Reset reproduction of a guest editorial in *AI Magazine*, Vol. 9, No. 1 (Spring 1988), pp. 19–22.

Current address: MIT Laboratory for Computer Science, Cambridge, Massachusetts.

Big Problems for Artificial Intelligence

Jon Doyle

Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

January 5, 1988

Abstract: We compare the big problems studied in artificial intelligence and related fields in order to understand some major changes—both internal and external—recently suffered by AI. The comparison finds AI with few problems to call its own, and we identify some further major changes that may occur soon.

This research was supported by the Defense Advanced Research Projects Agency (DOD), ARPA Order No. 4976, Amendment 19, monitored by the Air Force Avionics Laboratory under Contract F33615-87-C-1499. The views and conclusions contained in this document are those of the author, and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the Government of the United States of America.

Some old hands at artificial intelligence have been given recently to deploring the state of research in the field. Work in earlier days seemed heroic, filled with the excitement of being first to discover and conquer big problems, while much work today seems a bland pursuit of more of the same. Where, the old hands ask, have all the big ideas gone? This note attempts to respond to this complaint, for the answer sheds light on the nature of artificial intelligence.

1 Has AI changed?

This change in the character of the field is a real change with several causes, and not simply an illusion. Two factors immediately spring to mind:

- To some extent, it reflects the maturation of the field. As the easy parts of problems are solved, the remaining problems are harder, making progress slower and results smaller.

- In addition, much effort has been devoted to articulating and codifying knowledge and skills in many subjects, from medicine and manufacturing to discourse and physical imagination. This effort, while extremely valuable to civilization, sheds little or no light on understanding the mechanisms, as opposed to the substance, of thinking and intelligence, and hence carries no excitement for old hands, only the tedium of infinite repetition of familiar ideas. (Indeed, as it has matured, it has been largely divorced from artificial intelligence with its own conferences and journals.)

While these two factors explain much of the change in the field's character, there is something more to it, in fact, something more important than changes in the speed, size, and subjects of progress. Specifically, the source of problems for the field has changed as well, so that while artificial intelligence still has big problems demanding big ideas, many traditional paths of research do not lead to them, but only to backwaters.

2 Understanding AI problems

The fundamental observation we will use to understand the disquiet in artificial intelligence is that almost all of the field's big problems are problems shared with most or all of the mental and social disciplines of psychology, philosophy, economics, politics, sociology, and management. This is, put differently, a traditional thesis of artificial intelligence, namely that the hardware may vary but the basic problems of intelligent action remain the same.

For example, one big problem is rationality. This notion permeates all of artificial intelligence's relatives but less so artificial intelligence itself (see [Doyle 1987], which analyzes some familiar artificial intelligence notions in terms of rationality). But most of the fundamental problems involving rationality do not come from artificial intelligence at all. Rationality involves notions of consistency and completeness of mental attitudes. As is well known in political economy, these requirements are typically unattainable, and this infeasibility gives rise to fundamental notions of limited rationality. (See, for example, [Simon 1969], [Minsky 1986], [Mueller 1979], [Doyle 1988].) Artificial intelligence adds its own contribution to the study of rationality by using notions of computational complexity to suggest more refined notions of limited rationality. One day, these refined notions may help explain observed limitations in the everyday rationality of people. But artificial intelligence does not supply or solve the problem of rationality alone. Its contributions depend on older ideas from other fields, and gain much of their significance from the intellectual respectability and proven value of these older ideas. Other big problems, in my view, include motivation, growth (adaptation, improvement, learning), purpose (meaning, design), perception, representation (imagination), and communication. Of course, this list is not necessarily complete, nor especially novel. One can learn more about these topics by reading James' *Principles of Psychology* or Mill's *Principles of Political Economy* than most textbooks of artificial intelligence.

Of course, each field addresses a set of problems, some of which are shared with related fields, and some of which are its own characteristic problems, or characteristic methods of approaching the shared big problems, that distinguish it from related fields. In artificial intelligence, the characteristic problem (and simultaneously, the characteristic method) is mechanization, that is, the efficient implementation of the solutions of the shared big problems about thinking. This problem interacts with the others only to the extent that some big problems are demonstrably intractable with respect to accepted forms of mechanization. Whether mechanization really sheds much light on other big problems is controversial, though the utility of such mechanizations as are possible is less in question. (It is curious that artificial intelligence has tried to make its idiosyncratic problem an imperial one by recent work (e.g., [Genesereth and Nilsson

1987]) that claims, in contrast to the historical view in the field, that questions in other fields are *essentially* illuminated by the computational perspective.)

