Comments on the "STATE OF TECHNOLOGY IN ARTIFICIAL INTELLIGENCE" chapter of Duda, Nilsson, and Raphael.

Brian C. Smith
Massachusetts Institute of Technology

This commentary on the chapter entitled "State Of Technology In Artificial Intelligence" by Duda, Nilsson, and Raphael (DN&R) consists of three sections. In the first I briefly critique their presentation, and argue for a deeper analysis of the goals and structures of the field. In the second I present an alternative conception of AI, distinguishing the scientific and engineering dimensions of the field, and examining the importance of human intelligence and the role of the computer. Finally, in the third section, as an example of a more theoretical presentation of the state of the art, I discuss the current state of knowledge representation.

1. CRITIQUE

It is generally agreed that in order to understand a body of facts, you need a framework in terms of which to organize and comprehend them. One wishes that Duda, Nilsson, and Raphael took this insight more to heart in writing their paper. While they present an impressive list of facts, and identify a wide ranging set of tools, they do not explain the intellectual basis of the discipline we call "Artificial Intelligence". Consequently, the facts are hard to comprehend, and the tools difficult to assess. The article is more successful in setting forth data than in giving the reader a deep understanding of the underlying enterprise.

In addition, many of the claims set forth -- particularly those stating that certain problems have been understood or solved -- are simply inaccurate. I deny, for example, that "sets of rule-like quanta of knowledge form a sufficient basis for capturing the experienced judgments of experts about several domains including medicine, chemistry, and electronic circuit theory".[1] Although programs have been constructed in these areas which exhibit some degree of sophistication, none of them approaches the sublety, flexibility, common sense, or judgment of human experts.

I also reject the claim that Schank's conceptual dependency formalism is capable of capturing the "true meaning" of certain verbs (section 2a) -- in fact

1. From DN&R's section 2a. See also sections 2d on planning and 2e on perception.

I doubt whether Schank himself would make such a statement.[2] Not only is it unclear how such a claim would be validated, but we don't even have an accepted theory of semantics in which meaning is defined. No one, therefore, can be said to know formally what "true meaning" is. From a computational perspective meaning seems to be a more active notion than the static conceptions of "truth" and "falsehood" of classical logic, but beyond this the concept is still very dimly understood.

In an article on the social impact of computing,[3] Joseph Weizenbaum has expressed concern about a form of "technological mentality". His worry is that from a social and political point of view, computers will be dangerous not because they are powerful, but because society will *think* they are more powerful than they really are. As a result of this misconception, people are likely to relinquish responsibilities to computers for which those machines are inherently unsuited. My concern with the DN&R chapter is that it contributes to this kind of misconception by conveying a sense of surface impressiveness with no deep communication of what is behind (or what isn't behind) the claims. I worry that the reader will go away from the chapter either confused or with a serious misconception both of the state of the art of AI, and of the nature of intelligent behaviour.

What kind of presentation would serve better? One possible approach would be to present a deeper analysis of one or two parts of the field, and then to suggest by analogy how the rest of the field is similar. The chapter by Schank and Lehnert is a good example, giving an overview of the issues involved in natural language processing. In section 3 below I will consider another area -- the representation of knowledge -- and try to convey a sense of the state of our current understanding. Before turning to that discussion, however, I want to present an alternate, more theoretical, conception of AI as a whole.

2. AN ALTERNATE CONCEPTION OF ARTIFICIAL INTELLIGENCE

I define Artificial Intelligence as a study of the *structure of human intelligence* in terms of a formal and precise method of understanding processes called the *computational metaphor.* This characterization leads immediately to the following questions:

---

2. See the article by Schank and Lehnert elsewhere in this book.

3. Weizenbaum, Joseph: "On the Impact of the Computer on Society", The Technology and Culture Seminar of the Massachusetts Institute of Technology; Cambridge, Mass. February, 1972.

1. Why just *human* intelligence?
2. Why the *structure* of intelligence instead of intelligence as a whole?
3. Why the computational metaphor?

I will develop answers to each of these questions in the following discussion. The first step in constructing the argument is to examine the goals of AI research.

There are two fundamentally independent sets of goals in AI. One is concerned with the construction of sophisticated computer systems, designed to perform tasks which would require intelligence if performed by people. Such systems might help in industry, serve as personal assistants, entertain and instruct, etc. Although goals such as these are originally practical, they have a theoretical side as well. For one thing, in order to build better systems in the future, one must continually deepen one's understanding of basic mechanisms and techniques. In addition, the study of basic algorithms and methods for doing things can develop into a theoretical discipline of its own.

