

# Cognitive Science Meets Multi-Agent Systems: A Prolegomenon

Ron Sun  
CECS Department, University of Missouri - Columbia  
Columbia, MO 65211, USA

EMAIL: [rsun@cecs.missouri.edu](mailto:rsun@cecs.missouri.edu)

To appear in: *Philosophical Psychology*, Vol.14, No.1, pp.5-28, 2001.

# Cognitive Science Meets Multi-Agent Systems: A Prolegomenon

## **Abstract**

In the current research on multi-agent systems (MAS), many theoretical issues related to sociocultural processes have been touched upon. These issues are in fact intellectually profound and should prove to be significant for MAS. Moreover, these issues should have equally significant impact on cognitive science, if we ever try to understand cognition in the broad context of socio-cultural environments in which cognitive agents exist. Furthermore, cognitive models as studied in cognitive science can help us in a substantial way to better probe multi-agent issues, by taking into account essential characteristics of cognitive agents and their various capacities. In this paper, we systematically examine the interplay among social sciences, MAS, and cognitive science. We try to justify an integrated approach for MAS that incorporates different perspectives. We show how a new cognitive model CLARION can embody such an integrated approach through a combination of autonomous learning and assimilation.

## 1 Introduction

Multi-agent Systems are computational systems in which multiple autonomous agents work together to perform some set of tasks to satisfy some set of goals. These systems may involve computational agents that are homogeneous or heterogeneous; they may involve agents either having a common goal or having goals that are distinct or even contradictory; there may be varying degrees of communications during their interaction. The study of multi-agent systems raises many specific issues. These issues include how to develop coordination strategies (that enable groups of agents to solve problems together effectively), negotiation mechanisms (that serve to bring a set of agents together), conflict detection and resolution strategies, and other mechanisms whereby agents can contribute to overall system effectiveness while still assuming autonomy. The study of how organizations can be formed, structured, and utilized has also been undertaken. The broad fundamental issues that have been studied include those shown in Figure 1. The complexity of work on multi-agent systems is fast approaching the complexity dealt with by sociology and economics, although in practice, work is focused more on a smaller number of agents and usually without complex social organizations (with some exceptions, in e.g. computational organizational theory). Notably, however, most of the work is from a purely computational standpoint and assumes very rudimentary models of agents (e.g., Shoham 1993, Ross 1973).

On the other hand, although a significant amount of work has been done in cognitive science for studying individual cognition (Anderson 1983, Anderson and Lebiere 1998, Holland et al 1986, Klahr et al 1987, Osherson and Lasnik 1990, Rumelhart and McClelland 1986, Bechtel and Graham 1998, Sun et al 1998 a, b, 1999), the sociocultural processes and their relation to cognition have never been a focus of cognitive science. As a result, the sociocultural aspects of cognitive processes are not well understood (although there has been work on these aspects; e.g. Suchman 1987, Lave 1988, Barkow et al 1992, Hutchins 1995, Bruner 1995). There were even explicit calls for cognitive scientists to ignore such aspects in studying cognition (e.g., Fodor 1980).

However, we believe that MAS (and to some degree, social and economic theories) needs cognitive science, because it needs better understanding and better models of individual cognition, only based on which it can develop better models of aggregate processes (through multi-agent interaction) (cf. Ross 1973, Wilson 1975, Shoham 1993). Cognitive models may provide a more realistic basis for understanding multi-agent interaction, by embodying realistic constraints, capabilities, and tendencies of individual agents in their interaction with their environments, which include both physical and social environments. This point has been made by others too, e.g., in the context of cognitive realism of game theory (Kahan and Rapaport 1984).

Conversely, cognitive science needs MAS too. Cognitive science is very much in need of new theoretical frameworks and new technical tools, especially for analyzing sociocultural aspects of cognition and cognitive processes involved in multi-agent interaction. It needs computational models and theories from MAS, but it also needs deeper theories and broader conceptual frameworks that can only be found in sociological and anthropological work (which evidently has a long and honorable intellectual history).

In studying cognition, we may identify two kinds of knowledge that an agent possesses based on

Cooperation, coordination, and conflict  
Communication  
Negotiation (in both competitive and cooperative situations)  
Distributed consensus  
Resource allocation  
Learning and adaptation

Multi-agent cooperative reasoning  
Distributed search  
Multi-agent planning

Social structures  
Organization, and organizational knowledge

Figure 1: Issues in Multi-Agent Systems

their origins: one kind comes from individual sources and the other from social sources (which are admittedly two extreme cases, since much knowledge involves both sources). The individual sources are made up of learning processes involving an agent’s interacting with its environment (both sociocultural and physical) and the resulting (implicit or explicit) representations. The other kind of knowledge involves social sources, which result from concepts and conventions formed in broader settings and are somehow “transmitted to” (or even “forced upon”) individuals. This is an important process in cognition, when mixed with the individual process. Without it, it would be hard to account for human cognition in its entirety. We need to consider not only individually acquired concepts but also collectively acquired concepts, not only self-generated knowledge but also socially transmitted knowledge, and overall, not only an agent in a world alone but also interaction and sociocultural processes among multiple agents and through social institutions.

In understanding this process, we need to gain much better understanding of sociocultural dynamics. This is true for both MAS and cognitive science. We need to draw on sociological theories and insights, which will not only help us to develop sociocultural processes in MAS, but also help us to capture sociocultural influences on individual cognition. We need to understand both macroscopic and microscopic processes.

In this paper, we will argue for an integration of these strands. In section 2, we will discuss various issues concerning sociocultural processes, drawing ideas from branches of social sciences. In section 3, we will in turn look into the impact of such sociocultural processes on individual cognition. In section 4, we will look into an example cognitive model and see how sociocultural processes can be embodied by it. Section 5 will then speculate on how such a model can be used to investigate sociocultural issues related to MAS. Section 6 will conclude the paper.

## 2 Aspects of Sociality

Let us first examine some social issues raised and touched upon by MAS research, in the broader context of social sciences, so as to have a *deeper* look at the issues than what is commonly done in MAS.

## 2.1 Sociality

What counts as a sociocultural process? What is the difference between sociocultural and individual processes (for example, in action, learning, or concept formation)? These are the questions that we need to get a handle on.

Max Weber (1957) defined social actions to be those “where the actor’s behavior is meaningfully oriented to that of others. For example, a mere collision of two cyclist may be compared to a natural event. On the other hand, their attempts to avoid hitting each other, or whatever insults, blows, or friendly discussion might follow the collision, would constitute social ‘action’ ”. Social processes begin when there are multiple autonomous agents each of which acts on its own while taking account of the actions of other agents. When each agent tries to orient its behavior toward other agents simultaneously and continuously, a complex dynamic process results. A culture is formed when certain conventions and conceptions are commonly assumed by a social group. Culture shapes the dynamic process of social interaction in a profound way. Conversely, the process of social interaction also shapes culture itself, through individual or collective actions. Therefore, this nexus can be viewed as one single process: a sociocultural process above and beyond individual cognitive processes.

Sociology and anthropology, as well as as other branches of social science, have been dealing with these issues for more than a century. Sociology aims to describe, understand, and explain social facts with clear, neutral, and abstract concepts.<sup>1</sup> Various schools of thought exist that are radically different from each other. The rivalry can be between collectivist sociology (see, for example, Durkheim 1962 and Bourdieu and Wacquant 1992) and individualistic sociology (for example, phenomenological and interpretive studies; see Geertz 1973 and Schutz 1967); or between functionalism, system theory, and symbolic interactivism; or between structuralism and anti-structuralism. Some schools of sociology are close to philosophy (especially phenomenological philosophy) while others seek to be quantitative and exact.<sup>2</sup> However, for our purposes, each of them provides some useful insight and a vocabulary with which we can further our understanding. Sociology remains the main intellectual source of perspectives on the problems of culture, society, and civilization, which are inescapable in the study of both individual and collective cognition.

## 2.2 Society

Let us discuss in some detail major factors and aspects of society, in preparation for later discussions of these factors and aspects in cognition.

