The Economic Approach to Artificial Intelligence

MICHAEL P. WELLMAN

University of Michigan, Artificial Intelligence Laboratory, Ann Arbor, MI 48109-2110
⟨wellman@umich.edu⟩
http://ai.eecs.umich.edu/people/wellman/

To take an economic approach to *anything* typically invokes certain premises: First, that the fundamental problem to be solved is one of *resource allocation*; second, that it is useful to model behavior in terms of a *rationality abstraction*; and third, that it is essential to consider how authority and activity may be *decentralized*. All these premises are being increasingly adopted (explicitly or implicitly) in artificial intelligence, and growing numbers of AI researchers are working within the *economic paradigm*.

Resource Allocation

Computer scientists like to view a program as an abstract specification of a *machine*, describable behaviorally in terms of the input/output relationship resulting from its computation. The machine's *product* is its output, representing the value of a function at the point represented by its input. Often we find it helpful to view this product at a higher level, say, as the solution to some well-posed problem. Inevitably, this problem bears on what we are to do, that is, some *course of action* to be embarked upon. (Conceptions of computation as *answering questions* are relics of the era when human intermediaries were necessary to perform the transduction from computation to action.) In this view, the computer is a *decision machine*, where a *decision* is the resolution of a distinction among potential courses of action.

Without loss of generality, every decision—hence every computation [Doyle 1992]— is really about resource allocation. Choosing to do something entails an allocation of attention and other activity resources to that thing in lieu of others. Conversely, an allocation of resources defines the activities done and not done. Making such choices appropriately involves weighing the benefits of the activities done against the opportunity cost of forgoing those not done.

So far, we have only tautologies. Sure, every problem can be cast as one of the resource allocation, but what is the concomitant advantage? The answer is that, without considering resources explicitly, it is difficult to express the range of courses of action available, as defined by configurations of resources devoted to the various activities. Perhaps even more seriously, without acknowledging gradations in value (or likelihood of outcome allocations), it is impossible to take into account trade-offs among alternate activities.

It is widely recognized that many of the problem-solving techniques developed in AI research (e.g., so-called *classical* planning) need to be generalized to accommodate uncertainty and graded preferences. Work in *decision-theoretic planning* [Hanks et al. 1994] is beginning to address these problems, adopting a more comprehensive framework for principled resource allocation while attempting to retain useful computational and representational techniques from prior AI work.

Rationality Abstraction

Most of microeconomic theory assumes that individual agents are *rational* —acting so as to achieve their most preferred outcome, subject to their knowledge and capabilities. Indeed, this *rationality abstraction* is perhaps the single methodological feature that most distinguishes economics from the other social sciences.

This approach is highly congruent with much work in artificial intelligence. About fifteen years ago, Newell [1982] proposed that a central characteristic of AI practice is a particular abstraction level at which we interpret the behavior of computing machines. Viewing a system at Newell's *knowledge level* entails attributing to the system knowledge, goals, and available actions, and predicting its behavior based on a *principle of rationality* that specifies how these elements dictate action selection. Rationality as applied here is a matter of coherence, defining a relation in which the knowledge, goals, and actions must stand. This is exactly the Bayesian view of rationality (standard in economics), in which knowledge and goals (or beliefs and preferences) are subjective notions, constrained only by self-coherence (consistency) and coherence with resulting behavior.

In introducing the knowledge-level idea, Newell proposed a particular basic rationality principle:

> If an agent has knowledge that one of its actions will lead to one of its goals, then the agent will select that action.

This formulation can be criticized on several grounds, most having to do with its relegation of all matters of resource allocation and graded preferences to some rather ad hoc auxiliary principles. From the economic perspective, a satisfactory comprehensive rationality principle should address choice among alternate activities and resource allocations that accomplish goals to varying degrees. But regardless of the particular expression, the point is that some coherence-based rationality principle is required in order to make sense of the sorts of agent attitudes—knowledge, belief, preference, intention—commonly invoked in AI research.

Decentralization

In human societies, computational power is inherently distributed across many relatively small brains resident in separate skulls, connected by costly, low-bandwidth, error-prone communication channels. Moreover, authority over activity is separately controlled by the local computational units. It is therefore not surprising that economics focuses on the decentralized nature of decision making. A primary aim of the discipline is to explain the aggregate results of alternate configurations of interacting rational agents.

The case for decentralization in computational environments, where communication is usually more direct and configurations more controllable, is less straightforward. Nevertheless, a variety of technological and other factors are leading to computational environments that are increasingly distributed. At this writing, the development and promotion of "software agents" (not necessarily derived from AI technology) is a prominent activity. Although interpretations of software agency vary widely, typical conceptions involve autonomy of action, modularity of scope and interest, and interaction with other agents. Understanding and influencing configurations of software agents is directly analogous to the problem faced by economists.

Within economics, the problem of synthesizing an interaction protocol via which rational agents achieve a socially desirable end is called *mechanism design*. This is exactly the problem we face in designing distributed software systems, which suggests an opportunity to exploit existing economic ideas. One rich source of ideas is game theory, and some recent AI work directly appeals to game-theoretic concepts in designing multiagent interaction protocols [Rosenschein and Zlotkin 1994].

Sometimes we have no choice—the designers of agents may not have control over the interaction protocol, and agent interfaces may be undefined or open-ended. In such cases, we must often resort to economic interactions simply because—in our society—the market system is the de facto default interface. Our software agents should be equipped to deal with this interface. One can take this to the extreme, as I have done in recent investigations [Wellman 1993], and design distributed decision systems explicitly as competitive computational economies.

Conclusion

It is not possible in this short position paper to survey the large body of work on probabilistic reasoning, decision-theoretic planning, game-theoretic analysis of multiagent systems, and so on, that has made its way into AI over the last ten years. Suffice it to say that the field has been far more open than in the previous decade to ideas that could be broadly characterized as economic. That these ideas have had significant impact in particular subfields is reflected in the ubiquity of concepts of resource allocation and rationality in the recent AI textbook of Russell and Norvig [1995].

This is not a surprise. As I have attempted to point out, the goals of AI and those of economics overlap substantially and are analogous in many of the nonoverlapping regions. AI is the branch of computer science that is concerned with the substance of behavior and with deriving general principles for designing deciding agents. In so doing, AI unapologetically invokes rationality concepts and aims to render the rationality abstraction an operationally viable approximation. When activity is decentralized, AI considers interactions in social terms.

The point of all this is not, of course, to suggest that economics has all the answers to AI problems. But recognizing that AI's problems are in large part economic does help us to formulate the questions and opens to us a variety of concepts and techniques that offer a starting point on potential solutions. Success in AI would mean an account of the economics of computation, and one way toward this goal starts with some computation of economics.

REFERENCES

DOYLE, J. 1992. Rationality and its roles in reasoning. *Comput. Intell 8*, 376–409.

HANKS, S , RUSSELL, S., AND WELLMAN, M. P Eds. 1994. *Working Notes of the AAAI Spring Symposium on Decision-Theoretic Planning* AAAI Press.

NEWELL, A. 1982 The knowledge level. *Artif. Intell. 18*, 87–127

ROSENSCHEIN, J. S. AND ZLOTKIN, G. 1994 *Rules of Encounter: Designing Conventions for Automated Negotiation among Computers* MIT Press, Cambridge, MA.

RUSSELL, S. J. AND NORVIG, P 1995 *Artificial Intelligence: A Modern Approach.* Prentice-Hall, Englewood Cliffs, NJ.

WELLMAN, M. P. 1993. A market-oriented programming environment and its application to distributed multicommodity flow problems. *J. Artif. Intell. Res. 1*, 1–23