A second class of goals, of a more psychological and philosophical nature, relates to the general study of how the human mind works. From this perspective actual computer programs are of secondary interest; the primary purpose is to uncover the basic nature of people and of human rationality. Studies of this sort are commonly thought to be only theoretical, although in fact the ways in which people choose to understand themselves have tremendous practical import in society at large.

There are tensions that arise between the various practical and theoretical aspects of these different sets of goals. In building systems, for example, one generally strives for coverage and breadth; in theorizing, one focusses on a small area in order to distill its essence. Ideally, these tensions can be productive -- progress in one area can be carried over and found helpful in others. Confusion and competition, however, are also potential outcomes of such disparate interests. An important question to ask, therefore, is how the different endeavors can best come together to form a synergistic and cohesive whole.

The essential ingredient, I would argue, is for each individual component to be seen as distinct, and valued in a manner appropriate to its particular interests and goals. In the following discussion, therefore, I will attempt to carefully delineate the various enterprises, and identify their separate properties. In doing so, I use different names: by *cognitive science* I refer to the scientific study of the structure of human intelligence; by *knowledge engineering* I refer to both the theoretical and practical aspects of the production of sophisticated computational systems.[4] (By *artificial intelligence*[5] I will continue to refer to the more or less unified whole.)

With these distinctions in mind, we can turn to a consideration, within the general study of AI, of the role of human intelligence and the role of the computer. In particular, I will argue that human intelligence and computers have a different part to play, depending on whether one's primary interest is in knowledge engineering or cognitive science. In the course of this discussion, we will be led to the answers to our original three questions; in each case the answer will similarly depend on one's intellectual goals. In spite of these many differences, however, I will endeavor to show how a constructive relationship could be established between the dissimilar viewpoints.

Regarding human intelligence, several points come immediately to mind. First, we use our intelligence to define the area of study: that is, to define just what kinds of behaviour we consider to be examples of intelligence. Second, as with any discipline, we use our intelligence in carrying out our research. This is not simply to say that we "do the best we can", but that the investigation is inherently self-referential:[6] the same intelligence which is the object of study is simultaneously involved in deciding how to study it. We constantly apply our intellect, for example, in deciding what data to study, what counts as a theory, what counts as evidence, and which theories are better or more elegant than others. This will always be true, since we cannot escape our humanity and look at intelligence from an extra-human perspective.

There are also some reasons which have particular relevance for those engaged in knowledge engineering. Because humans are our most compelling (if not our only) model of intelligent behaviour, they serve as the best example from which we can learn. In addition, since we would like any systems we build to communicate with us, we would do well to endow them with a human kind of intelligence. It would be of very little use, for example, to build a medical

---

4. The DN&R article deals with AI purely as knowledge engineering, ignoring the cognitive science dimension completely. Any commentary on the state of the art of AI should of course include an account of current engineering; one might claim, given that the article is entitled "state of *technology*", that in their case nothing more should be expected. However, as I argued above, the engineering aspects of the field can be honestly assessed only from an understanding of the fledgling science on which they are based.

5. An unfortunate term, in my opinion, but one which I will use here for historical reasons.

6. And presumably, therefore, far from "objective".

assistance system whose diagnosis procedures were so radically different from those employed by a doctor that no communication could take place.

In addition to these fairly obvious points, there are others based on some deeper philosophical issues. A priori, there is no reason to suppose that the very idea of intelligence is not a uniquely human conception. Furthermore, even if there were other forms of intelligence, they might be radically unlike us, and there is no reason to assume that the human mind would be capable of understanding them. Someday, if we are ever able to model human intelligence successfully, we may be able to adjust the model's parameters or structures and see behaviour which differs from our own in certain aspects of its performance. It might, for example, solve problems in parallel, resembling human society more than a single person's mind. I very much doubt, however, that we will witness anything differing in *kind* from the intelligence of we who created it. For example, it seems implausible that we could construct a computational intelligence fit to perceive the universe in terms of a continuum of temporal dimensions, or one which did not happen to have embedded in it the arbitrary notion of a discrete object, since we cannot even imagine what either of those would mean. Similarly, our common-sense would probably be ill-suited for creatures a million times smaller than we are, or for mass-less beings accustomed to travelling at the speed of light. It is for this reason -- because of the inherent nature of things rather than for any lack of interest on our part -- that AI must be seen as the study of *human* intelligence.

