PART 1

INTRODUCTION

1

Prolegomena to Integrating Cognitive Modeling and Social Simulation

Ron Sun

1 INTRODUCTION

A multi-agent system (i.e., a society of agents) is a community of autonomous entities each of which perceives, decides, and acts on its own, in accordance with its own interest, but may also cooperate with others to achieve common goals and objectives. How to achieve meaningful coordination among agents in general, however, is a difficult issue and, to a very large extent, a mystery thus far (despite the fact that it has been extensively tackled).

Over the years, the notions of agent and agency have occupied a major role in defining research in social and behavioral sciences, including sociology, philosophy, economics, psychology, and many other fields. The notion of agent has also invaded computer science and engineering (in Internet computing and in robotics research in particular). Computational models of agents have been developed in both artificial intelligence and cognitive science. In AI, they appear under the rubric of "intelligent agents." In cognitive science, they are often known as "cognitive architectures," that is, the essential structure and process of a (broadly-scoped) domain-generic computational cognitive model. They are often used for broad, cross-domain analysis of cognition and behavior (Newell, 1990; Sun, 2002). Together, these strands of research provide useful paradigms for addressing some fundamental questions concerning human nature (Newell, 1990; Anderson & Lebiere, 1998; Sun, 2002).

In particular, although traditionally the main focus of research in cognitive science has been on specific components of cognition (e.g., perception, memory, learning, or language), relatively recent developments in computational modeling through cognitive architectures provide new avenues for precisely specifying a range of complex cognitive processes together in tangible ways.[1] Computational cognitive modeling, especially with cognitive architectures, has become an essential area of research on cognition (Anderson & Lebiere, 1998; Sun, 2004). Computational cognitive modeling has been gradually integrated into larger communities of social and behavioral sciences (Schunn & Gray, 2002). A particularly important aspect of this integration is that by now, mainstream experimental and theoretical psychology journals have started publishing computational modeling papers. This fact reflects the growing interest in computational cognitive modeling and cognitive architectures on the part of traditional psychological communities. Likewise, significant applications of computational cognitive models have found their way into some significant application domains (Pew & Mavor, 1998; Ritter et al., 2003). Such developments, however, need to be extended to issues of multi-agent interaction. There have been some promising initial developments in this regard (see, for example, a number of recent papers in this area in the journal *Cognitive Systems Research*).

Against this background, the present volume brings together cognitive scientists, social scientists, as well as AI researchers, with a wide range of background and expertise, to discuss issues in understanding the relation between cognition and social processes, through exploring the relation between computational cognitive modeling and social simulation (Axelrod, 1984; Gilbert & Doran, 1994; Gilbert & Conte, 1995; Epstein & Axtell, 1996; Conte et al., 1997; Moss & Davidsson, 2001; etc.). The questions that are of particular interest in this endeavor include:

- How do we extend computational cognitive modeling to multi-agent interaction (i.e., to social simulation)?
- What should a proper computational cognitive model for addressing multi-agent interaction be like?
- What are essential cognitive features that should be taken into consideration in computational simulation models of multi-agent interaction?
- What additional representations (for example, "motive," "obligation," or "norm") are needed in cognitive modeling of multi-agent interaction?
- What are the appropriate characteristics of cognitive architectures for modeling both individual cognitive agents and multi-agent interaction?

[1] A cognitive architecture provides a concrete framework for more detailed modeling of cognitive phenomena, through specifying essential structures, divisions of modules, relations among modules, and a variety of other essential aspects (Sun, 1999). It helps to narrow down possibilities, provides scaffolding structures, and embodies fundamental theoretical assumptions. The value of cognitive architectures has been argued many times before; see, for example, Newell (1990), Anderson and Lebiere (1998), Sun (2002), Sun (2004), and so on.

And on the other hand,

- How do we measure cognitive realism of multi-agent (social simulation) models?
- What can cognitive realism contribute to the understanding of social processes?
- How should we understand the relation between individual cognition and collective social phenomena in general?
- What are the fundamental ways of understanding and modeling multi-agent interaction? How much can they be reduced to individual cognition?
- How should we characterize the "collective mind"?
- How important is culture in shaping individual cognition and collective behavior? How can we model the effect of culture on cognition and behavior?
- How can we best characterize and model social relations, structures, and organizations in relation to cognition?
- How important is evolution in shaping individual cognition and collective social phenomena? How can we model that aspect?

So on and so forth. These issues are just a few particularly important ones among many others important issues.

It should be noted that here we use the term "cognition" in the broadest sense, including, but not limited to, thinking, reasoning, planning, problem solving, learning, skills, perception, motor control, as well as motivation and emotion. That is, we use it to denote everything going on in the mind.

