Knowledge And Experience In Artificial Intelligence

Bernard Meltzer

Via G. Galilei 5, 21027 Ispra (VA), Italy

The period since the last conference in this series has been characterized by the explosive expansion of AI out of the confines of institutions of basic research like university departments into the worlds of industry, business, and government (a development I had long expected). But it seems to me that there are plenty—perhaps an overabundance—of other occasions, other conferences, other workshops, and the like, at which the applications of AI would appropriately be considered. I will confine my remarks, therefore, to issues of basic research.

In fact, it is ironic—though perhaps it may be understandable—that precisely now, when the outside world has discovered and started showing its appreciation of AI and its potential, there is a widespread malaise among research workers in the field about the health of their subject. This malaise has to do not only with logistic issues such as the drain of very good people from research into applications, or some of the gross inadequacies of structural and funding support by governments. It has to do also with the very heart and methodology of the subject. For instance, not very long ago in the United States a questionnaire was sent out to the main relevant research groups requesting their judgments on the salient aspects of what has by common consent for many years been a major issue—namely, the representation of knowledge—and the main result of the exercise was that there is practically no common view on any one of these aspects. And there has been a plethora of impressions of the field published by leading workers like Newell, Nilsson, McCarthy, Waltz and Schank (particularly in the *AI Magazine*), often very distinctly voicing this malaise: in regard to a lack of accepted criteria in evaluating pieces of research, the dismal lack of progress in some

directions, or sometimes agonizing over the definition of artificial intelligence or questioning whether it can claim to be a scientific discipline at all. And certainly, looking at the work coming out of laboratories, one sometimes has the impression of a marking of time, as of a gramophone record that gets stuck in a groove, maybe turning out one Ph.D expert system after the other.

I have no intention of decrying expert systems: They are obviously a tremendous advance in the methodology of problem solving by computer over the more conventional approaches, even though they are based only on principles discovered in the earlier years of AI. But as far as I know no fully implemented expert system except perhaps one has any substantive capacity to learn from experience—surely a strange property is an "intelligent" system. In fact, in one of the surveys I referred to earlier, Roger Schank asserted that a program that couldn't learn was not an AI program—a judgment that would exclude most of the work ever done in our field including, incidentally, the excellent work of Schank's own group at Yale on modeling understanding of natural language. Although I doubt that Schank's view, taken literally, is tenable, it points to deeper issues in the development and history of our subject.

Right from the start of the subject about 40 years ago, John McCarthy emphasized, quite rightly, the centrality of the task of modeling commonsense reasoning and the knowledge required for it. In due course this led to intensive study and experimentation on ways of representing knowledge and modes of inference. A peculiarity of most of this work is remarkable: the modeling of our notions of space and time, which are basic to the commonsense view of the world, was hardly broached in any serious, principled way; at least until comparatively recently it was either ignored or dealt with in fragmentary and ad hoc ways. I

don't know if others in the field have had the impression I have sometimes had, looking into the design of these programs: they seem to model a ghostly, ethereal world in which there is not only no space or time but not even any physical objects—a solipsistic world insomnia, as we say in Italy. Why this lacuna? I suggest that it is one symptom of the fact that the issue of knowledge became so dominant that it was allowed to push into the background the issue of experience. I can best explain what I mean by taking a look at a few well-known research projects of the past.

First, Winograd's famous program for the understanding of a subset of English involved in dialogue about a simple so-called blocks world. It was a landmark in my view, not chiefly because it was the first really successful model in this area of language, nor because it was the first successful example of what came to be known as the procedural embedding of knowledge, but because, besides syntax and semantics, it exploited so-called "real-world" knowledge, and so by implication pointed to the importance of experience. In fact it even pretended to a sort of experience—not, it is true, of the real world of blocks, but a simulation of it which appeared on the screen of a cathode ray tube. As I intend to explain later, in the context of AI, modeling a simulation of a real world is something very different from the actual real world, but Winograd's motivation was directed toward experience of the latter.

My second example is David Marr's work on the modeling of visual perception—an even more important landmark in my view. He pushed the issue of knowledge into the background quite explicitly, and set himself the problem of modeling experience—that is visual experience. Let me explain. If light of various colors and intensities falling on my retina activates various receptors, neurons, and so forth, even giving rise to some more or less confused sensations, that does not become an experience until it has resulted in my becoming aware of some configuration or some object or objects in real space in the outer world. And it was precisely the process, or sequence of processes, conducting toward this state that Marr studied and tried to model. He was studying how perception produces knowledge, instead of the converse prevailing paradigm of high-level knowledge controlling the perceptual process. A notable feature of his approach was the search for generally necessary or highly probable principles that govern such processes in, for example, stereoscopic perception or that of texture. (By the way, I would think the search for general principles is not sufficiently pursued in AI research. For instance, what might be the implication of Immanuel Kant's doctrine, which I am inclined to think may be true, that our intuitions of space and time are not things we extract from experience but are the conditions for the very possiblity of experience?)

My last and perhaps for my present purpose most important example is of work started, I believe, well before that of Winograd's or Marr's—namely, that of the group at SRI International under the leadership of Nils Nilsson, which led to the construction and development of a robot on wheels that came to be quite well-known as Shakey. It was, and probably still is, the most advanced robot ever built, but that is by the way. It had some elementary perceptual and tactile abilities, could plan sequences of simple actions like pushing objects for particular purposes in and between rooms connected by a corridor, avoiding obstacles, moving a ramp to get on to a raised platform, executing and monitoring the execution of its plans, modifying them when things went wrong, and even displaying a modest capacity for serendipity—that is, taking advantage of unexpected features of its environment to perform its current task better.