I have identified a set of different roles played by human intelligence; they have different force depending on one's goals. Cognitive scientists, of course, study people by definition. As suggested by the practical reasons given above, knowledge engineers study people because it makes sense. And while the philosophical questions might make a difference if we knew how to answer them, such concerns have little effect on day-to-day research. In other words, while the underlying reasons are different, the practice is roughly the same. This fact is one of the reasons (as suggested above, and pointed out by DN&R and by Feldman in his comment) why the psychological and engineering perspectives on AI are seen to be mutually reinforcing.

Regarding the role played by "computers", we must first digress to distinguish the actual physical machines from the intellectual tools we have developed with which to understand them. The label "computer science" has been called a "patent misnomer": computer science is not the study of computers, but the study of formal processes. As we develop more sophisticated ways of understanding processes, we are able to build new physical embodiments of our theories; this is how more powerful computers have been developed. But underlying all computational theories is a consistent and basic methodology of understanding processes in terms of functional decomposition and the manipulation of symbols -- a methodology I call the *computational metaphor.* I will briefly examine some of the assumptions and consequences of this metaphor, before returning to consider its importance to the various endeavors of AI.

Fundamentally, the computational metaphor is a way of understanding processes in terms of the active interpretation of symbols by an "interpretive" process. The manner in which symbols are manipulated is based purely on their structure -- any meaning those symbols may have,[7] or any sense in which they may correspond to the world, is invisible to their interpreter. This is at the heart of the statement that computer science is "formal"; it is the form of symbols which influences their behaviour. Note that such a claim does not mean that symbols may not have significant internal structure, but merely that, again, it is the *form* of that internal structure that determines how the symbol will be interpreted.

This formality, inherent in the computational perspective, is what leads me to define AI as a study of the *structure* of human intelligence. In fact, any endeavor based on purely formal tools will necessarily study just the structure of its domain; one might for example suggest analogously that much of modern linguistics is a study of the structure of language, rather than being a study of language itself. This formal or structural approach is just one among many ways in which human thought might be studied.[8] For knowledge engineering, of course, understanding the structure of intelligence is the best route towards building computational systems. For cognitive science, the structure of intelligence is all that we will uncover because, given the glasses that we have chosen to wear, that is all that we can see.

We may now return to consider the general role of computation. For knowledge engineering, the computer is of course crucial, because the production of computational systems is the goal of that pursuit. Both actual computers, and the techniques with which we understand them (including the basic metaphors) are essential in this enterprise.

Regarding cognitive science, it is the metaphors and intellectual tools of computer science which have been proposed as relevant. That they might lead to insights about human cognition follows from the obvious fact that the human mind is both active and complex. It is therefore natural in studying it to apply tools which were specifically designed to deal with active, complex processes. Furthermore, people clearly think and communicate using symbols, and a metaphor that considers not only activity but symbol processing is an

7. Or, more precisely, any meaning that we may impute to them.

8. The formal style is so deeply embedded in the computational approach, however, that we are in danger of forgetting that there are others.

obvious candidate with which to understand such reasoning. Space does not permit considering this issue in more depth, but the cognitive science dimension of AI is founded on the hypothesis that this metaphor is indeed helpful in deepening our understanding of human cognitive processes.

It might seem, from this account, that practical computer programs would be of no interest to cognitive science. Certainly computational systems are not the products of scientific research; to the extent that research should be said to have products at all, theories or ideas would be the appropriate candidates. It turns out, however, that because the human mind is limited and imprecise, researchers find it useful to construct computer models of the process theories they are investigating. Computers can serve as testbeds for working out new ideas and for demonstrating their logical integrity. They are *active scratchpads*, much as sheets of paper are static scratchpads for mathematicians.

(It is essential in passing to dispell a common confusion about such computational models of cognitive theories, or indeed of theories of any sort. One cannot implement a whole theory on a computer; a theory -- in particular, a formal theory -- consists of a model, a world, and a statement of correspondence between the model and the world, of which only the first can be embodied in a computer program. The analogy with the mathematician's sheets of paper can again serve to illustrate this point. One can no more embed a theory on a computer than one can put a theorem onto a piece of paper. In each case what is encoded -- either written down or programmed in -- is a set of symbols which are to be understood as signifying the theory in question.)

