitment. Though in OISS communication takes place digitally, that is only possible within webs of commitment [Gasser86, Kling82, McCarroll90].

4.3 Principle 3: Multiple Perspectives

The advantages and disadvantages of multiple perspectives are well known in contemporary DAI research. Multiple views can be used to improve robustness, and several techniques for reaching reliable joint conclusions using many bits of unreliable data from multiple perspectives have been proposed. These include the Functionally Accurate, Cooperative (FA/C) problem solving approach of Lesser and Corkill [Lesser81], the Distributed ATMS approach of Mason and Johnson [Mason89], and the Ether problem-solving system of Kornfeld and Hewitt [Kornfeld81]. The disadvantages of multiple perspectives (e.g. for global coherence) are well recognized in DAI research, and many distributed coordination mechanisms are based on reducing disparities globally by exchanging self-descriptions among agents in a process often called negotiation (e.g., [Davis83, Durfee89, Kuwabara89, Sycara89]). But all current approaches rely on a global perspective on some level, whether it be semantics or communication protocols, and assume that the context of negotiation cannot itself be negotiated; thus DAI as yet has no complete theory. PGPs have been suggested as a foundation for multilevel negotiations, but not for reflexively negotiating communications protocols [Durfee89].

OISS provides a deeper understanding of the basic problems of multiple viewpoints than is currently extant in most DAI. In particular, OISS accounts for the fact that negotiations can be carried out at any level of the system, including negotiations about the appropriate context of negotiations. (Others share this view to various extents. See, e.g., [Durfee89, Ferber88, Gasser89b]). But a primary difficulty is that, despite defining negotiations as "Trials of Strength carried out using Representations," OISS provides no mechanism for integrating negotiations and more primitive (i.e., implementable) trials of strength. We do have illustrations of trials of strength at several levels of complexity, but no guidance in constructing these into multi-level negotiation mechanisms.

A *perspective* can be seen as a local organization of commitments that takes some aspects of the situation as variable or negotiable and others as fixed (cf. [Gasser89b]). Strong commitment webs are ways of making things seem invisible or taken for granted—unquestioned—in dealing with the world. For example, in a logic-based agent a perspective is manifested as the choice of a set of predicates an agent uses to describe its world, and their truth values, which the agent then uses, in a taken-for-granted way, as a world representation. It is also manifested in the decision processes the agent uses to weigh control choices it makes; these are typically commitments that cannot be changed by the agent. We can view these as *commitments* because the agent—or its designer—*could* change its representation, but that would take shifting other commitments in other contexts, e.g.

commitments to using some particular communication protocol, understood by others, that relies on those predicates, or to avoiding the effort of reprogramming. Thus, the advantage of metalevel control is that it allows an agent to take on different control perspectives reflexively, but at added cost [Durfee87b, Hayes-Roth85, Maes88].

Multiple perspectives can be seen as differences in commitments. Moving from one perspective to another, or aligning perspectives among agents involves changing some set of commitments - i.e. the commitments that define what the local perspectives are, e.g. commitments to what predicates to use, or what assumptions to allow, or what features of a situation are important, etc. In this way, the OISS concepts of Systems Commitment, Representation and Negotiation can be brought together, and used as a foundation for conceiving problems of disparate perspectives.

4.4 Principle 4: Resource Limitation

Lesser and Erman described the DAI problem as that of enabling a collection of problem solvers to exercise sufficient control to make use of available resources and knowledge to solve a problem, assuming that the knowledge and resources were adequate for some solution [Lesser80]. Some recent DAI work has turned to resource-bounded problem solving. The issue has been inherent if not explicit in DAI due to the ways global coherence has been measured. If work is divided among nodes with potential redundancy, then one node's activities must be temporally correlated with the responses of its associates. Otherwise, these nodes may perform necessary tasks themselves, believing that they haven't been done—leading to redundancy and lowered global performance. Thus time constraints can arise purely by the need for coordination. The primary distributed AI approach to explicit resource-bounded reasoning has been *approximate processing*, introduced by Lesser et al. [Lesser88].

Problem solving under resource constraints is not clearly accounted for in the OISS framework. Earlier notions of resource *sponsors* introduced by Kornfeld and Hewitt [Kornfeld81] have not been incorporated into OISS at this point. This naturally raises the question of how would OISS approaches fit in real-time settings? Resource limitations are not explicit in the OISS notion of systems commitments, though they could be made so.