### 2.2.1 Social Structures

Social structures are the enduring, orderly, and patterned relationship among elements in a society (such as groups and hierarchies). They are the results of both biological evolution and evolution of complex social interaction (Caporael 1995). Simple group behavior, especially in lower species, may be

---

<sup>1</sup>It has deep intellectual roots in philosophy, history, and legal theories. Since the term “sociology” was coined by Auguste Comte in early 19th century, it has been developed vigorously, especially in the 19th century by Emile Durkheim in France, Max Weber in Germany, as well as many sociologists in America.

<sup>2</sup>For example, the majority of American sociology is aimed at being a quantitative science that influences the formation of public policy.

considered the direct result of biological adaptation.<sup>3</sup> More complex kinds of social structures, such as those found in human societies, historically or current, are the results of complex biological, social, cultural, and cognitive adaptation that goes beyond simple biological processes (i.e., extragenetic). The interplay of these factors (biological, sociocultural, and cognitive), acting through individual agents, gives rise to a variety of different forms of structures (Caporael 1995); these structures in turn impose themselves on individuals (Eggertsson 1990). Complex social structures are maintained through social control by a combination of means (such as common values, coercion, and so on).

Social institutions, which themselves constitute a complex and important type of structure, reinforce and support existing social structures. They include kinship institutions, economic institutions, political institutions, stratification institutions, and cultural institutions. They are formed through certain social processes in which certain social practices become regular and continuous (i.e., “institutionalization”). Although social institutions are relatively stable (as they are means for maintaining stability of social structures), it is clear that they are not immutable. Established social institutions can drift overtime and change drastically occasionally. The question is how they persist, function, and change.

## 2.2.2 Culture

Culture is a (structured) system of socially formed and transmissible patterns of behavioral and cognitive characteristics of individuals, within a social group. It includes conventions, rules, social institutions, significant common symbols, common behavioral patterns, characteristic attitudes, shared values, shared skills, beliefs, knowledge, myths, rituals, and so on.

The notion of culture as a collective entity is akin to the notion of a *scientific paradigm* as proposed by Kuhn (1962): It is shared among the majority of participants; it is generally observed but not always followed; it is relatively stable but can sometimes undergo changes or even radical revisions; it is clear to those involved but may or may not be articulated in explicit ways (Caporael 1995). With regard to the last point, paradigms or culture may involve both explicit or implicit representations (for example, simpler parts of it being explicit and the rest being implicit). In this regard, Kuhn is clearly not in favor of the idea of a scientific paradigm as a set of explicit rules: “I have in mind a manner of knowing which is misconstrued if reconstructed in terms of rules that are first abstracted from exemplars and thereafter function in their stead” (Kuhn 1962). Culture, as a generalization of scientific paradigms, may also consist of an articulated (explicit or conceptual) part (Smith et al 1992) and an unarticulated (implicit or subconceptual) part. This view is somewhat similar to the Jungian notions of collective consciousness and collective unconsciousness (minus any mysticism that may have come to be associated with them; Jung 1959). Bourdieu also adopts such a metaphor and sees the “socio-analysis” as a collective counterpart to Freudian psycho-analysis: It helps to unearth the social unconscious embedded into institutions as well as lodged inside individual agents.<sup>4</sup>

With regard to the unity of paradigms, Kuhn (1962) asked: “Why is the concrete scientific achievement, as a locus of professional commitment, prior to the various concepts, laws, theories, and points of view that may be abstracted from it? In what sense is the shared paradigm a fundamental unit for the student of scientific development, a unit that cannot be fully reduced to logically atomic components

---

<sup>3</sup>For example, an ant colony is formed through genetically predetermined means (Wilson 1975, Fetzner 1985, Deneubourg and Goss 1989), and it involves large-scale communication and coordination through simple biological means obtained evolutionarily.

<sup>4</sup>However, we shall hold that the social unconscious is an emergent property rather than existent in and by themselves.

which might function in its stead?” We can ask the same questions about culture in general; as has been observed in sociology, there seem to be a similar unity in culture as well. This in fact has also been the position of some important schools of sociologists. Emile Durkheim claimed that collective culture is not causally connected to “states of the consciousness of individuals but by the conditions in which the social group, in its totality, is placed.” This leads to the methodological holism, which bases its work on the premise that social systems have emergent properties that cannot be reduced to their constituting parts, so that the understanding of culture (and society in general) must start at the whole-system level. There are, however, alternatives to this view. The opposite of methodological holism, the methodological individualism, holds that culture and society are explicable purely based on cognitive processes (broadly construed) of individual agents. Methodological “situationalism” transcends the difference between these two schools and takes instead the very properties emerged from situational social interaction as a starting point in understanding culture and society (Bourdieu and Wacquant 1992).

In sum, it might be useful to view culture as a collective entity, but we shall keep in mind that it is made up of actions and beliefs of individual agents. Hence, it involves both explicit and implicit components, the same as individual cognition.

### 2.2.3 Social Determinism?

Social determinism has been popular in some segments of social science. However, this position seems extreme in many ways, as in the following claim: “Collective representations, emotions, and tendencies are caused not by certain states of the consciousness of individuals but by the conditions in which the social group, in its totality, is placed. .... These individual natures are merely the indeterminate material that the social factor molds and transforms” (Durkheim 1895/1962). Pierre Bourdieu posited, as the determining factor, an individual agent’s place in a “field” (which constitutes a social environment and has a stable structure that places different agents in different positions; Bourdieu and Wacquant 1992); individual’s cognition is determined through internalization of such an objective “reality”. As he put it, “persons, at their most personal, are essentially the personification of exigencies actually or potentially inscribed in the structure of the field or, more precisely, in the position occupied [by the agents] within this field.”

On the other hand, against the onslaught of social determinism, interpretive anthropology and sociology (especially the hermeneutic and the phenomenological varieties) try to ground the “objectivity” (of social environments) in individual minds. Phenomenological sociologists, such as Alfred Schutz (e.g., 1967), attempted to analyze, using the terminology of phenomenological philosophy, the construction of social reality, from the point of view of an individual agent (see also Putnam 1975). Geertz and other ethnographers tried to understand how people of different cultures comprehend and organize their world, through analyzing their explanations of their life experience (Geertz 1973). More recently, cognitive anthropology (see e.g. D’Andrade 1989) generally holds that a society is fully determined by individuals and culture is composed of psychological structures by means of which individual agents guide their behavior. That is to say, a society’s culture is made up entirely of rules in individual’s head; therefore, as soon as one explicates all the rules, one can fully understand a society and its culture. This position has been accused of, rightfully I believe, extreme subjectivism (by e.g. Bourdieu and Wacquant 1992). Much of the criticism against classical cognitivism (especially as discussed in Sun 2000, Brooks 1991, Lave 1988, Varela et al 1993, Bickhard 1998) can be applied to this position too.

In sum, to avoid extremity on either end, a proper balance between “objective” social reality and individual cognitive processes is needed, in order to understand the true natures of sociocultural processes and cognition.

### 3 Sociocultural Processes and Cognition

We turn now to the very issue of how social processes influence cognition of individual agents and their individual and collective behaviors.

#### 3.1 Inter-Agent Processes

Let us examine inter-agent processes, i.e., group processes within a small group. First, we note that many concepts are “social concepts”, that is, concepts formed necessarily in a social context through interacting with other agents. As an extremely simple example, even in Agre’s (1988) model of purely reactive (situated) agents, there are such social concepts; for example, the-block-that-can-use-to-kill-a-bee, the-bee-that-is-chasing-me, and so on are socially formed concepts (if we consider the “society” as consisting of the agent and the bee), since, to form such concepts, there needs to be an antagonist bee around.

L.S. Vygotsky’s description of the development of cognitive agents is illuminating. Vygotsky (1986) emphasized the *internalization* of social interaction as a major factor in the development of thinking in agents. One aspect of internalization is through the genesis of verbal (i.e., conceptual) thoughts (Vygotsky 1962). Speech develops before the development of internal verbal thinking. It starts at the single-word level which merely serves the function of labelling (that is, linking signs/symbols to their meanings/denotations). Such labeling itself is sociocultural since it is based on the established convention of a sociocultural and linguistic community. However, when more complex speech develops, it directly serves a social function (e.g., to get someone to do something). When the speech loses its social functions (for example, when nobody responds to a child’s request, as discussed by Vygotsky 1962), it can be turned inward and becomes a request/command to oneself (or in Vygotsky’s term, egocentric speech). Speech can thus be transformed from an interpersonal function to an intrapersonal function. Egocentric speech can be further turned inward and become internal verbal thoughts. Internal thinking relies on the internalized signs/symbols/concepts from social contexts, but can be accomplished without overt utterances or actions. According to Vygotsky, the development of thinking and behavior is the interweaving of the biological processes (or “elementary processes”; Vygotsky 1986) and sociocultural processes (or “higher psychological functions”, as manifested in speech and its internalization; cf. Barresi and Moore 1995).