Thus both scientists and engineers construct programs: once again, while their purposes are different, their behaviour is similar. This is another reason, then, for a potential symbiosis between AI as cognitive science and AI as knowledge engineering. Such a relationship also explains the historical compatibility between the largely practical goals of those who fund AI research (funders are by and large interested in products) and the goals of researchers who are interested in AI as science. The difference in emphasis, however, makes sense out of DN&R's odd statement that the goals of AI research are different from the goals of people who carry out that research.

As cognitive science, AI is in its infancy. Most of the work to date has been from the engineering perspective, and even in terms of theoretical engineering, much of it has been ad hoc and un-scientific. This is partly because AI is a young paradigm, and has yet to define its conception of a theory, its criteria of elegance, its standards of rigour, etc. As the intellectual framework becomes better defined, work within it will undoubtedly become more disciplined. An understanding of this "just-emerging" sense of AI (and its cognitive science dimensions in particular) is essential, not only to understand what has been done in the field so far, but also to envisage what may be accomplished in the future.

3. THE REPRESENTATION OF KNOWLEDGE

In this third section I turn my attention to a specific sub-discipline within AI, in an effort to characterize the current state of our understanding of the issues it raises. My goal is to illustrate the depth of the problems involved, and to convey a sense both of what progress we have made, and of how far we have yet to go.

In the early days of the field, there were hopes that some general theory of intelligence could be shown to underlie a large fraction of intelligent behaviour. People envisaged a general algorithm, capable of behaving intelligently without the need of special purpose advice for each new problem and each new subject area. However this early approach quickly bogged down in a morass of difficulties, and people began to appreciate the importance of domain-specific knowledge. For example, in order to translate an article on meteorology from English to Russian, a machine must know a substantial amount of meteorology. In order to walk around a building, a robot must know about rooms and doors. The chapter by Schank and Lehnert explores the relevance of appropriate knowledge in natural language processing; the situation is similar in many of the other areas of AI.

Much attention, therefore, has been focussed on how to embed knowledge in a computational model. It turns out that this is far from simple: it is difficult to figure out what knowledge is relevant and to decide how to organize it. Some issues seem to be specific to particular problem areas, but many appear again and again in different contexts. Many of the most serious difficulties stem from the fact that the task is not simply to capture the knowledge in a computational formalism, but to do so in such a way that a process can reason intelligently with it. Ultimately, it is active, intelligent reasoning which is primary; all questions of representation are secondary goals that must fit into the larger concern.

For example, the representational formalism must encode knowledge in such a way that the appropriate facts can be accessed quickly in appropriate situations. Care must therefore be taken to establish explicit structural relationships between the representations of different facts, which may be different from the encoding of the logical connections between them. In addition -- for deep reasons of generality, flexibility, and economy -- only a small subset of all possible facts can be stored: formalisms must be provided to encode general principles, and routines built to access these appropriately and to deduce their obvious consequences.

One of the goals of knowledge representation research has been to provide a formalism which would be suitable for representing knowledge in many different domains. Classical logic is an obvious candidate: it has grown out of philosophical and mathematical traditions that have asked questions similar to those asked in AI, it has a well-developed theory of semantics, and it benefits greatly from being more thoroughly understood than any other system that has been proposed. In fact there is no doubt that logic has no serious competition, *with respect to the issues which it faces.*

Nevertheless, there are many perspectives on knowledge representation which are essential to AI, but which logic and predicate calculus simply do not address. For example, since computational models are inherently active, it is important to know how long it takes a program to do something. Consequently (as noted above) one needs to know how the knowledge is organized, and how concepts are accessible one from another. Logic has nothing to say about this kind of memory organization. In addition, people in AI are in general interested in the interactions between propositions and reasoning processes; again, logic provides no way in which to even ask this class of question.

Frustrated by these shortcomings, many researchers have designed their own representation systems. Some of the first schemes dealt almost solely with how concepts are accessible from other concepts, in a network-like scheme. Later suggestions combined this kind of organizational idea with notions of properties, as suggested by the propositional calculus. The property lists of LISP, and various extensions of that idea, have been used to experiment with this kind of representation. In such a scheme, indexed under the concept "CANOE" one might find indicators of several of its properties, such as being a kind of boat, or weighing about 70 pounds.

Out of work on the properties of objects came a suggestion which plays a significant role in much of the current research in knowledge representation. The idea was based on the obvious observation that most objects in the world are not atomic, but instead have an internal composition. This is true both of physical things (a table has a top, legs, etc.) and conceptual abstractions (a marriage involves a husband, a wife, and perhaps children; an act of giving involves a giver, a gift, a recipient, etc.). The suggestion, then, was to organize the facts about such objects in such a way as to mirror the structure that we perceive objects to have in the real world. For example, if a piece of sculpture had a central region and three long appendages, a formal representation of it might be constructed with four internal parts -- one for the main body and one for each limb.