The oversight of OISS in respect to Principle 4 is that commitments are ways of allocating resources, and any resource-bounded activity can be represented as negotiation among participants with conflicting commitments. As the availability of resources is always linked to the activities of other agents, it is clear that the commitments of the collection of agents is a influence on resource use. In fact, remaining consistent with the social notion of commitment introduced in Section 3, we can see a commitment as simply *the use of resources.* Commitment in this sense necessarily has future implications: actor A's use of resources for one purpose in the present constrains A's (and others') choices in the future. (The economic notion is "opportunity cost.") Commitments thus "flow" through resources. (OISS would say that resources participate in Systems Commitments.) Moreover, the possibilities for resource allocation now (e.g., the amount of resources available) is a result of other *prior* commitments of many agents, including those of A. Becker's notion of being committed to a course of action through a collection of "side bets,"—other courses of action related through resource dependencies—also falls out of this conception [Becker60]. So do the observations that resource constraint reduces the range of practical choice of a course of action, locking the agent in ("beggars can't be choosers") and slack resources reduce commitment by opening a greater number of practical courses of action ("The rich can do what they want").

4.5 Principle 5: Joint Qualification, Representation Incommensurability, Failure Indeterminacy.

These three problems are simply designed out of contemporary DAI systems—or rather brittleness in the face of them is designed in. There has been little or no attempt to grapple with them, in large part, because there has been so little attention to the automated formulation of problems [Bond88a] or with collaborative learning. Computational approaches to the construction of scientific knowledge and scientific explanation, have in general been quite naive about scientific practice and the nature of explanation [Social Studies89], a multi-agent arena in which failure indeterminacy appears routinely. In general, joint qualification and representation incommensurability are handled by assuming a global semantics for a system, and working within the constraints of the theory-relativity of that semantics. Failure indeterminacy has been dealt with via generalization [Huhns87] and model-based reasoning [Hudlicka87], but these are not essentially distributed approaches, nor have they been implemented under assumptions of joint qualification problems and representational incommensurability.

OISS allows us to consider several of the concepts embodied in Principle 6. First, OISS takes for granted that participants have fundamentally local and separate representations, and thus are subject to each of these problems. OISS presents a single framework—conflicting Systems Commitments—that integrates representational incommensurability with other levels of discrepancy mentioned in DAI Problem 5 of Section 2. The OISS approach to the joint qualification problem is to negotiate qualification discrepancies when they become manifest. OISS embeds the qualification problem in a situated process, and makes its solution responsive to local contingencies. Since negotiations can set precedents, the foundation for stable joint qualifications is laid in OISS. The OISS definition of *cooperation*, "mutually dependent roles in a Systems Commitment," is also a statement of the joint qualification problem. It does make the link to mutually supportive commitments, (i.e. those that allow resources to flow in both directions) which are the foundation of an approach to joint qualification (cf. the discussions of Principle 4 above). OISS deals with representation incommensurability through the mechanism of recursive negotiation, if at all. It is not clear that OISS recognizes representation incommensurability as a key problem, and any treatment it would have would be necessarily incomplete, because the treatments of commitment and action-at-a-distance are not well-integrated. Likewise, failure indeterminacy is not accounted for, again because the development and integration of commitment is weak.

4.6 Principle 6: Robust Joint Courses of Action and Knowledge.

Current DAI systems and theories achieve robustness through several mechanisms, which primarily are founded on either triangulation of multiple perspectives, redundancy and slack resources, or pre-specification of the causes and possibilities of failure. Several methods for robust problem solving under uncertainty that exploit multiple perspectives have been discussed in Section 4.3 above. A number of DAI systems rely upon redundancy available through parallelism to guard against failure or overload, and these have proven robust in practice. There have been few tests of the performance and overload limits of various DAI approaches, or the limits of organizational and coordination forms. Malone has given a characterization of the susceptibility of various organizational structures to node failure [Malone87] under particularly rigid interaction assumptions. Several approaches to multiagent planning attempt to iron out the contingencies of interaction before plan execution by interleaving partial orders of concurrent actions [Georgeff87]; these are not properly in the domain of open-systems approaches.

One reason multiple agents and openness matter to DAI is because of a basic difficulty in building reliable distributed decisionmaking systems, that is accounted for in the conceptual machinery of OISS, but not in conventional DAI. As the nationwide nine-hour telephone network service interruption of January, 1990 illustrates [McCarroll90], systems which depend upon shared knowledge, common semantics, and global conceptions of coherence (e.g., identical programs at each node of a network with identical decision rules) can be subject to cascading catastrophic failures (see also [Gasser89a, Huberman88]). On the other hand, without common semantics and global conceptions, interoperability and relability become difficult for other reasons. In OISS terms, greater self-reliance produces inherent conflicts of commitment, e.g, to decision rules or communication protocols. The OISS concepts of arms-length relationships, local and multiple authority, asynchrony, self-reliance and openness are useful conceptual tools here. The fact that the network service disruptions were never complete, and operations could be restored in nine hours (i.e. the network *was* a robust and continuable process) can only be explained by reference to the existence of multiple authorities and arms-length relationships, including at least the authority of telephone engineers at multiple sites over the behavior of network nodes, and their alternative decisionmaking activities. The network standing alone could not restore itself, and the OISS concepts help us to focus on the the actual actors doing the job, not just the network itself as a unit of analysis.

