Editorial: Economic Principles of Multi-Agent Systems

Craig Boutilier, University of British Columbia
Yoav Shoham, Stanford University
Michael P. Wellman, University of Michigan

May 7, 1997

An oft-noted feature of artificial intelligence research is its interdisciplinary nature. AI draws on, and contributes to, research in such diverse fields as philosophy, logic, psychology, linguistics, neuroscience, mechanical engineering, control theory—and of course many other areas of computer science. A discipline of great relevance to AI which has perhaps received less attention until recently is economics.

A primary goal of AI is the creation of decision-making artifacts, or agents, that act, with varying degrees of autonomy, on behalf of their designers, owners, or users. In service of this goal, the predominant approach in AI has been to endow the agents with *interests* of their own (coinciding with those of the party on whose behalf they act) as well as reasoning mechanisms based on principles of rationality of one kind or another. Naturally, then, when AI researchers have considered multi-agent systems, the agents have generally been similarly endowed with interests, rational decision-making capabilities, and—importantly—the ability to model the interests and capabilities of other agents.

Stated this way, it is at first puzzling why the connection between AI and economics has not been stronger heretofore. After all, the decision-making behaviors of self-interested agents as they interact with their environment (including, especially, other self-interested agents), the resulting emergent phenomena, and mechanisms that can influence these behaviors, are the province of economic theory. Why then has economics not had a more tangible impact on AI? We see two fundamental reasons, one substantive and one cultural.

The slight economic influence on AI to date may be due in part to the different objectives of the fields. Economics is primarily concerned with explaining the decisions and interactions of rational self-interested agents (or communities thereof), or designing policies that influence these interactions to further certain global objectives. AI, in contrast, is largely (though not exclusively) concerned with constructing self-interested agents, the very entities economic theory takes for granted. Although modeling computational entities as rational beings is standard AI practice, we do not generally have the luxury of *assuming* rationality—it is our burden to explain how to realize approximately rational behaviors in operational computational terms.

The different intellectual heritage of the two fields may also account for the dearth of connections between AI and economics. At its inception, AI was influenced by several intellectual movements. One was philosophical epistemology, reflected in the very influential logicist approach pioneered by McCarthy. Another was psychology, reflected in the work of Newell and Simon, and, to an extent, Minsky. Neither of these movements is strongly reflected in modern economic theory. In particular, modern microeconomics, decision theory, and game theory are all couched in a strict Bayesian setting, which until recently was foreign to both the logicist and psychologicist elements of AI.[1]

While the distance between AI and economics is considerable, it is shrinking, as researchers in both fields are finding light on the other side of the "rationality abstraction barrier." Within economics, there is an increased interest in practical, psychologically-oriented, or computational theories of decision making. Recent studies have taken account of computational limits of various sorts, expressly model learning processes, or measure computational or communication complexity of economic institutions and decisions.

Within AI, we see several reasons for interest in economic theories. First, economic (especially decision-theoretic) principles offer realistic and flexible models for the design of self-interested agents. These principles have been adopted in AI and given a computational flavor. Additionally, the Internet has emerged as a productive environment for deployment of computational agents, and its inherently distributed nature has naturally directed attention toward multi-agent systems. New media for agent interaction open research questions in the design of mechanisms and protocols, and methods for predicting or influencing the behavior of agents within these mechanisms. Participating agents have substantial incentive to model and reason about the others, and both economic and computational concepts bear directly on this task.

The overlap in problems and techniques has not in our view yet been matched by a proportional amount of expressly interdisciplinary effort. Nevertheless, these considerations clearly suggest that mainstream economic principles can play a crucial role in the study of multi-agent systems, and, indeed, in recent years they have increasingly done so.[2] Thus it should not be surprising that

[1] We remark in passing that this observation also resolves an apparent paradox: the connection between AI and economics has been weak despite the fact that one of the founders of AI—Herbert Simon—won a Nobel Prize in economics. However, Simon's brand of economics was strongly psychological in flavor, and by-and-large has not been reflected in mainstream economics work (but see later comments about recent trends in economics).

[2] Evidence of this trend is the granting of the 1995 Computers and Thought Award to two individuals for research significantly influenced by strands of economic thought. Because of this connection, we have invited the recipients—Stuart Russell and Sarit Kraus—to reprise their respective award lectures in this collection.

in response to our call for papers on "Principles of Multi-Agent Systems," the principles that emerged have been primarily economic. For this reason, we have named the resulting journal issue "Economic Principles of Multi-Agent Systems."