It should also be noted that the study of multi-agent interaction (e.g., in AI and in economics) raised some specific issues. These issues include how to develop coordination strategies (that enable groups of agents effectively to solve problems together), negotiation mechanisms, conflict detection and resolution strategies, and other mechanisms whereby agents can contribute to overall system effectiveness whereas still assuming a large degree of autonomy. Relatedly, issues concerning how organizations of agents (including teams) can be formed, structured, and utilized have also been raised. They are very relevant to addressing the questions enumerated earlier.

2 BACKGROUND

Two approaches dominate traditional social sciences. The first approach may be termed the "deductive" approach (Axelrod, 1997; Moss, 1999), exemplified by much research in classical economics. It proceeds with the construction of mathematical models of social phenomena, usually expressed as a set of closed-form mathematical equations. Such models may be simple and elegant. Their predictive power derives from the analysis of various states (equilibria) through applying the equations. Deduction is used to find consequences of assumptions in order to help achieve better understanding of relevant phenomena.

The second approach may be termed the "inductive" approach, exemplified by many traditional approaches to sociology. With such an approach, insights are obtained by generating generalizations from (hopefully a large number of) observations. Insights are usually qualitative in nature and describe social phenomena in terms of general categories and characterizations of these general categories.

However, a new approach has emerged relatively recently. It involves computer simulations of social phenomena.[2] It starts with a set of explicit assumptions. But unlike deduction, it does not prove theorems. Instead, simulations lead to data that can be analyzed inductively to come up with interesting generalizations. However, unlike typical induction in empirical social sciences, simulation data come from pre-specified rules, not from direct measurements of social phenomena. With simulation data, both inductive and deductive methods may be applied: Induction can be used to find patterns in data, and deduction can be used to find consequences of assumptions (that is, rules specified for simulations). Thus, simulations are useful for developing theories, in both directions and in their combinations thereof (Axelrod, 1997; Moss, 1999).

Among this third approach, a particularly interesting development is the focus on *agent-based* social simulations, that is, simulations based on autonomous individual entities, as defined earlier. Naturally, such simulations focus on the interaction among agents. From their interactions, complex patterns may emerge. Thus, the interactions among agents provide explanations for corresponding social phenomena (Gilbert, 1995). Agent-based social simulation has seen tremendous growth in the recent decade. Researchers frustrated with the limitations of traditional approaches to the social sciences have increasingly turned to "agents" for studying a diverse set of theoretical and practical issues.

Despite their stated goals, however, most of the work in social simulation still assumes very rudimentary cognition on the part of agents. Whereas often characterizing agents as "intelligent" actors, there have been relatively few serious attempts to emulate human cognition (Thagard, 1992). Agent models have frequently been custom-tailored to the task at hand, often amounting to little more than a restricted set of highly domain-specific rules. Although this approach may be adequate for achieving some limited objectives of some simulations, it is overall unsatisfactory. It not only limits the realism, and hence applicability of social simulations, but also

[2] It has sometimes been referred to as a "third way" of doing science, as contrary to the two traditional approaches (Axelrod, 1997; Moss, 1999).

precludes any possibility of resolving the theoretical question of the micro–macro link (to be discussed later). At the same time, researchers in cognitive science, although studying individual cognition in depth, have paid relatively little attention to social phenomena (with some notable exceptions of course). The separation of the two fields can be seen (1) in the different journals dedicated to the two different fields (e.g., *Journal of Artificial Society and Social Simulation*, *Emergence*, and *Computational and Mathematical Organization Theory* for social simulations, versus *Cognitive Science*, *Cognitive Systems Research*, and *Cognitive Science Quarterly* for cognitive modeling), (2) in the different conferences for these two different fields (e.g., *the International Conferences on Social Simulation* versus *the International Conference on Cognitive Modeling*), (3) in the different professional organizations (e.g., *the North American Association for Computational Social and Organizational Science* and *the European Social Simulation Association* versus *the Cognitive Science Society*), as well as (4) in the scant overlap of authors in these two fields. Moreover, most of the commonly available social simulation tools (e.g., *Swarm* and *RePast*) embody very simplistic agent models, not even remotely comparable to what has been developed within the field of cognitive architectures (Anderson & Lebiere, 1998; Sun, 2002).

We believe that investigation, modeling, and simulation of social phenomena (whether using multi-agent systems or not) needs cognitive science (Sun, 2001a,b), because we have reasons to believe that such endeavors need a better understanding, and better models, of individual cognition, only on the basis of which it can develop better models of aggregate processes through multi-agent interaction. Cognitive models may provide better grounding for understanding multi-agent interaction, by incorporating realistic constraints, capabilities, and tendencies of individual agents in terms of their cognitive processes (and also in terms of their physical embodiment) in their interaction with their environments (both physical and social environments). This point was argued at length in Sun (2001b). This point has also been made, for example, in the context of cognitive realism of game theory (Kahan & Rapaport, 1984; Camerer, 1997), or in the context of deeper models for addressing human–computer interaction (Gray & Altmann, 2001).

