ELSEVIER

Journal of Cognitive Systems Research 2 (2001) 1–3

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Editorial

Individual action and collective function: From sociology to multi-agent learning

Ron Sun

CECS Department, University of Missouri, Columbia, MO 65211, USA

Co-learning of multiple agents has been studied in many different disciplines under various guises. For example, the issue has been tackled by distributed artificial intelligence, parallel and distributed computing, cognitive psychology, social psychology, game theory (and other areas of mathematical economics), sociology, anthropology, and many other related disciplines.

These studies are often disparate. Different disciplines tend to ignore each other, although there has been cross-disciplinary work, such as AI models and cognitive studies using game theory (e.g., West & Lebiere, 2001), or sociological work incorporating psychological insights.

We believe that interdisciplinary interaction and integration are important, and cross-disciplinary communications can help to make better progress. Therefore, we want to take a close look at research on multi-agent learning, accentuating its interdisciplinary nature.

Many questions concerning multi-agent learning can be asked, in an interdisciplinary way:

- How do agents learn to cooperate with each others, especially under bounded rationality?
- What is the minimum cognitive capacity necessary for an agent to learn to cooperate with others?
- What are the realistic cognitive constraints in co-learning settings, and how do they help, or hamper, learning and cooperation?
- How do we characterize the process and the dynamics of co-learning, conceptually, mathematically, or computationally?
- how do social structures and relations interact with co-learning of multiple agents?

and so on.

A key question, however, is the following. As Adam Smith (1976) put it:

> *He generally, indeed, neither intends to promote the public interest, nor knows how much he is promoting it . . . . He intends only his own gain, and he is led by an invisible hand to promote an end which was not part of his intention.*

This paradox have been troubling sociologists and economists for many decades, and now computer scientists and psychologists as well. The issue may be formulated as the apparent gap between the individual intention in deciding his/her own action and the (possibly largely unintended) social function of his/her action. For example, how may self-interested action benefit social welfare? Or, how may cooperation be established through each individual maximizing his/her own gain (Axelrod, 1984)? As Castelfranchi put it: ''The real problem is modeling how we play our social roles, while being unaware of the functional effects of our actions, not only with our routine actions but even when doing something

E-mail addresses: rsun@cecs.missouri.edu (R. Sun), http://www.cecs.missouri.edu/~rsun (R. Sun).

1389-0417/01/$ – see front matter
PII: S1389-0417(01)00012-2

deliberately for our own subjective motives'' (Castelfranchi, this volume).

Is this situation similar to the ''paradox'' of the firing of individual neurons and the computation of a network of neurons? Each neuron fires at its own ''will'' and apparently for its own ''gain''. But, together, a network of neurons accomplishes complex functions unknown to individual neurons.

There is, clearly, a strong similarity there. However, when human actions are concerned, there is the issue of conscious intention of human actors, as well as explicit beliefs and goals of theirs (Sun, 1999). Such explicit mental representation, which may not have any direct connections with social function, poses a serious theoretical dilemma, as pointed out, for example, by Elster (1982). The problem is how we bridge the gap between explicit individual intentions and unintended social function.

As has been suggested before, the understanding of individual learning and collective evolution may be the key to a satisfactory explanation of this problem. Castelfranchi (this volume) looks into various forms of *emergence*, from simple pattern formation, to cognitive emergence. Among them, cognitive emergence (or implicit-to-explicit bottom-up explication, as termed by Sun, 1999 and Sun et al., 2001) is important. Along with collective evolution, the notion of cognitive emergence may reconcile the afore-mentioned difference between individual intention and collective social function of human action. In a nutshell, the hypothesis is that collective social function may be lodged in the cognitive unconscious of the human mind, through long evolutionary processes in social environments, and through reinforcement agents receive on an individual basis. Such hidden motives, through the cognitive unconscious, may serve as Adam Smith's ''invisible hand'', giving rise to the emergent structures of social function. Then, through cognitive emergence (Sun et al., 2001), they may become consciously known to agents as well, although correct conscious interpretations may not always be the case (as discussed by Castelfranchi, this volume).

Not only the notion of the individual cognitive unconscious need to be explored, the notion of the collective unconscious need to be explored as well (Sun, 2001), in our attempt to answer some of the afore-identified open questions. For example, culture may largely consist of unarticulated (implicit, subconceptual) processes, in addition to articulated (explicit, conceptual) processes (Sun, 2001). This perspective is similar to the Jungian notions of collective consciousness and collective unconsciousness (Jung, 1959). Pierre Bourdieu (see Bourdieu & Wacquant, 1992) also adopts such a metaphor and sees the ''socio-analysis'' as a collective counterpart to psycho-analysis: It helps to unearth the social unconscious embedded into social institutions and lodged inside individual minds. However, it should be noted that the social unconscious is an ''emergent'' property rather than existent in and by itself (Castelfranchi, this volume). The question is how we should understand and characterize the structures of the social unconscious and its ''emergence'', computationally or otherwise. Answering this question is a crucial step in establishing the micro–macro link, as highlighted earlier, between individuals agents and society.