When considering the broad distinguishing characteristics of AI research vis-à-vis economics, it is worthwhile to ask where AI has diverged from economics, where AI methods can exploit economic principles and theories, and what AI has to offer to the further development of such theories.

The complexity of solving decision problems forced AI researchers to adopt simplifying assumptions in the earliest days of the field. A clear example is the almost exclusive focus—until recently—of planning research on sequential decision problems with fixed goals, deterministic actions, and complete knowledge. The economic view of self-interested agents as utility maximizers, while not rejected outright, was not adopted in its full generality, as classical planning concentrated on combinatorial issues. Work in decision-theoretic planning over the last several years has over the last several years has considered more general decision makers and has embraced at least part of the economic perspective on agents. This move has been forced by the need to design agents that deal with uncertainty, competing objectives, and so on. Techniques from decision theory, operations research, and control theory have all been adopted in this enterprise.

The combinatorial nature of decision problems has consistently plagued AI methods, more so than economic models, again due to the difference in outlook. In formulating a model to explain some economic activity or phenomenon, an economist can carefully craft the problem specification, ignoring irrelevant features, abstracting away details, approximating details and solutions where appropriate. This process itself requires a considerable degree of intelligence. The AI researcher has no such luxury. An intelligent agent must be equipped to deal with a large number of problem features thrown its way. The resulting combinatorial explosion threatens to make reasoning and deciding impossible without a designer or user nearby to help, say, rule out irrelevant details. AI research has focused on two general, complementary approaches to this problem.

The first is the use of intensional, feature-based problem representations that allow large decision-making scenarios to be specified concisely. Work in knowledge representation—using both logical and probabilistic formalisms—has emphasized precision in problem specification and solution characterization, tradeoffs in expressive power and tractability, and the use of structural properties laid bare by good representation schemes to ease the computational burden of reasoning and decision making. Poole's paper in the collection, *The independent choice logic for modelling multiple agents under uncertainty*, addresses certain representational issues along these lines. Specifically, he presents a logic-programming-style language and methodology for the specification of self-interested agents and their interactions with each other and their environment.

The second approach explicitly accounts for time and other costs of deliberation or computation in determining appropriate courses of action. If the calculation of an "optimal" action is not worth the computational effort, the rational behavior is not optimal in the idealized sense. Thus one must often decide what (and how long) one will think about before thinking about it. This notion is strongly tied to the decision-theoretic concept of value of information, and more generally the economic concept of opportunity cost. It is just one of the ways that rationality notions differ for bounded and idealized agents. In *Rationality and intelligence*, Russell argues that the bounded case is the relevant one for AI, and offers a conceptual framework for analyzing the rationality of bounded agents.

This last point suggests an interesting feature of the applicability of economic theories to computational systems. Most economic models assume idealized, rational decision makers interacting in narrow, precisely prescribed ways. These assumptions, while critical to the tractable exposition and implementation of any theory, often fail the test of descriptive adequacy. However, what may be unrealistic with respect to rich environments populated by imperfectly understood interacting human agents, may often provide adequate descriptions of restricted environments populated by formally specified interacting computational agents. Moreover, to the extent that AI develops well-characterized models of rational or approximately rational computational agents, we can provide ideal domains for investigating and applying economic theories.

We have described a strand of AI research that deals with designing self-interested agents, which merely gets us to the "base-level" assumptions of most economic models. Economics is a social science, and as such, the primary role of this individual level is as a foundation for analyzing agent interactions, emergent properties of these interactions, and mechanisms that influence the interactions and their results. As noted above, considerable recent work in AI has addressed interactions within systems of computational agents. One reason (but by no means the only application) is the emergence of networked communications and interactions, especially the Internet, as an importance facet of everyday life. The opportunity to have computational agents do real work on behalf of human clients and users delivers a pressing need for the development of software agents and interaction mechanisms to facilitate complex decision making in such settings. The existence of collections of approximately rational *computational* agents, sorely lacking in the past, is in part available to the AI researcher as computational agents of increasing sophistication emerge (a trend that is sure to accelerate in the near future).