A related issue in inter-agent interaction is communication. Communication can in part be accomplished through direct and unmediated comportment, the same way as other routine activities (Heidegger 1927a, Dreyfus 1992). A basis of this process is the shared world of the interacting agents and the shared understanding in this world. This shared background may or may not be explicitly represented, and may not be explicitly representable (Dreyfus 1972, 1992). We similarly need to see inter-agent interaction and communication partially as comportment or everyday coping (Dreyfus 1992), instead of being completely explicit.

## 3.2 Social Processes

Beside direct inter-agent interaction, we shall also consider the impact of sociocultural processes, social structures, and social institutions on the cognitive process of an agent. Although it has been recognized that the relation between an individual agent and society is complex and goes in both ways, it has also been recognized that, fundamentally, it is the influence from the society on individuals that matters more. This notion is best expressed as “power asymmetry” between individuals and societies (Geertz 1973). Individuals usually find themselves already “current in the community” (Heidegger 1927 b). Thus, their cognitive processes and the resulting behaviors are shaped by their surrounding social environments. Although such “shaping” may vary in degrees (since some are more engrossed in their particular social environments and cultures than others), the very existence of it as a major determining factor in individual’s cognition is undeniable.

In this regard, Bourdieu’s notion of “field” is illuminating. A social environment with a stable structure constitutes a “field” that places different social agents in different places playing different roles (much like a soccer field). An agent’s place in a “field” is (partially) determined by the structure of the field and the rule of the game, which are external “objective” reality.<sup>5</sup> In relation to cognition, on top of a “field”, there is always a sociocultural system of signs/symbols. The system is not only an instrument of knowledge, but also an “instrument of domination”. That is, it is an instrument for establishing and maintaining a “field” — a “peck order” among agents (e.g., with regard to distribution of resources; Wilson 1975, Fetzer 1985). The stability of social structures is helped by “the orchestration of categories of perception of the social world which, being adjusted to the divisions of the established order ....., impose themselves with all appearances of objective necessity” (Bourdieu 1984). Perceiving social reality as mere aggregates of individuals (and their volition, action, and cognition) misses the fact that the social structures are far more resilient than such a view would suggest and they *seem* to possess an objective and permanent character in their configurations rather than being instantaneously and arbitrarily constructed (Sowell 1996). Social structures determine to a large extent cognition of individuals.

Although the influence from societies to individuals is overwhelming, the influence in the other direction can nevertheless be discerned. As emphasized by phenomenological sociologists, social reality is an “ongoing accomplishment” actively constructed by agents through organized practices of everyday life. Social reality is, in some ways, an emergent product of the decisions, actions, and thinking of individual agents each of which has a direct, meaningful interaction with its world. The answer to this apparent dilemma is a proper mixing of the two perspectives. As Bourdieu (1984) put it, “objectivism and subjectivism, mechanism and finalism, structural necessity and individual agency are false antinomies. Each term of these paired opposites reinforce the other; all collude in obfuscating the anthropological truth of human practice.”

We believe that, in a mixed perspective, a basic element at the microscopic (or individual) level should be Heidegger’s notion of *facticity*, that is, the way the sociocultural (and natural) world appears to individuals as solid, taken for granted, and unalterable. In everyday existence, they can never get clear about their facticity, and therefore can never get rid of that facticity (Heidegger 1927a). The existential experience of an agent is *partially* structured by sociocultural signs/symbols and concepts/categories, which are formed within particular social structures and thus reflect them (as “instruments of domination”). These signs/symbols internalized by agents (cf. Vygotsky 1962) help to

---

<sup>5</sup>Bourdieu (see e.g. Bourdieu and Wacquant 1992) may take a more extreme position and claim that an individual agent’s place in a “field” is *fully* determined by the structure of the field and the rule of the game.

determine the cognition of an agent from within (while external stimuli provide external constraints). Certainly, as an autonomous agent, an individual may generate internally its own concepts/categories. But even internally generated concepts/categories of an agent may also reflect, to some degree (albeit a lesser degree), “objective” social reality, in the sense that these concepts/categories are influenced by given sociocultural signs/symbols and given social structures that have already been internalized. Through internalization, the agent’s conceptual processes are mediated (Wertsch 1998) by externally given concepts/signs/symbols (as well as their associated perspectives and biases).

Can we say the same about everyday routines activities or comportment (only involving subconceptual, or implicit, representations; Heidegger 1927 a) of an agent in a social world? We can, if we recognize that such routines are (in part) constrained by internalized social concepts/categories used in perceiving the world and they are cultivated in a particular social environment with given social structures that the agent sees. These everyday routines are thus in part sedimented social reality, and can only be applied within the particular sociocultural environment. Even the most immediate experience of agents cannot completely escape that characterization. This view of everyday routines (with implicit, subconceptual representations) is similar to the notion of *habitus* in Bourdieu’s (1984) theory (minus his social determinism). See also Wertsch (1991, 1998), Sowell (1996), and Eggertsson (1990).

In the other direction, in terms of forming particular social structures (“fields”) from the interaction of agents, there are some ideas from e.g. game theory (see Osborne and Rubinstein 1994) as well as from more recent MAS work (e.g., d’Inverno et al 1997, Ketchpel 1996, Sandholm and Lesser 1997, Salustowicz et al 1998). However, not enough emphasis has been placed on the *cognition* of individual agents in this process. Thus, existing models are simplistic and formalistic (although some evolutionary approaches seem more promising; Cosmides and Tooby 1989, Pinker 1994). In terms of generating sociocultural concepts/signs/symbols within social structures from a cognitive standpoint, there is however a shortage of good ideas and theories. We do not know clearly what kind of process is in the working and how we can best characterize it. We need better models, especially computational models, in order to study such an issue.

## 4 A Cognitive Model Addressing These Issues

Below we will present a computational cognitive model compatible with these above considerations. We will show the *connection* of the model to sociocultural processes studied in MAS, sociology, anthropology, and so on. We present this model as a case study of how theories of social sciences can benefit cognitive science, through enhancing cognitive models by taking into account social factors in cognition (which is deemed necessary for any accurate cognitive model), especially internalization of external sociocultural aspects.

### 4.1 The CLARION Model

First, we will describe a cognitive model. The model, named CLARION, has been described extensively in a series of previous papers, including Sun (1997), Sun and Peterson (1998 a, b), Sun et al (1998a, b, 1999). The model has also been successfully applied to capture various cognitive phenomena and processes (see e.g. Sun et al 1998a, b, 1999). Essentially, it consists of two levels, whereby the top level is conceptual and the bottom level subconceptual (comportment-oriented) (see also Smolensky

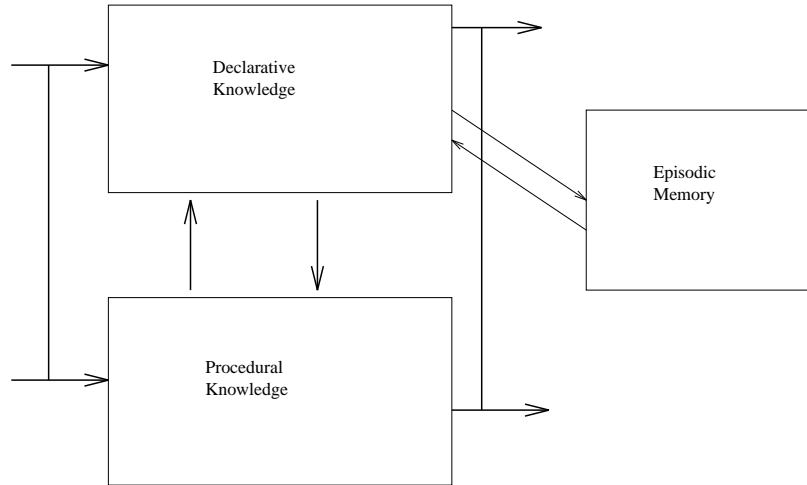


Figure 2: The CLARION Architecture

1988, Schneider and Oliver 1991, Sun 1995, 1997, Schacter 1987, Shallice 1972). (In addition, there is also a separate memory system as described in Sun 1997.) The two levels interact by cooperating in actions, through a combination of the action recommendations from the two levels respectively, as well as cooperating in learning through a bottom-up and a top-down process (to be discussed below). See Figure 2.