The OISS perspective makes use of the concept "in place" but this is not well-enough defined to be sensible. The notion of robustness, defined as keeping commitments in the face of conflict, is coherent with a concept of a continuable joint action. Further elaboration and mechanisms, however, are lacking. Earlier we spoke of commitment as the outcome of joint courses of action woven together. The notion of Systems Commitments being in place could be defined by using metacommitment (commitment to commitment) but this is not been done in OISS.

What OISS needs is a way of linking particular negotiation contexts and particular kinds of commitment to particular ways of achieving robust joint courses of action. Latour [Latour88] uses the image of an army made invincible by association with numerous allies, as a way of explaining robust joint courses of action. OISS must integrate similar images with computational mechanisms.

5 A Synthesis

A key missing link in OISS and the other new approaches discussed in this paper, at the moment, is how to make them computational. Because commitment has been posed as a foundational concept, let us briefly examine some of the computational questions surrounding it, to see the directions we might take to construct a more computational theory based on extensions to OISS. Cohen and Levesque's construction of commitment, which is to date the most sophisticated mathematical model, is based on representing commitment using notions of "belief" and "goal," and then computing whether an agent is committed based on the logical entailments of its beliefs. Their commitment is laced into a series of decisions, and at any one of them an agent has to deduce whether it is still committed to a goal. As we discussed in Section 3, this idea of being committed is something local to the agent, and local to its viewpoint. In contrast, the OISS notion of commitment as *Systems Commitment*, though ill-defined, has roots in a basically distributed framework of multiple agents being committed together. But how can such a notion be made both computational and non-local?

One way is to begin to develop theories that dissolve the distinction between open and closed systems, that consider all systems as fundamentally open ones, and that focus on mechanisms for weaving webs of commitment as ways of achieving robustness, joint action, and plausible knowledge. When an actor is committed in the social sense of Becker and Gerson it is constrained to a course of action because of its particular local sovereignty. Establishing commitments in a manner that is both social and computational means setting up numerous side bets that constrain an agent's field of choice. Computing commitment means setting up relationships of mutual influence with additional agents[13]. There are two ways to do this that are already familiar to the world of DAI: passing self-descriptions, and developing checks and balances. Currently, these are only minimal parts of the OISS analysis.

A promising approach to distributed computational commitment is based on agents modeling one another and exchanging *self descriptions*. This approach is a foundation of the MACE system, has been exploited in DATMS [Mason89], PGPs [Durfee89], and various network protocols, and has foundations in what Mead [Mead34] saw as a concept that could unify concepts of the self and of society: the process of "taking the role of the other." Self descriptions can be as simple as an address at which to receive messages, or as complex as a rich knowledge model of an agent. Variance in the ability to self-describe and to incorporate self-descriptions into action differentiate the interactive power of participants.

The flexible composition mechanism of *actor configurations* introduced by Agha and Hewitt [Agha86], is based on actors' ability to pass self-descriptions—their mail addresses. The boundaries of a configuration are defined purely in terms of various actors' access to the addresses of other actors within and outside the configuration; this makes a configuration both a flexible and distributed, and in a sense defines limits of interaction and thus provides commitment.

Self-descriptions are also ways of embedding participants in many situations simultaneously. Commitment is generated to the extent those self-descriptions actually become a part of the calculus of action in those situations. If one agent takes another's self-description into account, it becomes committted to acting in a more constrained way. For example, in the network service interruption case, most nodes' decision algorithms did react to the overload indications in passed self-descriptions, which involved them all in joint courses of action that, on the whole, interrupted service. Passing self-descriptions and meta self-descriptions can also increase sovereignty in distributed ways, by increasing local awareness of how to adapt (e.g., [Durfee87b, Lesser83]); this was the aim of the phone network self-descriptions but there was a missing link: checks and balances.