These formal objects lead to a powerful way to represent abstractions. One can construct an abstract prototypical object, which stands for all the objects in some class, and then describe individuals in terms of that prototype. General properties can then be inherited automatically from more general concepts to more specific ones. For example you might represent "FIDO" with a simple link indicating that it is an instance of the concept called "DOG". Then, if the system knew that the prototypical dog has four legs and tends to bark, it could conclude that Fido has four legs and can probably bark. Similarly, if it knew that dogs are animals, and that animals breathe, it could conclude that Fido breathes. All of these facts could be found directly by following links "upwards" from "FIDO".

Many subtleties arise in representing such abstract objects, only a few of which are addressed by current representation systems. For example:

1. It is sometimes powerful to describe a single individual in terms of more than one abstraction. For example, you might want to describe Randy as "A LAWYER" and also as "A WOMAN". Providing for this kind of *multiple description* allows great power and flexibility, although it introduces complexities in keeping track of inheritance paths. For example, if you want to know Randy's sex, you would have the choice of looking for the answer under "LAWYER" and under "WOMAN". We people know that "WOMAN" is the more appropriate abstraction, but it is not always easy to control a program so that it does the sensible thing.

2. Sometimes properties of abstract classes are not held by all known instances. Birds in general fly, although penguins are birds which don't. Representations of many concepts need to include *default* information that is taken to be true of any instance, unless there is specific reason to believe otherwise.

3. Different individuals sometimes share a common subpart. A typical living room and a typical dining room, for example, may share a door between them. If a living room and a dining room are separately represented, the system must know that the door between them is shared.

As mentioned above, as well as representing objects and concepts, people are interested in specifying how programs should think about these objects. Issues of control structure are related to issues of knowledge representation, since one often wants to associate with a particular concept a special-purpose program to perform some specific task. For example:

4. Suppose you wanted to check, each time you found out where someone lived, whether you knew anyone else from that town. You might want to associate with the general representation of a person's address a special program to perform the check. Your intent might be to have this program run quickly whenever a new address was discovered, no matter what the more global reasoning context might be. Some representation schemes provide for this kind of *procedural attachment*, but no one yet knows how to use such facilities in powerful but structured ways.

While we do not yet understand all the subtleties regarding structured objects, many systems deal with some of them, and there may be some convergence within the next few years. However there are many more difficult representational issues, which have at best been explored in research systems, but about which there is no general consensus. For example: Continuous objects (such as mashed potatoes or water or strings or continuous expansion) are notoriously difficult to represent and reason about. One possible explanation may be that computational symbols -- at least any we are familiar with -- are themselves inherently discrete, but this is not well understood. Other open topics include hypothetical situations, models of other people's knowledge, time, states and events (somehow anything that seems to be an event can turn into a process if you look at it closely enough), and complex quantificational descriptions (such as: "Most of the people who went on the Sierra Club trip seemed to know at least several of the other members.").

More importantly, we have only the most tentative ideas about how people reason with such knowledge. Researchers are currently exploring theories of instantiation, reasoning by example, reasoning by analogy, comparison in terms of mutually exclusive categories, reasoning by consideration of a set of alternative hypotheses, etc.

None of these issues is very clearly understood. In the final analysis, however, understanding how people reason is of more importance than simply representing the world that they think about; the task of representing knowledge is subsidiary to the primary goal of representing *knowing*. It is striking to observe that the word "knowledge" is a noun derived from the verb "know"; that "meaning" comes from the verb "mean"; that "thought" comes from the verb "think"; and even that "reason" was a verb before it was a noun. Intelligence, after all, is first and foremost an activity. The process models of computer science may be helpful in uncovering its essence, but coming to grips with that activity is still our most elusive goal.

4. CONCLUSION

My purpose in this comment is not to be pessimistic; I would merely argue that the achievements of AI should be viewed with some perspective, and its goals with some respect. The human mind and the complexity and subtlety of human intelligence are awesome, and the fact that we do not yet come close to understanding them is more a reflection of their power than of our failings.

Many thanks to Barbara Grosz, David Levy, Mitch Marcus, and Terry Winograd for their help, both in the preparation of this note, and in the continuing development of the ideas.