We intended for the cognitive model to satisfy some basic requirements as follows. It must be able to learn from scratch (Berry and Broadbent 1988, Reber 1989, and Sun et al 1998a). The model also has to learn continuously from on-going experience in the world (as indicated by Medin et al. (1987), Nosofsky et al (1994) and others, human learning is often gradual, on-going, and concurrent). As suggested by Anderson (1983) and many others, there are clearly different types of knowledge involved in human learning (e.g., procedural vs. declarative, implicit vs. explicit, or subconceptual vs. conceptual): While one is generic and easily accessible, the other is embodied and specific. Moreover, different types of learning processes are involved in acquiring different types of knowledge (Anderson 1983, Keil 1989, Smolensky 1988, Stanley et al 1989). Humans are able to learn subconceptual (implicit) knowledge through trial and error (without a priori knowledge). On top of that, conceptual (explicit) knowledge can be acquired also through on-going experience in the world (see Stanley et al 1989). Furthermore, it is beneficial for conceptual knowledge to be learned through the mediation of subconceptual knowledge (i.e., bottom-up learning; see Sun 1997, Keil 1989, Stanley et al 1989, Willingham et al 1989, and Dreyfus and Dreyfus 1987).

Based on the above considerations, we developed CLARION. An overall pseudo-code algorithm that describes the operation of CLARION is as follows:

1. Observe the current state  $x$ .
2. Compute in the bottom level the Q-values of  $x$  associated with each of all the possible actions  $a_i$ 's:  $Q(x, a_1), Q(x, a_2), \dots, Q(x, a_n)$ . Select one action or a few based on Q-values.
3. Find out all the possible actions ( $b_1, b_2, \dots, b_m$ ) at the top level, based on the input  $x$  (sent up from the bottom level) and the rules in place.
4. Compare the values of the selected  $a_i$ 's with those of  $b_j$ 's (sent down from the top level), and choose an appropriate action  $b$ .
5. Perform the action  $b$ , and observe the next state  $y$  and (possibly) the reinforcement  $r$ .
6. Update Q-values at the bottom level in accordance with the *Q-Learning-Backpropagation* algo-

rithm

7. Update the rule network at the top level using the *Rule-Extraction-Refinement* algorithm.
8. Go back to Step 1.

In the bottom level, a Q-value is an evaluation of the “quality” of an action in a given state:  $Q(x, a)$  indicates how desirable action  $a$  is in state  $x$  (which consists of some sensory input). We can choose an action in any state based on Q-values. To acquire the Q-values, we use the *Q-learning* algorithm (Watkins 1989), a reinforcement learning algorithm. It basically compares the values of successive actions and adjusts an evaluation function on that basis (without explicitly involving and representing goals, states, and outcomes). For details of Q-learning, see Watkins (1989), as well as more recent developments such as Kaelbling et al (1996).

We use a four-layered network for implementation in which the first three layers form a backpropagation network for computing Q-values and the fourth layer (with only one node) performs stochastic decision making. The combination of Q-learning and backpropagation facilitates the development of subconceptual knowledge in the bottom level, which can potentially be done solely on the basis of acting in the world.

This level is modular (that is, a number of small networks can co-exist each of which is adapted to specific modalities, tasks, or groups of input stimuli). This coincides with the modularity claim (Fodor 1983, Karmiloff-Smith 1986, Cosmides and Tooby 1994, Hirschfield and Gelman 1994) that much processing is done by limited, encapsulated (to some extent), specialized processors that are highly efficient. These modules can be developed in interacting with the world (computationally through various decomposition methods; e.g., Jordan and Jacobs 1994, Humphrys 1996, Sun and Peterson 1999). Some of them, however, are formed evolutionarily, that is, given a priori to agents, reflecting their hardwired instincts and propensities (Hirschfield and Gelman 1994).

In the top level, knowledge is captured in a simple rule form. To facilitate correspondence with the bottom level (Clark and Karmiloff-Smith 1993), we use a localist network model for representing these rules. Basically, we connect the nodes representing conditions of a rule to the node representing the conclusion (Sun and Peterson 1998 a). That is, we translate the structure of a set of rules into the structure of a network. See e.g. Sun (1995, 1997) for details of localist encoding.

#### 4.1.1 Autonomous Generation of Symbolic Structures

First, incorporating symbolic rule learning at the top level can be very useful in enhancing reinforcement learning at the bottom level. CLARION extracts symbolic knowledge to supplement neural networks. The basic process is as follows: if an action decided by the bottom level is successful, then the agent extracts a rule that corresponds to the action selected by the bottom level and adds the rule to the top level. Then, in subsequent interaction with the world, the agent verifies the extracted rule by considering the outcome of applying the rule: if the outcome is not successful, then the rule should be made more specific and exclusive of the current case; if the outcome is successful, the agent may try to generalize the rule to make it more universal (Michalski 1983, Mitchell 1982). After rules have been learned, a variety of explicit reasoning methods may be used. (The detail of the algorithm can be found in Sun and Peterson 1998a, b.)

In addition, we can improve the usefulness of results from reinforcement learning by extracting explicit plans at the top level that can be used in an open-loop fashion from closed-loop policies resulting from reinforcement learning at the bottom level. We devise a two-stage bottom-up process, in

which first reinforcement learning is applied to acquire a closed-loop policy and then explicit plans are extracted (Sun and Sessions 1998). The extraction of symbolic knowledge improves the applicability of learned policies when environmental feedback is unavailable or unreliable.

Third, to enhance reinforcement learning and to improve learning results, we form a coalition of multiple reinforcement learning modules. We partition a state space to form regions with each assigned to a different module to handle. Incorporating symbolic methods can be very useful here. We partition regions on-line with, but separately from, reinforcement learning using symbolic descriptions to facilitate the overall process. Partitioning regions symbolically reduces learning complexity, as well as simplifies individual modules (and their function approximators), thus facilitating overall learning.

Finally, yet another process involves learning to segment action sequences to form subroutines to be handled by different modules. It segments sequences to create hierarchical structures of subroutines (temporal modules) to reduce (non-Markovian) temporal dependencies in order to facilitate the learning of tasks.

See Sun (1997) and Sun and Peterson (1998 a, b, 1999) for the details of these processes. Briefly, rule extraction is done on line based on an information gain measure that generates useful rules (different from Towell and Shavlik 1993). Plan extraction is done based on beam search using value functions as a guide. Spatial partitioning is done in a modular reinforcement learning setting based on a measure of error consistency. Temporal segmentation is done using a modular reinforcement learning setup that in effect compares different ways of segmentation (through competition) and selects the best one that maximizes the expected reinforcement.

#### **4.1.2 Assimilation of Externally Provided Symbolic Structures**

Although CLARION can avoid relying on a priori or external knowledge, it can make use of it when such knowledge is available. Knowledge from other agents can guide the exploration by the agent and speed up learning (Towell and Shavlik 1993). To deal with instructed learning, externally provided knowledge, in the forms of symbolic structures such as rules, plans, subroutines, categories (object groupings), and so on, should be assimilated into reinforcement learning and be combined with autonomously generated symbolic structures (i.e., internalization). There are the following possibilities.

When individual rules are given by other agents, these rules can be used at the top level to assist in action decision making as described before. In addition to that, supervised learning based on rules can be performed (with, e.g., backpropagation) for gradual assimilation of the rules into reinforcement learners at the bottom level (Maclin and Shavlik 1994). This process can be done before autonomous learning continues, or it can be done on-line along with reinforcement learning.