Including a collection of checks and balances (plurality) in a DAI system, so that different participants have control over different resources in critical interactions and no participant can be ignored, is another computational approach to OISS. To some extent this notion has already been

[13]It means also being honest about what those alliances are - whether they're a property of the programmer of the physical world, or of the "knowledge" in the system.

built into the convergent multiperspective approaches such as FA/C problem-solving, PGPs, and the DATMS. Building in plurality also means that self-descriptions are necessarily involved in joint courses of action. This gives us preliminary tools for implementing a balance between skepticism and involvement or self-reliance and interdependence.

Finally, by creating system-building mechanisms that treat the nature of systems as fundamentally open, we construct for ourselves another paradox—namely, do we really need new mechanisms? Suppose we change focus from problems of reasoning *across* participants with their own microtheories (which is the focus of OISS) to instead understanding the *processes of establishing and changing the boundaries of microtheories* (e.g., by understanding how to change the mix of actors participating in pragmatically common viewpoints, or by understanding how to effect closure by building denser webs of commitment). Then we have also removed the distinction between two approaches to system building (i.e. using microtheories or using OISS methods), and replaced it with a distinction in points of view toward *any* systems we build. Said another way, a viewpoint based on OISS and social foundations for DAI will provide new ways of explaining how and why *existing* reasoning paradigms work, and how the rethink their boundaries and the participants and work they leave out. In effect, we will be saying that we don't necessarily need new programming foundations, but we do need new theoretical foundations for explaining how and why existing programming foundations have the effects they do.

6 Conclusions

It is clear and not completely surprising that there are several problems with using deductive logic as a foundation for problem solving in open systems; these include Hewitt's deductive indeterminacy problem, as well as others such as the failure indeterminacy, representational incommensurability, and joint qualification problems. Since any deductive theory depends upon precursors such as a universe of discourse, a model, etc. it doesn't seem unreasonable to say that when multiple viewpoints are at stake, logic may fail.

Defining and exemplifying the problems of deduction, open systems, or DAI is an exercise - finding solutions appears to require some new foundations for knowledge and interaction, if deductive logic can't be used. OISS tells us to build our analytical foundations on Trials of Strength, and Systems Commitments. But it doesn't tell us how to win particular Trials or how to organize particular sets of Systems Commitments, and this is typically what engineers want. OISS and others do tell us that we cannot hope to be sure of organizing and winning some of them.

Neither this paper nor OISS is trying to criticize useful reasoning mechanisms that work within bounded microtheories. It does appear to me that OISS, coupled with some of the conceptions outlined in this paper, can begin account for both the processes of delimiting microtheories in practice and the processes of reasoning employed within them, while the converse is not the case. Our major focus is upon open large-scale multi-agent systems. Both OISS and I propose approaches based upon commitments, resource allocations and interaction, and a notion of meaning independent of Tarskian semantics or possible worlds. Though thoughtful, the proposal is just that. Nonetheless, we must make a start somewhere, and the place to begin again seems to be an examination of the processes of human interaction, social organization, and conccurrent systems—and thus the distributed foundations of knowledge and interaction.

7 Acknowledgements

Continuing discussions over many years with Elihu Gerson of the Tremont Research Institute have been invaluable in formulating the ideas in this paper. Carl Hewitt was generous enough to provide a number of clarifications of his work and critical comments on mine. I am also grateful for the comments of Phil Agre, Alan Bond, Geoff Bowker, Phil Cohen, Kari T. Eloranta, Rick Hull, Dean Jacobs, Izhar Matzkevitch, and Nick Rouquette for helping to improve the presentation of several concepts, and to David Kirsh for gently shepherding the process. I thank Susan Williams for her comments and for vigorous, unflinching editing, and M. Sue Gerson for her great herb tea at key moments. This research is partially supported by a grant from the AT&T Affiliates Program.

References

[Agha86] Gul Agha. *Actors: A Model of Concurrent Computation in Distributed Systems.* MIT Press, Cambridge, MA, 1986.

[Agre85] Philip E. Agre. Routines. MIT AI Lab Memo 828, MIT, Cambridge, MA, 1985.

[Agre88] Philip E. Agre. *The Dynamic Structure of Everyday Life.* PhD thesis, Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, 1988.

[Becker60] Howard S. Becker. Notes of the Concept of Commitment. *American Journal of Sociology*, 66:32–40, July 1960.

[Becker86] Howard S. Becker. *Doing Things Together.* University of Chicago Press, Chicago, IL, 1986.

[Bentley54] Arthur F. Bentley. The Human Skin: Philosophy's Last Line of Defense. In Sidney Ratner, editor, *Inquiry into Inquiries*, chapter 10, pages 195–211. Beacon Press, Boston, MA, 1954.

[Bond88a] Alan H. Bond and Les Gasser. An Analysis of Problems and Research in Distributed Artificial Intelligence. In *Readings in Distributed Artificial Intelligence.* Morgan Kaufmann Publishers, San Mateo, CA, 1988.