On the other hand, Burns and Gomolinska (this volume) address the social roles of agents involved in multi-agent interaction and learning. Based on the notion of rule complex (from knowledge representation in artificial intelligence), they describe how individuals change their beliefs (that is, how they learn) in the context, and through the filter, of social relationships, social roles, and social institutions they are involved in. Thus, multi-agent learning is not merely a matter of ''straight'' learning, but a matter involving complex patterns of social interaction and cognitive processes, which leads to complex collective functions.

Beside broad theoretical issues, various technical aspects of (emergent) collective function of individual action need to be explored as well. Work on multi-agent learning in artificial intelligence is particularly pertinent, in that they provide useful tools, techniques, and concepts that can benefit the effort at a broad, multi-disciplinary understanding of individual action and collective function. In this volume, various aspects of multi-agent learning are addressed.

For example, Michael Littman (this volume) deals with value function reinforcement learning in certain types of multi-agent co-learning settings. His focus is on games (von Neumann & Morgenstern, 1944) especially games with competitive equilibria or with cooperative equilibria, which constitute a subset of possible game types. The basic approach is rein-

forcement learning, through estimating the values of different actions. Several different types of reinforcement learning techniques that help to increase the likelihood of achieving Nash equilibria (i.e., stable and rational outcomes) are discussed. Formal results concerning their convergence are given. Such results are useful for advancing formal, mathematical understanding of interaction among multiple agents in co-learning.

There are also models based loosely on economic (market) principles. For example, the work by Baum (1997) relies on an artificial economy for evolving an effective constellation of agents for accomplishing complex tasks. The work by Sun & Sessions (1999), on the other hand, focuses on a simple and effective bidding mechanism for establishing multi-agent cooperation through reinforcement learning, in order to solve complex tasks efficiently.

Hu and Wellman (this volume) dealt with on-line learning about other agents in double auction markets. Their interesting finding is that learning agents making minimum assumptions about other agents actually perform better than agents making more elaborate assumptions about their opponents (which may be wrong or otherwise misleading). It is an open question whether this conclusion is generally true of other circumstances as well, or it is more of a matter of case-by-case analysis. Some recent work suggests the latter (Sun & Qi, 2000).

Cooperation in robotic teams has been the central theme of Maja Mataric's research. In this volume, Mataric discusses a general approach: behavior-based representation for individual robots as well as collections of robots. A variety of techniques are surveyed that range from the use of communication channels to ameliorate partial observability of environments, to the representation of behavior history as a way of coordination, and furthermore to imitative learning by robots from humans and other robots. The on-going explorations of these techniques lead toward practical ways of constructing cooperative robotic teams that learn to coordinate and accomplish tasks jointly.

Overall, there are many issues, problems, and approaches concerning multi-agent cooperation. Among other things, learning (broadly defined) is essential for multi-agent cooperation, and worth much further exploration. Multi-disciplinary collaboration is, by all means, the most promising way of making rapid progress. In fact, centuries of theoretical work on sociology, anthropology, and economics has now been incorporated, in various ways, into the work on multi-agent systems, including multi-agent learning, by many researchers. We are hopeful that this cross-disciplinary cooperation will lead to a better understanding of the sociology of collective function and individual action, as well as a better understanding of human cognitive processes that underlie this process, which may in turn lead to better designs of artificial multi-agent systems. The present volume is a snapshot and a sampling of work in this direction.

References

Axelrod, R. (1984). The Complexity of Cooperation. Princeton University Press, Princeton, NJ.

Baum, E. (1997). Manifesto for an evolutionary economics of intelligence. In Neural Networks and Machine Learning. Ed. C. M. Bishop. Springer-Verlag, Berlin.

Bourdieu, P. & Wacquant, L. (1992). An Invitation to Reflexive Sociology. University of Chicago Press, Chicago.

Elster, J. (1982). Marxism, functionalism, and game theory: the case for methodological individualism. *Theory and Society 11*, 453–481.

Jung, C. G. (1959). The Archetypes and the Collective Unconscious. Pantheon Books, New York.

Smith, A. (1976). The Wealth of Nations. Clarendon Press, Oxford.

Sun, R. (1999). Accounting for the computational basis of consciousness: A connectionist approach. *Consciousness and Cognition 8*, 529–565.

Sun, R. (2001). Cognitive science meets multi-agent systems: a prolegomenon. *Philosophical Psychology 14*(1), 5–28.

Sun, R., Merrill, E. & Peterson, T., 2001. From implicit skills to explicit knowledge: a bottom-up model of skill learning. Cognitive Science, 25(2).

Sun, R. & Qi, D. (2000). Rationality assumptions and optimality of co-learning. In: Zhang, C. & Soo, V. (Eds.), Design and Applications of Intelligent Agents. Lecture Notes in Artificial Intelligence, Vol. 1881. Springer-Verlag, Heidelberg, Germany, pp. 61–75.

Sun, R. & Sessions, C. (1999). Reinforcement learning with bidding for automatic segmentation. Intelligent Agent Technology: Systems, Methodologies, and Tools. World Scientific, Singapore, pp. 84–93.

von Neumann, J. & Morgenstern, O. (1944). Theory of Games and Economic Behavior. John Wiley and Sons, New York.

West, R., & Lebiere, C. (2001). Simple games as dynamic, coupled systems: Randomness and other emergent properties. *Cognitive Systems Research 1*(4), 221–240.