As anticipated in early work of Doyle, and reinforced subsequently with work by Rosenschein and his colleagues, the framework of game theory has exerted significant influence on formal models of multi-agent systems.[3] One of the more active contributors of game-theoretic models to the AI literature is Kraus, whose

[3]One could argue that game theory's influence on AI dates from the first minimax search chess algorithm developed by Shannon circa 40 years ago. However, prevalent adoption of the general game-theoretic formulation of multi-agent action is considerably more recent.

Negotiation and cooperation in multi-agent environments surveys much of this work. In this article, Kraus also considers other "multi-entity methodologies," and discusses their relevance to the design of systems of interacting agents. Sandholm and Lesser, in *Coalitions of computationally bounded agents*, apply game-theoretic concepts to the particular issue of coalition formation. Specifically, the authors characterize the propensity of bounded computational agents to act as cooperative subgroups in a well-defined class of interaction problems. Their analysis illustrates the value of combining economic (game-theoretic) and computational (complexity) analyses in situations where computational properties affect the value structure of the problem.

Distributed coordination among multiple agents is of central importance to both game theory and AI. In *On the emergence of social conventions: models, analysis, and simulations*, Shoham and Tennenholtz introduce a stochastic model in which local behavior rules lead to coordination over time. While similar in flavor to several models within game theory, and even couched in a game theoretic setting, this paper is distinguished by its computational orientation. In particular, most of the results in the paper concern the rate with which conventions evolve, and most of these results were obtain through computer simulations.

Koller and Pfeffer's paper, *Representations and solutions for game theoretic problems*, similarly places a computational spin on familiar game theoretic notions. Game theoretic formulations are attractive in part because of their generality, but this generality comes at a price. The search for joint agent strategies satisfying a solution criterion (e.g., Nash equilibrium) can grow combinatorially with the individual strategy spaces, which in turn may be exponential in size of the domain. Koller and Pfeffer propose a logic-programming-style language with which to naturally and compactly specify games by exploiting the structure of the game. They also describe an implemented system, Gala, which efficiently interprets these specifications and computes solutions of the game.

This type of work suggests one sort of product that AI has to offer to economics. Work on specification of agent interactions—perhaps, as in Gala, expressed in terms of programming constructs—can be expected to expose further regularities in multi-agent situations, and thus provide further opportunities to exploit game structure.

Other potential contributions of AI to economics draw on AI's models of the agents themselves. To the extent that research in the design of self-interested agents informs (and is informed by) descriptive theories of such agents, the results of economic and AI research mutually benefit the other. Specifically, notions such as computation cost and bounded rationality, as well as the use of specific forms of representation and their influence on reasoning, should impact economic models as they become more realistic and accurate, just as they have the design of self-interested agents in AI.

In many economic models (e.g., of bargaining), not only is the specific structure of agents largely unanalyzed, so are the means by which agents interact or communicate. Once again, the emphasis in AI on building agents that interact requires that concrete theories and models of this interaction be developed. Such normative or practical models can clearly have an impact on the descriptive models required in economics.

That agents act on behalf of their designers or users is a critical assumption underlying much research in AI. It requires techniques for interacting with people (or other agents) in order to determine models of the environment and preferences of the user. Knowledge acquisition, especially the automation of the process, is a key element of the AI enterprise that does not have an exact counterpart in economic theories, where agents are assumed to the know the context in which decisions must be made. Models of preference elicitation and revealed preference do play a role in decision analysis and economics, but are especially crucial in AI. In AI the focus of some research is on interactions with users that is straightforward and efficient. Reasoning directly with qualitative information about beliefs and preferences is an area where AI seems poised to contribute to economic understanding. Brafman and Tennenholtz's article, *Modeling agents as entities with a mental state*, formalizes a form of belief ascription within this category.

We see that, while the aims of economics and AI are somewhat different, the models and solution techniques should have much in common. It is surprising that the two fields have not had tighter connections throughout the years, but it is encouraging to see the fields moving closer together. The papers in this volume are representative of this effort. Each embraces some aspect of some economic theory in its approach to an AI problem. Each also offers some insight into the model being adopted, illustrating the potential for AI research to have an influence on economics. We sincerely hope that this volume, or more precisely, the research of which this collection is a small sample, presages substantial interaction between research and researchers in economics and AI.