However, external instructions tend to be generic. Although helpful, generic knowledge alone is insufficient in achieving efficient performance of a task. The agent needs to develop more specific skills that can be more efficiently applied. This is a complex problem in general. It may require operationalization, that is, turning instruction into a form compatible with the existing knowledge of an agent and directly usable. Operationalization may involve terminological mapping (mapping into the terminology used by the agent), qualitative-to-quantitative mapping, and factual knowledge of task domains. It may also involve the recoding of the knowledge into an internal representation. More flexible instructions may even require a substantial general reasoning capability and a capability for interaction with other agents to progressively refine instructions.

Information concerning categories of objects may be provided by other agents. Such information may be provided through a series of exemplars. The information can be assimilated through the aforementioned spatial partitioning method, which divides up the state space for reinforcement learning into regions that correspond to categories. When information about a category is accepted by an agent, a separate spatial module may be set up for handling that category specifically, and the module is trained with the exemplars provided. Then (multi-module) reinforcement learning can continue as before.

Information concerning (sub)routines may be taught by other agents. When such routines are detailed enough, the afore-mentioned temporal segmentation method may be used to assimilate the information. When the information is provided, a sequential segment module can be set up specifically for the given routine and trained with the symbolic information associated with the routines. Multi-module reinforcement learning can then continue as before.

When a skeletal plan (a sketch of a routine) is given by other agents, assimilation can be done the same way as rule assimilation. That is, the plan can be used at the top level to direct actions whenever they are applicable, and through the process, the reinforcement learner at the bottom level is trained (i.e., biased) by the externally provided plan.

### 4.1.3 Sociocultural vs. Self-Generated Concepts

The assimilation approach above has been called *advice taking* or *learning by being told* in AI (Michalski 1983). The other approach, the *autonomous learning* approach, can also be referred to as *learning by exploration*. While many models explore these two types of learning individually (Gelfand et al 1989, Maclin and Shavlik 1994, Anderson 1983, Sun and Peterson 1998 a, b), we explore the integration of advice taking and autonomous learning with top-down and bottom-up processes. Through integration, together, we show how an agent's conceptual system emerges through interacting with the world.

The issue of sociocultural concepts vs. self-generated concepts has largely escaped the attention of L.V. Vygotsky and Vygotskian psychologists. In the main, Vygotsky (e.g., 1962) and Vygotskian psychologists viewed concepts as completely socially/culturally generated. This is also the position taken by many sociologists, especially collectivists such as Durkheim. But CLARION reveals another side of the story.

In CLARION, although concepts (as represented in the top level) can be acquired from external sources, they can also be acquired internally through extraction from the bottom level (explication of implicit knowledge). In other words, there is clearly a self-generated component in cognition. Of course, ultimately, self-generated concepts are based mostly on the external world (both the physical and sociocultural world), since they are developed through the very process of interacting with the world, and affected by assimilated sociocultural concepts. But they are not completely determined by the external world, in the sense that there are many alternative ways of describing the external world and thus by self-generating a particular set of concepts, an agent puts its own stamp on things and develops its own idiosyncrasy in its dealing with the world. This is advantageous, because self-generation contributes to the formation and continuous revision of a rich, diverse, and useful set of concepts and beliefs that are shared by a society. CLARION illustrates how this is possible and how the two sources — external and internal — interact to accomplish individual cognition.

## 4.2 Accounting for Sociocultural Cognition

How may this model capture sociocultural cognition we described in the previous section? First of all, the ideas reviewed in section 2 concerning social processes of both comportment (or subconceptual processing) and conceptual processing correspond well with the analogous division in the model (i.e., the two levels in CLARION). An agent in normal circumstances is almost wholly absorbed into a social role (as described by Bourdieu), “instinctly” following the rule of the game, through bottom-level implicit, subconceptual processes. In addition, explicit concepts and explicit conceptual processes, for dealing explicitly with social roles and structures, are also formed in the top level, in part based on externally given sociocultural concepts. This match should not come as a surprise, given all the preceding discussion on the learning of everyday routines and the extraction of conceptual representations, since the social/cultural environment is part of the everyday world that an agent deals with and thus the same cognitive processes shall be involved.

Based on the earlier exposition, let us consider social processes in acquiring concepts. Vygotsky’s notion of internalization is central to such processes. Internalization can be accomplished in CLARION through the top-down assimilation process described earlier, which matches well the phenomenological characterization of internalization. The direct acceptance of external rules, plans, and concepts in the top level of the model captures the initial stage of internalization. The assimilation into the bottom level as described earlier, however, captures a deeper process by which external symbolic structures are meshed with the existing routines, reflexes, and behavioral propensities so as to effectively affects agents’ comportment with the world. The implicit, subconceptual learning at the bottom level also captures the internalization of implicitly represented cultural aspects through interacting with them.

Internalized sociocultural signs/symbols enable agents to develop richer representations, utilizing signs/symbols that are formed through sociocultural processes (Wertsch 1991). The internalized external signs/symbols are not innocuous: They carry with them particular social/cultural perspectives and biases. Through internalization, in CLARION, the agent’s behaviors (and the agent’s thinking that directs its behaviors) are mediated by externally given signs/symbols, as well as their associated perspectives and biases (as in phenomenological analyses).

We can distinguish several senses of mediation (Wertsch 1991). The first one is mediation as think-in-terms-of. That is, with externally given concepts/categories, an agent accepts a particular division of the world into categories and thinks in terms of these categories, as opposed many other possible divisions. Another sense is mediation as perspective-taking: Along with accepting a particular division, the agent is forced to accept implicitly the perspective and the world view that are associated with it. The third sense is mediation as biases (whereby the notion of “bias” is the one used in the machine learning literature; Shavlik and Dietterich 1991). It basically means that a given representation may lead the agent to learn certain concepts more readily than others. Although these concepts may not be explicitly given, the agent is *biased* toward learning them (as opposed to their alternatives). All of these senses are applicable in describing the impact of the sociocultural concepts/symbols/signs on cognition in CLARION.

## 4.3 Representing Self and Others

We will discuss two additional aspects that are incorporated into the model to better address sociocultural processes: That is, how an agent represents the notion of *self* and how it represents the notion(s) of *others* (Ashmore and Jussim 1997, Turner et al 1987). The reason why an agent needs to represent

self and others is basically the same as that for representing individual physical objects: It is often advantageous to be able to distinguish different objects, so as to act or react appropriately (which was often neglected in early work on autonomous agents; e.g., Agre 1988). Moreover, it may also need to represent itself, so as to make appropriate inferences and decisions when “self” needs to be figured into decision making expressly.

### 4.3.1 Developing Self Representation

Self understanding is achieved primarily through comportment, direct and unmediated interaction (Heidegger 1927a), with the world and all of its aspects (see also Dreyfus 1992). This is because the primordial way is essential to the understanding of the world and thus also self. This is in sharp contrast to a common approach in AI knowledge representation in which self and its desires, intentions, and beliefs are explicitly represented in some form of logic with which deliberate inference is conducted and is followed by always purposive actions in relation to self-referential goals (Davis 1990, Shoham 1993).

However, although self understanding through transparent everyday activities (comportment) is essential, it can result in misinterpretation, as pointed out by Heidegger (1927a), because the hidden and the distorted cannot be clarified due to a deep “involvement” of the agent with the world (i.e., the *facticity* of the agent’s existence in the world). One way in which this tendency can be counteracted is to establish an *explicit* self representation that can be explicitly manipulated, rectified, introspected, and adjusted, in accordance with proper criteria and explicit reasoning.

Combining this explicit representation of self with the implicit (primordial) representation, we again have a two-level situation, in much the same way as other representations in CLARION. Self representation is thereby accommodated in the CLARION framework.

### 4.3.2 Developing Representation of Others

Similar to representing *self*, the representing of *others* can also be accomplished through a combination of the two types of representations.

First and foremost, *others* are accessible to an agent through their shared world and shared activities in the world, in the same way as *self* is understood in a primordial (implicit and subconceptual) way. Thus, at the bottom level of CLARION, routines for interacting with individuals or types of individuals are developed. At the top level, an explicit representation may also be established for an individual or a type of individual, that is linked to relevant concepts regarding the characteristics of the individual(s) represented. According to Heidegger, such knowledge of others are disclosed through “concernful solicitude”; in CLARION, for example, top-level representations of others can be accomplished through either extraction from the bottom level or being given externally.