[Bond88b] Alan H. Bond and Les Gasser. *Readings in Distributed Artificial Intelligence.* Morgan Kaufmann Publishers, San Mateo, CA, 1988.

[Bond90] Alan H. Bond. Commitment: A Computational Model for Organizations of Cooperating Intelligent Agents. In *Proceedings of the 1990 Conference on Office Information Systems*, Cambridge, MA, April 1990.

[Buss87] Leo W. Buss. *The Evolution of Individuality.* Princeton University Press, Princeton, NJ, 1987.

[Chandrasekn81] B. Chandrasekaran. Natural and Social System Metaphors for Distributed Problem Solving: Introduction to the Issue. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-11(1):1–5, January 1981.

[Chapman87] David Chapman and Philip E. Agre. Abstract Reasoning as Emergent from Concrete Activity. In Michael P. Georgeff and Amy L. Lansky, editors, *Reasoning About Actions and Plans: Proceedings of the 1986 Timberline Workshop*, pages 411–424. Morgan Kaufmann, San Mateo, CA, 1987.

[Clarke90] Adele E. Clarke and Elihu M. Gerson. Symbolic Interactionism in Social Studies of Science. In Howard S. Becker and Michael McCall, editors, *Symbolic Interactionism and Cultural Studies.* University of Chicago Press, Chicago, IL, 1990.

[Cohen79] Philip R. Cohen and C. Raymond Perrault. Elements of a Plan-Based Theory of Speech Acts. *Cognitive Science*, 3(3):177–212, 1979.

[Cohen87] Philip R. Cohen and Hector J. Levesque. Intention = Choice + Commitment. In *Proceedings of AAAI-87*, pages 410–415, 1987.

[Cohen90] Philip R. Cohen and Hector J. Levesque. Intention is Choice with Commitment. *Artificial Intelligence*, 42(3), 1990.

[Collins85] Harry M. Collins. *Changing Order: Replication and Induction in Scientific Practice.* Sage, Beverly Hills, CA, 1985.

[Collins90] Harry M. Collins. *Artificial Experts: Social Knowledge and Intelligent Machines.* MIT Press, Cambridge, MA, 1990.

[Corkill86] Daniel D. Corkill, Kevin Q. Gallagher, and Kelly E. Murray. GBB: A Generic Blackboard Development System. In *Proceedings of AAAI-86*, pages 1008–1014, August 1986.

[Davis80] Randall Davis. Report on the Workshop on Distributed AI. *SIGART Newsletter*, 73:42–52, October 1980.

[Davis83] Randall Davis and Reid G. Smith. Negotiation as a Metaphor for Distributed Problem Solving. *Artificial Intelligence*, 20(1):63–109, 1983.

[Decker87] Keith S. Decker. Distributed Problem-Solving Techniques: A Survey. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-17(5):729–740, 1987.

[Decker89] Keith S. Decker, Edmund H. Durfee, and Victor R. Lesser. Evaluating Research in Cooperative Distributed Problem Solving. In L. Gasser and M.N. Huhns, editors, *Distributed Artificial Intelligence, Volume II*, pages 487–519. Pitman/Morgan Kaufmann, London, 1989.

[Dewey96] John Dewey. The Reflex Arc Concept in Psychology. *Psychological Review*, III:357–370, July 1896.

[Durfee87a] Edmund H. Durfee and Victor R. Lesser. Using Partial Global Plans to Coordinate Distributed Problem Solvers. In *Proceedings of IJCAI-87*, pages 875–883, 1987.

[Durfee87b] Edmund H. Durfee, Victor R. Lesser, and Daniel D. Corkill. Coherent Cooperation Among Communicating Problem Solvers. *IEEE Transactions on Computers*, C-36:1275–1291, 1987.

[Durfee89] Edmund H. Durfee and Victor R. Lesser. Negotiating Task Decomposition and Allocation Using Partial Global Planning. In L. Gasser and M.N. Huhns, editors, *Distributed Artificial Intelligence, Volume II*, pages 229–244. Pitman/Morgan Kaufmann, London, 1989.

[Erman88] Lee D. Erman, Jay S. Lark, and Frederick Hayes-Roth. ABE: An Environment for Engineering Intelligent Systems. *IEEE Transactions on Software Engineering*, SE-14(2):1758–1770, December 1988.

[Ferber88] Jacques Ferber and Jean-Pierre Briot. Design of a Concurrent Language for Distributed Artificial Intelligence. Technical Report LITP-88-57-RXF, Laboratoire Informatique, Theorique, et Programmation (LITP), Université Pierre et Marie Curie, 4 Place Jussieu, 75252 Paris, France, 1988.