One consequence of these observations is that since the big problems draw much of their substance from outside artificial intelligence, good formulations of problems involving these ideas must make sense in all of the fields, not just artificial intelligence. Formulating problems about learning in terms of CAR's, CDR's, frames, or even sentences, is not likely to be fruitful, any more than is formulating problems about sorting in terms of FORTRAN integer arrays. Another consequence is that verifications of the importance of most problems lies outside of artificial intelligence. That is, the artificial intelligence researcher must have some acquaintance with other fields to understand the proper place of artificial intelligence problems. Problems that appear in each of the fields are likely to be central and difficult ones. Problems that arise only in the discourse of one field are suspect as major problems. Either they are non-problems that stem from local or historical confusions, or they are one of the characteristic problems (whether large or small) that distinguish the field from its relatives.

3 How AI's situation has changed

The observation that big problems are shared among many fields explains several elements of the old hands' depression. One element of the old hands' despair is that it is often difficult for artificial intelligence researchers to have ordinary confidence in their competence: insecurity and overconfidence are quite common. The method of verifying the importance of one's work leads one to expect extreme degrees of confidence for those who look only within artificial intelligence to judge their work, for the field encompasses a zoo of schools and factions, each working on seemingly incommensurable problems, with the big problems of one school being trivialities or nonsense to other schools. To be confident in the soundness of one's work, one must be confident of the importance of the problems one addresses, and looking purely within artificial intelligence, one sees no monarch granting patents, only warring fiefdoms. Artificial intelligence alone cannot supply respectable formulations of problems that admit respectable solutions. Only by looking outside of artificial intelligence can one verify the respectability of problems that makes confidence in research possible. This makes life increasingly insecure for anyone whose gaze remains fixed on traditional artificial intelligence problems. Most old hands realize this, at least unconsciously, and are facing the painful choice between the easy path of continuing work as usual, which may mean continued insecurity and possible sterility, and the harder path of increasing one's understanding and confidence by giving up some old problems and ways of thought for more informed problems and ways of thinking.

Another element is that for much of its history, artificial intelligence has worked in a virtual intellectual vacuum, isolated (for reasons that need not concern us here) from most work in related fields. This meant that the standards of novelty and competence were much lower than today. In the past, one could discover and publish important ideas new to artificial intelligence though long known elsewhere. And in the past, one's audience for these ideas was composed of other innocents. But no more. Starting around 1980, artificial intelligence began to change from a largely academic and speculative field into one with a large industrial and applied component. Large numbers of new people have entered, considered, or employed the field. This change is still in progress, but two consequences are already visible.

First, the character of work in the field is changing, becoming more formal and precise. While in the past artificial intelligence was almost purely speculative, with no immediate influence in external productive affairs, its ideas and techniques now appear in systems relied upon by people and organizations, directly and indirectly affecting people's lives. Past work was often creative but rarely precise, since all that mattered was that the general idea of the work be suggestive enough to other (usually personally known) researchers so that they could figure out how to place it within their own favorite vocabulary and methodology. This arrangement allowed some progress in spite of widely differing ideologies within the field about the substance of the field. In the future, however, the field will have to offer to its applied, engineering component those intellectual foundations required by all engineering disciplines: sound, precise understandings of at least the fundamental ideas and techniques, ways of judging reliability, stability, powers and limits of designs, and communicable, replicable, and verifiable specifications of designs and realizations. The field will have to supply this level of precision to its engineers, even if its scientists prefer a less formal language and method, lest the engineers produce it themselves and render the scientists superfluous.

Second, much of this transformation of the field may be accomplished by newcomers, not by previously established members of the field. The great publicity accorded artificial intelligence at present is already attracting outsiders competent in other fields who see immediate opportunities to "clean up" the treatment of the shared big ideas within artificial intelligence. These newcomers include theoretical computer scientists, logicians, decision theorists, statisticians, operations researchers, mathematicians, philosophers, linguists, and others. Even if established members of the field have the competence and the inclination, the sheer amount of work to be done means that large fractions will be done by the newcomers. Indeed, the publishers prize for best paper at the 1985 International Joint Conference on Artificial Intelligence was won by two theoretical computer scientists, Fagin and Halpern, who applied their expertise in the theory of distributed computing systems to the problems of incomplete knowledge.