Representing agents (self or others) is more complex than representing physical objects, because, associated with an agent representation, there is not only information about the appearance of the agent, but also information about the “inside” of the agent: its state of mind, its behavioral propensity, its particular way of thinking (a “theory” of its mind), and so on. Each of these items can be handled by a combination of top-level explicit representations and bottom-level implicit representations. To capture such complexities, beside generic representations, specialized modules can be developed in CLARION (in the bottom level), each of which is related specifically and exclusively to one individual

(or one type of individual), dealing with its state of mind, behavior propensity, and idiosyncrasy. Note that those and only those individuals (or types of individuals) that are highly relevant to the existential experience of the agent are given a specialized representation in a specialized module.

## 5 Further Work — Modeling Sociocultural Processes

As future work, we may speculate on how computational cognitive models of agents, incorporating both autonomous learning and assimilation, can in turn help us to better understand sociocultural processes themselves, above and beyond the current state of the art in MAS.

Computational studies have been undertaken to understand social structures. Beside statistical models used in sociological research (which, however, is of minimum relevance to a computational understanding), research on multi-agent systems has dealt with simplified versions of the problems of social structures and social institutions. The formation of social structures is, in a way, similar to coalition formation in multi-agent systems (Ketchpel 1996, Kahan and Rapoport 1984, Sandholm and Lesser 1997), whereby continuous interaction through which each agent trying to maximize its own “gain” (defined in some way) leads to the formation of protocols for interaction or coalitions that maximize the expected gain of each agent. Artificial Life (Levy 1992) is also relevant, in which the evolution of interaction of simple organisms and the establishment or emergence of structures and regularities therein (through adaptive processes) are considered. However, these areas assume only very simple agents, which are not cognitively realistic (Kahan and Rapoport 1984, North 1998). After all, social interaction is the result of individual cognition (which include instincts, reflexes, and routines, as well as high-level conceptual processes). Therefore, details of individual cognition cannot be ignored in studying sociocultural processes. At least, the implications of such details should be understood before they are abstracted away. A detailed agent model that are sufficiently complex to capture essential features of individual cognitive processes, especially the combination of autonomous learning and assimilation, should be adopted. This is in contrast to simpler simulations that are constructed based on a set of variables that are abstract and do not translate well into mechanisms of individual cognition. This is the key difference between the kind of computational experiment we suggest here and the kind of computational simulation conducted in ethology, sociobiology, and quantitative sociology. In this regard, CLARION has a distinct advantage.

With a more realistic cognitive model, we should investigate how individual agents interact to give rise to some emergent properties, and from the emergent properties, we should investigate how individual agents fare in the emerged structures. But, moreover, we should investigate how the temporal course of such emerged structures and processes takes shape, and what its effects are on individual agents over its entire course.<sup>6</sup> Due to the complexity of these issues, computational modeling based on detailed cognitive models has a large role to play in this endeavor. Many important issues can be studied on the basis of the CLARION model.

We may extend the CLARION framework to deal with additional aspects of sociocultural processes. Many existing theories in relevant disciplines, ranging from ethology and sociobiology to sociology, can suggest important directions to be explored in computational modeling.

---

<sup>6</sup>In other words, like Bourdieu (1984), we should investigate the bidirectional interaction of “subjective” (i.e., individual) and “objective” (i.e., collective) structures, but we should not place one above the other methodologically (for instance, placing the “objective” over the “subjective” as Bourdieu did; or in the other way as Schutz 1967 did).

There is also the issue of extragenetic vs. genetic factors (Wilson 1975, Fetzer 1985). Society and culture, in general, are extragenetic. But deeper down, it has come to light that society and culture are also shaped by genetic factors formed in the evolutionary process (Cosmides and Tooby 1989), since they rely on cognitive agents, in a collective way, in its formation, transmission, and maintenance. There is, therefore, complex interaction between extragenetic factors on the one side and genetic and ontogenetic factors on the other side (Caporael 1995): Society and culture can affect ontogenetic development of individuals and thus affect the evolutionary process (through natural selection), and conversely, genetic and ontogenetic processes of individual agents, collectively en masse, affect society and culture in various ways.

## 5.1 Simulating Social Processes

We need to have a *set* of CLARION agents. These agents interact not only with the physical environment but also with other agents in order to survive (or to maximize some measure of its own gain). Given appropriate spatial proximity among agents (so that there can be interaction) and survival pressure (such as physical resource limitations), proper forms of cooperation will develop in an either biological way (through evolving specialized “instincts” in a phylogenetic way; Wilson 1975, Fetzer 1985, Cosmides and Tooby 1989) or a cognitive way (through the evolvement of a generic cognitive apparatus and through its development in individual ontogenetic courses). Forms of cooperation can develop into complex social structures that ensure effectiveness of cooperation, in the same way as an interacting dynamic system of many elements that settles into and out of attractors in its state space. The agents in this system will develop not only behavioral routines and conceptual representations appropriate to the physical environment but also those appropriate to the social environment that consists of all the agents.

New CLARION agents can also be assimilated into an established set (with its established social structures). This process can be studied to understand the implication of sociocultural systems on individuals, as well as, to a lesser degree, how individual beliefs affect sociocultural systems and even social structures. That is, we can study how the “power asymmetry” works.

Below, let us look into a few specific issues.

## 5.2 Individual Beliefs and Sociocultural Beliefs

The question is how individual belief systems, which are formed as a mixture of sociocultural concepts and individual self-generated concepts, and the social/cultural belief system, which is the “average” (not the “sum total”) of individual beliefs, interact. According to previous discussions, the influence is mutual: Individual beliefs affect sociocultural beliefs (which is constituted from individual beliefs somehow); sociocultural beliefs also affect the formation, sustenance, and revision of individual beliefs of an agent. It is generally accepted that the influence overwhelmingly goes from established sociocultural beliefs to individual beliefs. We can study this computationally, based on the detailed cognitive model of agents — CLARION:

What are the fundamental ways an individual can affect systems of sociocultural beliefs?  
How do sociocultural beliefs change, computationally? As an entire system or by parts?  
How much “effort” is needed to make a change to a sociocultural belief, computationally? How much “power” does a sociocultural belief exert on individuals, computationally?

When there is an individual belief that conflicts with sociocultural beliefs, how is the conflict resolved?

What is the relationship between explicit, conceptual representations and implicit, sub-conceptual representations, individually as well as collectively? What are their respective roles (Sun 1997)?

### 5.3 Inter-Agent Interaction

We can also investigate, computationally, inter-agent interaction as has been described by, for example, Vygotsky (1962). With a set of CLARION agents, we can study many facets of this process (Thomas and Malone 1979). For example,

How does the collective behavior of a set of agents result from interaction computationally (Mataric 1993)? How does interaction affect individual agents' behaviors? How does inter-agent communication develop in the process?

How do concepts (e.g., Agre 1988) emerge, computationally, in inter-agent interaction? How much does that process depend on particular aspects of an agent model (e.g., CLARION)? Is the process best characterized by the internal properties within agents or by the global emergent properties?

How much is the outcome affected by the social structure of the group (cf. Bourdieu and Wacquant 1992)? (See also the next subsection.)

### 5.4 Forming Social Structures

We can investigate computationally how social structures are formed, enforced, and modified (either gradually or radically) through the interaction of individual agents, based on detailed models of agents. Although there are simplified models in sociobiology and economics (e.g., game theoretical models), there is not much done in understanding this aspect through detailed cognitive models. CLARION can help with the investigation of the following issues:

How can social structures be formed, dynamically in a distributed manner, without external intervention or central control (Hammond et al 1995, Osborne and Rubinstein 1994)? What are the different types of social structures that can be formed, in relation to various observed forms of human and animal societies (Wilson 1975)? What are the conditions for the formation of each type of social structure (Wilson 1975)? What are the conditions for the maintenance of each type of social structure? How do factors such as population size, population density, means of production, technological sophistication, etc. affect the form of social structures (i.e., the scaling effects; Wilson 1975)?

How critical does a social structure (its formation, maintenance, and revision) depend on some particular mechanisms of cognition in individual agents (for example, a particular learning mechanism such as assimilation)? Do some particular aspects of social structures depend critically on evolutionarily developed specialized instincts (as captured in CLARION in the specialized modules in the bottom level) or on generic cognitive abilities of agents (Cosmedis and Tooby 1989)? (That is, the genetic vs. extragenetic issue; Wilson 1975.)