[Fikes72] Richard E. Fikes, Peter E. Hart, and Nils J. Nilsson. Some New Directions in Robot Problem Solving. In Bernard Meltzer and Donald Michie, editors, *Machine Intelligence 7*, pages 405–430. John Wiley and Sons, New York, 1972.

[Fikes82] Richard E. Fikes. A Commitment-Based Framework for Describing Informal Cooperative Work. *Cognitive Science*, 6(4):331–347, 1982.

[Fujimura87] Joan H. Fujimura. Constructing Doable Problems in Cancer Research: Articulating Alignment. *Social Studies of Science*, 17:257–293, 1987.

[Garfinkel67] Harold Garfinkel. *Studies in Ethnomethodology.* Prentice-Hall, Englewood Cliffs, NJ, 1967.

[Garfinkel81] Harold Garfinkel, Michael Linch, and Eric Livingston. The Work of a Discovering Science Construed with Materials from the Optically Discovered Pulsar. *Philosophy of the Social Sciences*, 11:131–158, 1981.

[Gasser86] Les Gasser. The Integration of Computing and Routine Work. *ACM Transactions on Office Information Systems*, 4(3):205–225, July 1986.

[Gasser87] Les Gasser, Carl Braganza, and Nava Herman. MACE: A Flexible Testbed for Distributed AI Research. In Michael N. Huhns, editor, *Distributed Artificial Intelligence*, pages 119–152. Pitman Publishing/Morgan Kaufmann Publishers, San Mateo, CA, 1987.

[Gasser89a] Les Gasser and Michael N. Huhns, editors. *Distributed Artificial Intelligence, Volume II.* Pitman/Morgan Kaufmann, London, 1989.

[Gasser89b] Les Gasser, Nicholas F. Rouquette, Randall W. Hill, and Jon Lieb. Representing and Using Organizational Knowledge in DAI Systems. In L. Gasser and M.N. Huhns, editors, *Distributed Artificial Intelligence, Volume II*, pages 55–78. Pitman/Morgan Kaufmann, London, 1989.

[Gasser90] Les Gasser and Randall W. Hill. Engineering Coordinated Problem Solvers. In *Annual Reviews of Compuer Science*, volume 4. Annual Reviews, Inc., Palo Alto, CA, 1990.

[Georgeff87] Michael P. Georgeff. Planning. *Annual Reviews of Computer Science*, 2:359–400, 1987.

[Gerson76] Elihu M. Gerson. On Quality of Life. *American Journal of Sociology*, 41:793–806, October 1976.

[Gerson77] Elihu M. Gerson. Scientific Work and Social Worlds. *Knowledge*, 4:357–377, 1977.

[Hall85] Rogers P. Hall and Dennis Kibler. Differing Methodological Perspectives in Artificial Intelligence Research. *AI Magazine*, 6(3):166–178, Fall 1985.

[Hayes-Roth85] Barabara Hayes-Roth. A Blackboard Architecture for Control. *Artificial Intelligence Journal*, 26:251–321, 1985.

[Hewitt77] Carl E. Hewitt. Viewing Control Structures as Patterns of Passing Messages. *Artificial Intelligence*, 8(3):323–364, 1977.

[Hewitt84] Carl E. Hewitt and Peter de Jong. Open Systems. In *On Conceptual Modeling*, pages 147–164. Springer Verlag, New York, 1984.

[Hewitt85] Carl E. Hewitt. The Challenge of Open Systems. *Byte*, 10(4):223–242, April 1985.

[Hewitt86] Carl E. Hewitt. Offices are Open Systems. *ACM Transactions on Office Information Systems*, 4(3):271–287, 1986. Also in Huberman, B.A. *The Ecology of Computation*, Elsevier Science Publishers/North Holland, Amsterdam, 1988.

[Hewitt91] Carl E. Hewitt. Open Information Systems Semantics for Distributed Artificial Intelligence. *Artificial Intelligence*, This Volume, 1991.

[Hinke90] Thomas Hinke. *Query by Committee: Distributed, Data-Driven Problem Elaboration.* PhD thesis, Department of Computer Science, University of Southern California, Los Angeles, CA, (in preparation) 1990.

[Huberman88] Bernardo Huberman, editor. *The Ecology of Computation.* Elsevier Science Publishers/North Holland, Amsterdam, 1988.

[Hudlicka87] Eva Hudlicka and Victor R. Lesser. Modeling and Diagnosing Problem Solving System Behavior. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-17(3):407–419, May/June 1987.