Conversely, cognitive science also needs multi-agent systems, social simulation, and social sciences in general. Cognitive science is in need of new theoretical frameworks and new conceptual tools, especially for analyzing sociocultural aspects of cognition and cognitive processes involved in multi-agent interaction. It needs computational models and theories from multi-agent work in AI, and also broader conceptual frameworks that can be found in sociological and anthropological work (as well as in social psychology to some extent). In particular, computational cognitive modeling, as a field, can be enriched through the integration of these disparate strands of ideas.

This present volume is concerned exactly with such integration of the studies of the social and the cognitive. The underlying goal of what we are collectively doing here is evident: What we are working towards is not just a slightly better social simulation, or a more "believable" multi-agent system. Much beyond these, what we are actually working towards, whether we acknowledge it or not, is *cognitive social science* (or "cognitivized" social science) – a social science that bases its methodology and theory on the in-depth study of the human mind. The study of the human mind is the essential ingredient of any social science and, one may argue, should be the basis of such science (although we clearly realize that there are opposing views on this issue, which may be well entrenched). Going even beyond that, we are actually working towards *computational cognitive social science* – with computational approaches being adopted as the primary means (Prietula et al., 1998; Sun, 2001b).

3 ONE HIERARCHY AND MANY LEVELS

As alluded to before, one striking feature, apparent upon examining the state of the art in social and cognitive sciences, is the lack of integration and communication among disciplines. Each discipline considers a particular aspect and ignores the rest (more or less). Each is substantially divorced from other, related disciplines. Generally, they do not work together. Consequently, they often talk past each other, instead of to each other.

Here, instead, let us take a broader perspective and look at multiple "levels" of analysis in social and cognitive sciences. These levels of analysis can be easily cast as a set of related disciplines, from the most macroscopic to the most microscopic. These different *levels* include: the sociological level, the psychological level, the componential level, and the physiological level (see Table 1.1). In other words, as has been argued in Sun et al. (2004), we may view different disciplines as different levels of abstraction in the process of exploring essentially the same broad set of questions (cf. Newell, 1990).

TABLE 1.1. *A Hierarchy of Four Levels.*

Level	Object of Analysis	Type of Analysis	Model
1	inter-agent/collective processes	social/cultural	collections of agent models
2	agents	psychological	individual agent models
3	intra-agent processes	componential	modular construction of agent models
4	substrates	physiological	biological realization of modules

First of all, there is the sociological level, which includes collective behaviors of agents (Durkheim, 1895), interagent processes (Vygotsky, 1986), sociocultural processes, social structures and organizations, as well as interactions between agents and their (physical and sociocultural) environments.

Although studied extensively by sociology, anthropology, political science, and economics, this level has traditionally been very much ignored in cognitive science. Only recently, cognitive science, as a whole, has come to grips with the fact that cognition is, at least in part, a sociocultural process (Lave, 1988; Hutchins, 1995). To ignore sociocultural processes is to ignore a major underlying determinant of individual cognition. The lack of understanding of sociological processes may result in the lack of understanding of some major structures in, and constraints on, cognition.[3]

The next level is the psychological level, which covers individual experiences, individual behaviors, individual performance, as well as beliefs, concepts, and skills employed by individual agents. In relation to the sociological level, the relationship of individual beliefs, concepts, and skills with those of the society and the culture, and the processes of change of these beliefs, concepts, and skills, independent of or in relation to those of the society and the culture, may be investigated (in inter-related and mutually influential ways). At this level, we may examine human behavioral data, compared with models and with insights from the sociological level and details from the lower levels.

The third level is the componential level. At this level, we study and model cognitive agents in terms of components, with the theoretical language of a particular paradigm, for example, symbolic computation or connectionist networks, or their combinations thereof. At this level, we may specify computationally an overall architecture and the components therein. We may also specify some essential computational processes of each component as well as essential connections among components. Ideas and data from the psychological level, that is, the psychological constraints from above, which bear significantly on the division of components and their possible implementations, are among the most important considerations. This level may also incorporate biological/physiological facts regarding plausible divisions and their implementations; that is, it can incorporate ideas from the next level down – the physiological level, which offers the biological constraints. This level results in *mechanisms*, though they are computational and thus somewhat abstract compared with physiological-level details. The importance of this level has been argued for, for example, in Sun et al. (2004) and Gray and Altmann (2001).

[3] See Sun (2001b) for a more detailed argument for the relevance of sociocultural processes to cognition and vice versa.