## 5.5 Social Structures in Cognition

Social structures are forced on individuals. Each social institution reinforces its views on agents through explicit and implicit means. CLARION provides means for studying such an effect through controlled computational experiments:

How much and in what way do agents differ, from individual to individual, in their assimilation into social roles? How much social control is needed in order to get an individual agent into a social role? How much social control is needed in order to keep an individual agent in its role for an extended period? What is a change of social role like? How much self initiative is needed for accomplishing a change? How much external control is necessary to force a change? What is the interaction of the two forces? How does an assigned social role affect an agent's behavior (especially its routine activities, or habitus; Bourdieu and Wacquant 1992)? How does it affect an agent's conceptual representations (Durkheim 1895)?

In sum, CLARION provides a fertile ground for studying an extremely wide range of issues in MAS and social cognition, not only those issues that have been examined before by social scientists, which CLARION can further clarify, but also those issues that have never come up before because of lacking detailed computational models of individual agents, which CLARION can help to remedy.

## 6 Concluding Remarks

In this paper, we have argued for the integration of studies of multi-agent interaction and single-agent cognitive modeling. This is because such an integration helps us, on the one hand, to better understand social interaction among agents and thus contribute to multi-agent systems work and, on the other hand, to better model individual agents by taking account of sociocultural aspects of their cognition.

In this article, possible ways of integration have been suggested, which notably take into account important past thinking on social issues (e.g., Weber 1957, Durkheim 1962, Bourdieu 1984). We highlighted the following points: (1) Although existing cognitive models are mostly developed to study individual cognition in isolation, they can be enhanced to handle sociocultural aspects of cognition as well (through incorporating both autonomous learning and assimilation). (2) Such cognitive models can be integrated with models of multi-agent interaction (either small group processes or more complex social interaction). (3) The integration of the two strands can enable us to investigate various forms of sociocultural processes under realistic assumptions about individual cognitive processes. (4) The integration can also enable us to investigate the contributions of various cognitive capacities on sociocultural processes in multi-agent interaction. (5) The integration can enable us to investigate the influence of sociocultural processes on individuals.

ACKNOWLEDGEMENT: Thanks to Mark Bickhard for his comments on an early draft. Thanks also to the two anonymous reviewers for their detailed and thoughtful comments.

## References

- [] P. Agre, (1988). The Dynamic Structure of Everyday Life, Technical Report, MIT AI Lab. Cambridge, MA.
- [] J. R. Anderson, (1983). *The Architecture of Cognition*, Harvard University Press, Cambridge, MA
- [] J. Anderson and C. Lebiere, (1998). *The Atomic Components of Thought*, Lawrence Erlbaum Associates, Mahwah, NJ.
- [] R. Ashmore and L. Jussim, (eds.) (1997). *Self and Identity: Fundamental Issues*. Oxford University Press, New York.
- [] J. Barkow, L. Cosmides and J. Tooby, (1992). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press. New York.
- [] J. Barresi and C. Moore, (1995). Intentional relations and social understanding. *Brain and Behavioral Sciences*. Vol.?
- [] W. Bechtel and G. Graham, (eds.) (1998). *A Companion to Cognitive Science*. Blackwell, Oxford, UK.
- [] D. Berry and D. Broadbent, (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology*. 79, 251-272.
- [] M. Bickhard, (1998). Interaction and Representation. <http://www.lehigh.edu/~bickhard>
- [] P. Bourdieu, (1984). *Distinction: A Social Critique of the Judgement of Taste*. Harvard University Press, Cambridge, MA.
- [] P. Bourdieu and L. Wacquant, (1992). *An Invitation to Reflexive Sociology*. University of Chicago Press. Chicago.
- [] R. Brooks, (1991). Intelligence without representation. *Artificial Intelligence*. Vol.47. 139-159.
- [] J. Bruner, (1995). *Acts of Meaning*. Harvard University Press, Cambridge, MA.
- [] L. Caporael, (1995). Sociality: coordinating bodies, minds and groups. *Psychology*, February, 1995.
- [] A. Clark and A. Karmiloff-Smith, (1993). The cognizer's innards: a psychological and philosophical perspective on the development of thought. *Mind and Language*. 8 (4), 487-519.
- [] L. Cosmides and J. Tooby, (1989). Evolutionary psychology and the generation of culture II. *Ethology and Sociobiology*. 10, 51-97.
- [] L. Cosmides and J. Tooby, (1994). Beyond intuition and instinct blindness: toward an evolutionarily rigorous cognitive science. *Cognition*. 50, 41-77.
- [] R. D'Andrade, (1989). Cultural cognition. In: M. Posner, (ed.) *Foundations of Cognitive Science*. MIT Press, Cambridge, MA.
- [] E. Davis, (1990). *Representations of Commonsense Knowledge*. Morgan Kaufman, San Mateo, CA.
- [] J. Deneubourg and S. Goss, (1989). Collective patterns and decision making. *Ethology, Ecology, and Evolution*. 1: 295-311.
- [] M. d'Inverno, M. Luck, and M. Wooldridge, (1997). Cooperation structures. *Proc. of IJCAI'97*.
- [] H. Dreyfus, (1972). *What Computers Can't Do*. Harper and Row. New York.
- [] H. Dreyfus and S. Dreyfus, (1987). *Mind Over Machine: The Power of Human Intuition*. The Free Press, New York, NY.

- [] H. Dreyfus, (1992). *Being-in-the-world*. MIT Press, Cambridge, MA.
- [] W. Durkheim, (1895/1962). *The Rules of the Sociological Method*. Free Press, Glencoe, IL.
- [] T. Eggertsson, (1990). *Economic Behavior and Institutions*. Cambridge University Press, Cambridge, UK.
- [] J. Fetzer, (1985). *Sociobiology and Epistemology*. Reidel Publishing, Dordrecht, Netherlands.
- [] J. Fodor, (1975). *The Language of Thought*. Crowell.
- [] J. Fodor, (1980). Methodological Solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences*. Vol.3, 417-424.
- [] J. Fodor, (1983). *The Modularity of Mind*. MIT Press, Cambridge, MA.
- [] S. Freud, (1937). *A General Selection from the Works of Sigmund Freud*. Hogarth Press. London.
- [] J. Gelfand, D. Handelman and S. Lane, (1989). Integrating knowledge-based systems and neural networks for robotic skill acquisition, *Proc.IJCAI*, pp.193-198. Morgan Kaufmann, San Mateo, CA.
- [] C. Geertz, (1973). *The Interpretation of Culture*. Basic Books, New York.
- [] K. Hammond, T. Converse, and J. Grass, (1995). The stabilization of environments. *Artificial Intelligence*. Vol.72, 305-328.
- [] M. Heidegger, (1927a). *Being and Time*. English translation published by Harper and Row, New York. 1962.
- [] M. Heidegger, (1927b). *The Basic Problem of Phenomenology*.
- [] L. Hirschfield and S. Gelman, (eds.) (1994). *Mapping the Mind: Domain Specificity in Cognition and Culture*. Cambridge University Press, Cambridge, UK.
- [] J. Holland, N. Nisbitt, T. Thagard and J. Holyoak, *Induction: A Theory of Learning and Development*, MIT Press, 1986
- [] M. Humphrys, (1996). W-learning: a simple RL-based society of mind. Technical report 362, University of Cambridge, Computer Laboratory.
- [] E. Hutchins, (1995). How a cockpit remembers its speeds. *Cognitive Science*. 19, 265-288.
- [] W. James, (1890). *The Principles of Psychology*. Dover, New York.
- [] M. Jordan and R. Jacobs, (1994). Hierarchical mixtures of experts and the EM algorithm. *Neural Computation*. 6, 181-214.
- [] C. G. Jung, (1959). *The Archetypes and the Collective Unconscious*. Pantheon Books, New York.
- [] L. Kaelbling, M. Littman, and A. Moore, (1996). Reinforcement learning: a survey. *Journal of Artificial Intelligence Research*, 4, 237-285.
- [] J. Kahan and A. Rapoport, (1984). *Theories of Coalition Formation*. Erlbaum, Mahwah, NJ.
- [] S. Ketchpel, (1996). Forming coalitions in the face of uncertain rewards. *Proc. of AAAI*.
- [] A. Karmiloff-Smith, (1986). From meta-processes to conscious access: evidence from children's metalinguistic and repair data. *Cognition*. 23. 95-147.
- [] A. Karmiloff-Smith, (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. MIT Press, Cambridge, MA.
- [] F. Keil, (1989). *Concepts, Kinds, and Cognitive Development*. MIT Press. Cambridge, MA.
- [] D. Klahr, P. Langley, and R. Neches, (eds.) (1987). *Production System Models of Learning and Development*. MIT Press, Cambridge, MA.