[Hughes83] Thomas Hughes P. *Networks of Power: Electric Supply Systems in the US, England, and Germany.* Johns Hopkins University Press, Baltimore, MD, 1983.

[Huhns87] Michael N. Huhns, Uttam Mukhopadhyay, Larry M. Stephens, and Ronald D. Bonnell. DAI For Document Retrieval: the MINDS Project. In Michael N. Huhns, editor, *Distributed Artificial Intelligence*, pages 249–284. Pitman Publishing/Morgan Kaufmann Publishers, San Mateo, CA, 1987.

[King90] John Leslie King and Susan Leigh Star. Conceptual Foundation for the Development of Organizational Decision Support Systems. In Jay F. Nunamaker Jr., editor, *Proceedings of the Twenty-Third Annual Hawaii International Conferenc on System Sciences*, volume III, pages 143–151, Los Alamitos, CA, January 1990. IEEE Computer Society Press.

[Kling82] Rob Kling and Walter S. Scacchi. The Web of Computing: Computer Technology as Social Interaction. In Marshall C. Yovits, editor, *Advances in Computers*, pages 1–89. Academic Press, New York, 1982.

[Kornfeld81] William A. Kornfeld and Carl E. Hewitt. The Scientific Community Metaphor. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-11(1):24–33, January 1981.

[Kuwabara89] K. Kuwabara and V.R. Lesser. Extended Protocol for Multistage Negotiations. In Miroslav Benda, editor, *Proceedings of the 9th Workshop on Distributed Artificial Intelligence*, pages 129–162, Bellevue, Washington, September 1989.

[Lakatos76] Imre Lakatos. *Proofs and Refutations: The Logic of Mathematical Discovery.* Cambridge University Press, New York, 1976.

[Latour79] Bruno Latour and Steve Woolgar. *Laboratory Life: The Social Construction of Scientific Facts.* Sage, Beverly Hills, CA, 1979.

[Latour87] Bruno Latour. *Science in Action.* Harvard University Press, Cambridge, MA, 1987.

[Latour88] Bruno Latour. *The Pasteurization of France, with Irreductions.* Harvard University Press, Cambridge, MA, 1988.

[Lave88] Jean Lave. *Cognition in Practice.* Cambridge University Press, New York, 1988.

[Laymon85] Ronald Laymon. Idealization and the Testing of Theories by Experimentation. In Peter Achinstein and Owen Hannaway, editors, *Observation, Experiment, and Hypothesis in Modern Physical Science*, chapter 6, pages 147–173. Bradford Books/MIT Press, Cambridge, MA, 1985.

[Lenat75] Douglas B. Lenat. Beings: Knowledge as Interacting Experts. In *Proceedings of IJCAI-75*, pages 126–133, 1975.

[Lesser80] Victor R. Lesser and Lee D. Erman. Distributed Interpretation: A Model and Experiment. *IEEE Transactions on Computers*, C-29(12):1144–1163, December 1980.

[Lesser81] Victor R. Lesser and Daniel D. Corkill. Functionally Accurate, Cooperative Distributed Systems. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-11(1):81–96, January 1981.

[Lesser83] Victor R. Lesser and Daniel D. Corkill. The Distributed Vehicle Monitoring Testbed: A Tool For Investigating Distributed Problem Solving Networks. *AI Magazine*, pages 15–33, Fall 1983.

[Lesser87] Victor R. Lesser and Daniel D. Corkill. Distributed Problem Solving. In Stuart C. Shapiro, editor, *Encyclopedia of Artificial Intelligence*, pages 245–251. John Wiley and Sons, New York, 1987.

[Lesser88] Victor R. Lesser, Jasmina Pavlin, and Edmund H. Durfee. Approximate Processing in Real Time Problem Solving. *AI Magazine*, 9(1):49–61, Spring 1988.

[MacFarlane78] Alan MacFarlane. *The Origins of English Individualism.* Cambridge University Press, New York, NY, 1978.

[Maes88] Pattie Maes. Computational Reflection. *The Knowledge Engineering Review*, 3(1):1–19, March 1988.

[Malone87] Thomas W. Malone. Modeling Coordination in Organizations and Markets. *Management Science*, 33(10):1317–1332, 1987.

[Mason89] Cindy L. Mason and Rowland R. Johnson. DATMS: A Framework for Distributed Assumption-Based Reasoning. In L. Gasser and M.N. Huhns, editors, *Distributed Artificial Intelligence, Volume II*, pages 293–318. Pitman/Morgan Kaufmann, London, 1989.