Because of these two changes, one can expect the field to look very different in 5–10 years time, even if the basic set of theoretical ideas and computational techniques remains largely the same. For these external reasons, research in the field will not be viable if conducted as in the old days. Today, novel ideas must be novel in all fields at once, not just in artificial intelligence, and treatments of these ideas must satisfy standards of competence established in each of these fields. This has greatly increased the threshold of achievement needed to produce the pleasure of discovery or conquest, making impossible the innocent pleasures of the past.

4 Will AI disappear?

As people in other fields recognize their own problems within artificial intelligence, they will reformulate them, taking away the special trappings of their former treatments. What will remain if all the big problems are stripped from artificial intelligence and accorded to other fields?

While it is possible that artificial intelligence will disappear altogether, what seems more likely is a redivision of the field, some of which may already be observed to be taking place. If one considers the distinct intellectual tasks to be accomplished, a redivision along the following lines seems possible.

- First, each of the standard fields related to artificial intelligence will continue to absorb the method of looking for computational complexity explanations for their phenomena. The parts of artificial intelligence most directly affected by this absorption are the branches concerned with human psychology and linguistics. A good example is how Marr's [1982] work transformed AI work on vision.

- Second, work on *articulating intelligence*, on codifying common and expert knowledge on all topics will continue, increasingly as an activity separate from computation (though often with the ultimate aim of automation). This sort of work has been going on since long before computers arrived. The areas of expert systems and commonsense knowledge (e.g., naive physics, discourse conventions) fall into this category.

- Third, the mathematical and theoretical studies in artificial intelligence, logic, economics, statistics, etc. of possible psychological structures and organizations will draw together into a new mathematical science. [Doyle 1983] calls this field *rational psychology*, meaning the conceptual investigation of psychology by means of the most appropriate mathematical notions, with the aim of understanding the underlying nature of and connections between psychological concepts.

- Fourth, and finally, much of current artificial systems and techniques of mechanization will draw together as the engineering discipline corresponding to the mathematical field of rational psychology and to the scientific fields of psychology and economics. The aim of *psychological engineering*, as [Doyle 1983] calls it, is parallel to the aim of any engineering field, namely to find economical designs for implementing or mechanizing agents with specified capacities or behaviors.

For example, the characteristic concern of artificial intelligence, mechanization, is treated differently in human psychology, in rational psychology, and in psychological engineering. Human psychology, of course, seeks both to discover the materials—chemical and neural—from which human minds are constructed, and how human behaviors are realized in these materials. Rational psychology, like the theory of computation and computational complexity, would consider questions of existence of mechanizations of specific classes of psychologies by means of specific classes of machines or materials. Psychological engineering, like the field of design and analysis of algorithms, would be concerned with inventing and comparing machines that realize specified psychologies.

Thus artificial intelligence will not disappear, but it will likely be greatly transformed in ways that will demand adopting new standards of competence and learning new ways of thinking and writing about its problems.

Acknowledgments

I thank David McAllester, Johan de Kleer, Geoffrey Hinton, Ramin Zabih, Gregory Provan, Gerald Sussman, and Joseph Schatz for valuable discussions of this matter.

References

Doyle, J., 1983. What is rational psychology? toward a modern mental philosophy, *AI Magazine*, V. 4, No. 3, 50-53.

Doyle, J., 1987. Artificial intelligence and rational self-government, Pittsburgh: Department of Computer Science, Carnegie Mellon University.

Doyle, J., 1988. On universal theories of defaults, forthcoming.

Genesereth, M. R., and Nilsson, N. J., 1987. *Logical Foundations of Artificial Intelligence*, Los Altos: Morgan Kaufmann.

Marr, D., 1982. *Vision*, San Francisco: W. H. Freeman.

Minsky, M., 1986. *The Society of Mind*, New York: Simon and Schuster.

Mueller, D. C., 1979. *Public Choice*, Cambridge: Cambridge University Press.

Simon, H. A., 1969. *The Sciences of the Artificial*, Cambridge: MIT Press.