Now let me first list a few of the achievements of this quite outstanding research program. Its treatment of planning set the stage for most subsequent work in that area. Its representation of space in my view was on the right general lines. It developed a data structure, so-called triangle tables, which could be used not only for storing plans but also generalizing them and using the results to react to unexpected consequences of its actions—so, indeed, Shakey had been provided with a learning capability, essentially by means of what came to be known as "procedural abstraction." Last, but far from least, in the work of Sacerdoti, it broached for the first time in any thoroughgoing way the absolutely fundamental issue in all effective thinking: how to reason at different levels of abstraction; i.e., at different degrees of detail in the representation of a world.

Now the point I want to make is about research methodology: The SRI people did not follow the predominant paradigm of AI research. These remarkable results stemmed from studying the problems of direct experience of the physical world by an agent. So, for example, learning became an issue not because a bright student wanted to do a Ph.D thesis on learning, but because Shakey was laboriously having to work out anew time and again rather similar sequences of actions; and Sacerdoti's hierarchy of abstraction spaces or something like it became a necessity because without it the search space for plan construction became impossibly large as the range of activities of the robot was extended.

Interestingly enough, I understand, in the first stage of the SRI project, they, like Winograd, used a simulated model of the robot's world, but soon became convinced that it was inadequate for their purposes, since the essense of intelligence in thought and behavior is coping with the new and unexpected. It is almost impossible to provide that capacity satisfactorily by simulation. As I believe Raphael put it, they found that the best model of the world was the world—it was cheaper!

The end of this particular story is a sad one, and I want to make a brief digression to recount it, for it is instructive in regard to the politics of basic science. The major part of the funding of the project, when still in

mid-career, was stopped—apparently because the funding agency, part of the defense sector of the U.S. government, was unconvinced about its usefulness for military applications in respect to their time scales. One might draw the conclusion that support for basic science should not be controlled by the military—a proposition I have always strongly agreed with. Unfortunately, it is also the case that at about the same time the main British government funding agency also did its best to throttle the nascent AI robotics research in Edinburgh on the advice of a very distinguished civilian scientist. The proper lesson seems to be that the practitioners in any field of science are the best judges of where to invest limited resources (in that field). Sir James Lighthill was an applied mathematician who knew nothing about AI until, I believe, about two months before he made his recommendation.

It should now be clear what is the moral of the three case studies I have sketched. I have suggested that in the past AI has suffered from not concerning itself sufficiently with experience. And it needs to do this not only by modeling certain specific kinds of experience as David Marr did, but even more—as a standard methodology for research into cognitive processes—it should give its computational models experiences of the real world, for example with the aid of sensor and motor adjuncts on the SRI paradigm. For in company with most respectable philosophers, I am reasonably convinced that all our knowledge at the last arises from experience and can only be validated by experience. Furthermore, our modes of cognition (which AI is trying to discover) stem from our experiences of the world as individuals and genetically from the experiences of our species.

I had originally intended to limit my remarks to the issue of knowledge and experience; however, possibly what I have been talking about is only one important symptom of a more general aspect of the dominant attitudes in AI research—namely, a tendency not to look outside the immediate ambit of computer programs (perhaps one should term it "hacker's psychosis" or better still, "authackerism"). This kind of disciplinary introversion was notably not characteristic, for example, of David Marr, whose approach to the problems of perception was much influenced by scientific discoveries outside the field of AI , such as those of Hubel and Wiesel (found by experiments on the firing of single cells in the visual cortex of cats and other animals). Let me however give one or two examples of this narrowness of focus.

Aaron Sloman a long time ago pointed out that perfectly valid reasoning can and is in fact carried out without the aid of the logical or quasi-logical apparatus of propositions, functions, and truth values—for example, the use of Venn diagrams for reasoning about sets, or the use of schematic drawings of mechanisms to reason about their operation. Why has nobody tried to use or study how to use such representations inside a reasoning system, for instance by exploiting one of the so many presently available graphics systems in a more interesting way than merely as user interfaces to computers? Such studies could contribute to solving one of the major open questions in psychology: the nature of visualization.

My second and final example is perhaps a considerably more serious one. With very few exceptions, all of AI until now has been concerned with what Freud termed the secondary processes of the mind—that is, those processes concerned with logical, rational, reflexive, or potentially reflexive, commonsense thinking. It has neglected the primary processes—that is, those concerned with apparently nonrational, nonreflexive thinking that results, for instance, in new metaphors, shafts of wit, jokes, dreams, poems, brainwaves, neuroses, and psychoses. There has been an almost complete ignoring of the wealth of insight into such processes that Freud's investigations, based on hundreds of case studies, provided. For instance in his classic interpretation of dreams, he not only identified six of these basic processes (all, I think, in principle representable computationally), but also gave a penetrating analysis of the salient differences between secondary and primary modes of thought—for example the different ways of dealing with negation and implication. I would think some of the puzzles of language use that worry linguists, computational and otherwise, could be clarified in terms of some of these processes. It is interesting—as an example of reinventing the wheel—that one of the best early language programs (MARGIE, developed by Schank's group) implemented, in a rather fragmentary way admittedly, one or two of these processes. If and when AI comes round to tackling seriously the task of modeling creativity, it would become absolutely essential to test the functioning of these primary unconscious processes computationally.

Let me stop there. I know that there are few people more irritating than backseat drivers. I hope I haven't irritated too many of those of you who are actually, and devotedly and laboriously, doing the coding of AI programs, good and bad.