[McCarroll90] Thomas McCarroll and Paul A. Witteman. Ghost in the Machine. *Time Magazine*, 135(5):58–59, January 29, 1990.

[McCarthy77] John McCarthy. Epistemological Problems of Artificial Intelligence. In *Proceedings of IJCAI-77*, pages 1038–1044, 1977. Reprinted in Ron Brachman and Hector Levesque, eds., *Readings in Knowledge Representation*, Morgan Kaufmann Publishers, 1985, pages 24–30.

[Mead34] George Herbert Mead. *Mind, Self, and Society.* University of Chicago Press, Chicago, IL, 1934.

[Moore77] Robert C. Moore. Reasoning about Knowledge and Action. In *Proceedings of IJCAI-77*, pages 223–227, Cambridge, MA, August 1977. International Joint Conferences on AI.

[Nii89] H. Penny Nii, Nelleke Aiello, and James Rice. Experiments on CAGE and POLIGON: Measuring the Performance of Parallel Blackboard Systems. In L. Gasser and M.N. Huhns, editors, *Distributed Artificial Intelligence, Volume II*, pages 319–384. Pitman/Morgan Kaufmann, London, 1989.

[Quine81] W. V. Quine. *Theories and Things.* Harvard University Press, Cambridge, MA, 1981.

[Rosenschein82] Jeffrey S. Rosenschein. Synchronization of Multi-Agent Plans. In *Proceedings of AAAI-82*, pages 115–119, 1982.

[Rosenschein85] Jeffrey S. Rosenschein and Michael R. Genesereth. Deals Among Rational Agents. In *Proceedings of IJCAI-85*, pages 91–99, August 1985.

[Sathi89] A. Sathi and M.S. Fox. Constraint-Directed Negotiation of Resource Reallocations. In L. Gasser and M.N. Huhns, editors, *Distributed Artificial Intelligence, Volume II*, pages 163–194. Pitman/Morgan Kaufmann, London, 1989.

[Sethi89] Ravi Sethi. *Programming Languages: Concepts and Constructs.* Addison Wesley, Reading, MA, 1989.

[Smith80] Reid G. Smith. The Contract Net Protocol: High-Level Communication and Control in a Distributed Problem Solver. *IEEE Transactions on Computers*, C-29(12):1104–1113, December 1980.

[Smith86] Brian Cantwell Smith. Varieties of Self-Reference. In Joseph Y. Halpern, editor, *Proceedigs of the 1986 Conference on Theoretical Aspects of Reasoning About Knowledge*, pages 19–43. Morgan Kaufmann, March 1986.

[Social Studies89] Miscellaneous authors. Special Issue Symposium on 'Computer Discovery and the Sociology of Scientific Knowledge'. *Social Studies of Science*, 19(4):563–695, November 1989.

[Star89a] Susan Leigh Star. *Regions of the Mind: Brain Research and the Quest for Scientific Certainty.* Stanford University Press, Stanford, CA, 1989.

[Star89b] Susan Leigh Star. The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving. In L. Gasser and M.N. Huhns, editors, *Distributed Artificial Intelligence, Volume II*, pages 37–54. Pitman/Morgan Kaufmann, London, 1989.

[Strauss87] Anselm L. Strauss. *Qualitative Analysis for Social Scientists.* Cambridge University Press, New York, 1987.

[Suchman87] Lucy Suchman. *Plans and Situated Actions: The Problem of Human-Machine Communication.* Cambridge University Press, New York, 1987.

[Sycara89] Katia P. Sycara. Multiagent Compromise via Negotiation. In L. Gasser and M.N. Huhns, editors, *Distributed Artificial Intelligence, Volume II*, pages 119–138. Pitman/Morgan Kaufmann, London, 1989.

[Teil90] Genevieve Teil, Madeleine Akrich, Bertrand Michelet, and Bruno Latour. The Hume Machine: Can Association Networks do More That Formal Rules? Centre de Sociologie de l'Innovation, Ecole Nationale Supérieure des Mines de Paris, 1990. Submitted for Publication.

[Wesson81] Robert B. Wesson, Frederick A. Hayes-Roth, John W. Burge, Cathleen Stasz, and Carl A. Sunshine. Network Structures for Distributed Situation Assessment. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-11(1):5–23, January 1981.

[Wimsatt80] William Wimsatt. Reductionist Research Strategies and Their Biases in the Units of Selection Controversy. In T. Nickles, editor, *Scientific Discoveries: Case Studies*, pages 213–259. D. Reidel, Boston, 1980.

[Winograd86] Terry Winograd and Fernando Flores. *Understanding Computers and Cognition.* Ablex, Norwood, NJ, 1986.