Although this level is essentially in terms of intra-agent processes, computational models developed therein may be used to capture processes at higher levels, including interaction at a sociological level whereby multiple individuals are involved. This can be accomplished, for example, by examining interactions of multiple copies of individual agent models or those of different individual agent models. We may use computation as a means for constructing agent models at a sub-agent level (the componential level), but we may go up from there to the psychological level and to the sociological level (see more discussions of mixing levels later on).

The lowest level of analysis is the physiological level, that is, the biological substrate, or the biological implementation, of computation. This level is the focus of a range of disciplines including biology, physiology, computational neuroscience, cognitive neuroscience, and so on. Although biological substrates are not our main concern here, they may nevertheless provide useful input as to what kind of computation is likely employed and what a plausible architecture (at a higher level) should be like (Piaget, 1971). The main utility of this level is to facilitate analysis at higher levels, that is, analysis using low-level information to narrow down choices in selecting computational architectures as well as choices in implementing componential computation.

Work at this level is basically the reverse-engineering of biological systems. In such a case, what we need to do is to pinpoint the most basic primitives that are of relevance to the higher-level functioning that we are interested in. Although many low-level details are highly significant, clearly not all low-level details are significant or even relevant. After identifying proper primitives, we may study processes that involve those primitives, in mechanistic/computational terms.

To more clearly illustrate this view of cascading levels, Figure 1.1 shows the correspondences among levels, with a cascade of maps of various resolutions.

4 CROSSING AND MIXING LEVELS

Although analysis in modeling and simulation is often limited to within a particular level at a time (inter-agent, agent, intra-agent, or substrate), this need not be the case: Cross-level analysis and modeling could be intellectually enlightening, and might even be essential to the progress of science (Sun et al., 2004). These levels proposed earlier do interact with each other (e.g., constraining each other) and may not be easily isolated and tackled alone. Moreover, their respective territories are often intermingled, without clear-cut boundaries.

For example, the cross-level link between the psychological and the neurophysiological level has been strongly emphasized in recent years (in the form of cognitive neuroscience; see, for example, LeDoux, 1992; Damasio,

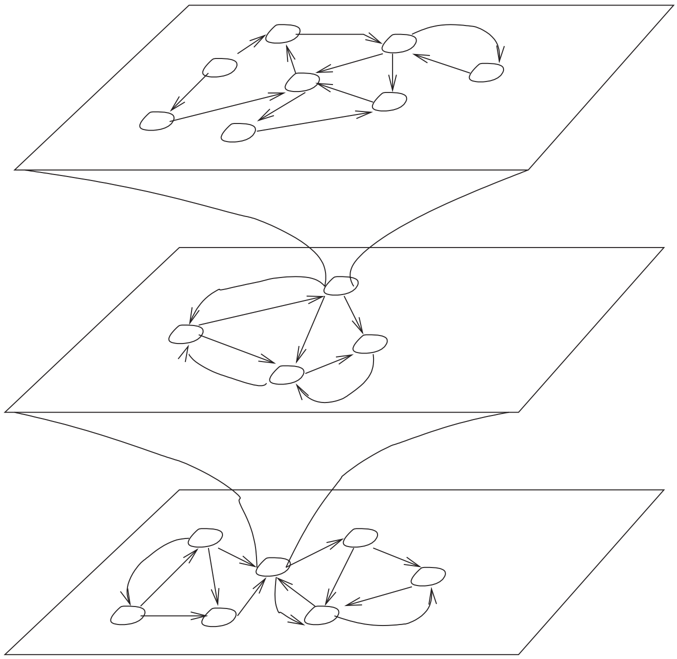

FIGURE 1.1. The cascading levels of analysis.

1994; Milner & Goodale, 1995). For another example, the psychological and the social level may also be crossed (and may even be integrated) in many ways, in order to generate new insights into social phenomena on the basis of cognitive processes (Boyer & Ramble, 2001) and, conversely, to generate insights into cognitive phenomena on the basis of sociocultural processes (Hutchins, 1995; Nisbett et al., 2001). In particular, in the field of cognitive work analysis, in order to facilitate the design of physical work environments and group structures that improve work performance, work activities are analyzed in terms of the cognitive processes involved (such as memory requirement, visual perception, etc.) to shed light on possible areas of improvement. In all of these cases, the ability to shift freely between levels, or to understand the mapping between levels, is a critical part of scientific work.

Note that when crossing levels, there is no fixed path, from either the highest level to the lowest level, or vice versa. Instead, analysis at multiple levels can, and should, be pursued simultaneously and be used to constrain and to guide each other.

Beyond cross-level analysis, there may be "mixed-level" analysis (Sun et al., 2004). The idea of mixed-level analysis may be illustrated by the research at the boundaries of quantum mechanics. In deriving theories,