- [] T. Kuhn, (1962). *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago.
- [] J. Lave, (1988). *Cognition in Practice*. Cambridge University Press.
- [] S. Levy, (1992). *Artificial Life*. Jonathan Cape, London.
- [] R. Maclin and J. Shavlik, (1994). Incorporating advice into agents that learn from reinforcements. *Proc.of AAAI-94*. Morgan Kaufmann, San Mateo, CA.
- [] M. Mataric, (1993). Kin recognition, similarity, and group behavior. *Proc.of Cognitive Science Society Annual Conference*. 705-710.
- [] D. Medin, W. Wattenmaker, and R. Michalski, (1987). Constraints and preferences in inductive learning: an experimental study of human and machine performance. *Cognitive Science*. 11, 299-339.
- [] R. Michalski, (1983). A theory and methodology of inductive learning. *Artificial Intelligence*. Vol.20, 111-161.
- [] T. Mitchell, (1982). Generalization as search. *Artificial Intelligence*, 18, 203-226.
- [] D. North, (1998). Institutions and economics. in: *A Companion to Cognitive Science*, W. Bechtel and G. Graham, (eds.) 713-721. Blackwell, Oxford, UK.
- [] M. Osborne and A. Rubinstein, (1994). *A Course on Game Theory*. MIT Press, Cambridge, MA.
- [] D. Osherson and H. Lasnik, (1990). *An Invitation to Cognitive Science*. MIT Press, Cambridge, MA.
- [] S. Pinker, (1994). *The Language Instinct*. W. Morrow and Co., New York.
- [] M. Posner, (ed.) (1989). *Foundations of Cognitive Science*, MIT Press, Cambridge, MA.
- [] H. Putnam. 1975. The meaning of 'meaning'. in K. Gunderson (ed.) *Mind and Knowledge*. Mpls, MN: University of Minnesota Press.
- [] A. Reber, (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*. 118 (3), 219-235.
- [] S. Ross, (1973). The economic theory of agency. *American Economics Review*. Vol.63, 134-139.
- [] D. Rumelhart and J. McClelland, eds. (1986). *Parallel Distributed Processing I*. MIT Press, Cambridge, MA.
- [] R. Salustowicz, M. Wiering, and J. Schmidhuber, (1998). Learning team strategies: soccer case studies. *Machine Learning*.
- [] T. Sandholm and V. Lesser, (1997). Coalition among computationally bounded agents. *Artificial Intelligence*.
- [] W. Schneider and W. Oliver, (1991), An instructable connectionist/control architecture. In: K. VanLehn (ed.), *Architectures for Intelligence*, Erlbaum, Hillsdale, NJ.
- [] A. Schutz, (1967). *The Phenomenology of the Social World*. Northwestern University Press.
- [] D. Schacter, (1987). Implicit memory: history and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 13, 501-518.
- [] T. Shallice, (1972). Dual functions of consciousness. *Psychological Review*. Vol.79, No.5, 383-393.
- [] J. Shavlik and T. Dietterich, (1991). *Readings in Machine Learning*. Morgan Kaufmann Publishers.
- [] Y. Shoham, (1993). Agent-oriented programming. *Artificial Intelligence*. Vol.60, 51-92.
- [] E. Smith, C. Langston, and R. Nisbett, (1992). The case for rules in reasoning. *Cognitive Science*, 16, 1-40.

- [] P. Smolensky, (1988). On the proper treatment of connectionism, *Behavioral and Brain Sciences*, 11, 1-43.
- [] W. Stanley, R. Mathews, R. Buss, and S. Kotler-Cope, (1989). Insight without awareness: on the interaction of verbalization, instruction and practice in a simulated process control task. *Quarterly Journal of Experimental Psychology*. 41A (3), 553-577.
- [] T. Sowell, (1996). *Knowledge and Decisions*. Basic Book, New York.
- [] L. Suchman, (1987). *Plans and Situated Actions*. Cambridge University Press, Cambridge, UK.
- [] R. Sun, (1994). *Integrating Rules and Connectionism for Robust Commonsense Reasoning*. Wiley, New York.
- [] R. Sun, (1995). Robust Reasoning: Integrating Rule-Based and Similarity-Based Reasoning. *Artificial Intelligence*, 75, 2. 241-296.
- [] R. Sun, (1997). Learning, action, and consciousness: a hybrid approach towards modeling consciousness. *Neural Networks*, special issue on consciousness. 10 (7), 1317-1331.
- [] R. Sun, (2000). Symbol grounding: a new look at an old issue. *Philosophical Psychology*, Vol.13, No.3, 403-418.
- [] R. Sun and T. Peterson, (1998 a). Some experiments with a hybrid model for learning sequential decision making. *Information Sciences*. Vol.111, 83-107.
- [] R. Sun and T. Peterson, (1998 b). Autonomous learning of sequential tasks: experiments and analyses. *IEEE Transaction on Neural Networks*, Vol.9, No.6, pp.1217-1234.
- [] R. Sun and T. Peterson, (1999). Multi-agent reinforcement learning: weighting and partitioning. *Neural Networks*, Vol.12, No.4-5. pp.127-153.
- [] R. Sun and C. Sessions, (1998). Learning to plan probabilistically from neural networks. *Proceedings of IEEE International Conference on Neural Networks*, pp.1-6. IEEE Press, Piscataway, NJ.
- [] R. Sun, E. Merrill, and T. Peterson, (1998 a). A bottom-up model of skill learning. *Proc. of 20th Cognitive Science Society Conference*, 1037-1042, Lawrence Erlbaum Associates, Mahwah, NJ.
- [] R. Sun, E. Merrill, and T. Peterson, (1998 b). Skill learning using a bottom-up hybrid model. *Proc. of The Second European Conference on Cognitive Modeling*, Nottingham, UK. April, 1998. 23-29. Nottingham University Press.
- [] R. Sun, E. Merrill, and T. Peterson, (1999). From implicit skills to explicit knowledge: a bottom-up model of skill learning. *Cognitive Science*, in press.
- [] E. Thomas and T. Malone, (1979). On the dynamics of two-person interactions. *Psychological Review*. 86 (4), 331-360.
- [] G. Towell and J. Shavlik, (1993). Extracting Refined Rules from Knowledge-Based Neural Networks, *Machine Learning*. 13 (1), 71-101.
- [] J. Turner, M. Hogg, P. Oakes, S. Reicher, and M. Wetherell, (1987). *Rediscovering the Social Group: Self-Categorization Theory* Blackwell, Oxford, UK.
- [] F. Varela, E. Thompson, and E. Rosch, (1993). *The Embodied Mind*. MIT Press, Cambridge, MA.
- [] L. Vygotsky, (1986). *Mind in Society*. Harvard University Press, Cambridge, MA.
- [] L. Vygotsky, (1962). *Thought and Language*. MIT Press, Cambridge, MA.
- [] C. Watkins, (1989). *Learning with Delayed Rewards*. Ph.D Thesis, Cambridge University, Cambridge, UK.

- [] D. Willingham, M. Nissen, and P. Bullemer, (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 15, 1047-1060.
- [] M. Weber, (1957). *The Theory of Social and Economic Organization*. The Free Press, Glencoe, IL.
- [] J. Wertsch, (1991). *Voices of the Mind: A Sociocultural Approach to Mediated Action*. Harvard University Press, Cambridge, MA.
- [] J. Wertsch, (1998). Mediated action. In: W. Bechtel and G. Graham, (eds.) *A Companion to Cognitive Science*, 518-525. Blackwell, Oxford, UK.
- [] E. Wilson, (1975). *Sociobiology*. Harvard University Press, Cambridge, MA.