AI Magazine Volume 5 Number 4 (1984) (© AAAI)

The Professor's Challenge

Pierre Bierre
Clairvoyant Systems, Mountain View, CA 94043

Abstract

The AI field needs major breakthroughs in its thinking to achieve continuous, sensory-gathered, machine learning from the environment on unlimited subjects. The way to motivate such dramatic progress is to articulate and endorse research goals for machine behavior so ambitious that limited-domain, problem-solving knowledge representation methods are disqualified at the outset, thus forcing ourselves to produce valuable new "thoughtware". After exploring why the tendency to associate intelligence with problem-solving may be a mental roadblock to further progress in AI science, some preliminary thinking tools are introduced more suitable for sensory learning machine research These include lifelong sensorimotor data streams, representation as a symbolic recording process, knowledge transmission, and the totality of knowledge.

One day, Professor Nokemoff, a distinguished AI professor of robotics at a leading institution, called in a grad student to hand out a thesis project. "Build me a robot that can sort two different parts coming down a conveyor belt. Give me a call when you're ready for a demonstration." Six months later, the student called the professor down to the lab, and demonstrated the requested skill by having the eye-hand system look at each part, then lift it and place it on one of two respective pallets. The student received his Ph.D.

Next semester, the professor called in another grad student looking for a thesis topic.

"Build me a robot that can grasp a part out of a jumbled bin of identical parts. When you get something working, give me a call." Nine months later, the student called the professor down to the lab and, sure enough, the robot repeatedly dipped its hand into the bin and each time came up with exactly one part. The student received his Ph.D and a very lucrative job in industry.

The following semester, the professor called in still another grad student and announced this thesis project:

"Build me a robot that can completely assemble a Sears 10-speed bicycle. When you're ready to show me something, give me a call." Undaunted by the challenge, and supported by a generous equipment budget, the student pressed forward and, one year later, called the professor down to the lab to witness a complicated three-hour procedure where, starting with all the parts and assembly tools neatly laid out on a table, the bike was put together on a bike stand piece by piece to completion. The student received his Ph.D and both he and the robot went on to lucrative jobs in industry.

The next semester rolled around, and Professor Nokemoff, reflecting on his students' past string of noteworthy successes, was beginning to bristle slightly at the thought of giving out projects that were too easy. It just so happened that around this same time a new grad student, a real eager beaver, was looking for a thesis topic. Smelling an opportunity, the professor called him in.

"Build me a robot that can ride a bicycle across town and back, go down to Motor Vehicles and obtain a California driver's license, play six innings of baseball, cook a gourmet dinner, and finish up with a rendition of Stephen Foster's 'Oh Susannah' played on piano." The student, taken aback but not totally deflated as he got up and headed for the door, offered sarcastically "Anything else?" "Oh..." the professor shot back, "and when you think you're finished, don't call me—have the robot come tell me in person."

In the preceding story, the professor's frustration stemmed from a gut feeling, one shared by a growing number of Nokemoff's nonfictional colleagues, that even after finding solutions to the parts-bin problem and light assembly tasks

Clairvoyant Systems is a new AI research group dedicated to long-term progress in continuous sensory learning, and is currently seeking team members including patrons, scientists, managers, and students. Send inquiries and comments to 1921 Rock St , Suite 17, Mountain View, CA 94043

Illustrations by Gordon S Bernstein, Sondra Bierre and author on the order of a 10-speed bike, or hundreds of other similar problems that subsume a well-defined goal in a narrowly constrained environment, we may emerge no better in terms of understanding how intelligence works.

Indeed, the operative paradigm in experimental AI of selecting a domain and building a system that solves a problem within it may have nearly outlived its usefulness, from the standpoint of gathering new scientific knowledge about intelligence. This article begins with the professor's challenge, and moves on to examine why the "problem-solving" paradigm may be a roadblock to further progress in AI. It suggests an alternative paradigm whose emphasis is on machine versatility, or the ability to do a great number of tasks in a great number of environments. Finally, it proposes some specific conceptual tools with which to embark into this new, uncharted territory.

The Challenge of Versatility

The thesis project the professor hands out at the end of the story roughly articulates the next major challenge facing the field of AI research —versatile behavior in unconstrained environments. The professor's intent was not merely to have his student program five separate "experts" at bicycle riding, automobile driving, baseball playing, gourmet cooking, and piano playing, all packaged into one device. Were his student to head off in this direction, the professor would merely lengthen the list (as he nonchalantly does in the story punchline), or better yet, impose increasingly vague requirements such as,

> "I really meant that I want the robot to be able to do complex sensory-motor tasks like ride a bike and drive a car. I'll tell you the exact tasks I want to see performed the day of your final project review."

The professor's intent in naming five or six areas of competence instead of one was clearly to challenge his student to come up with a unified approach that can underlie the attainment of such diverse skills.

Another aspect of Nokemoff's previous students' projects that bothered him was the degree to which his students were redesigning the robot's work environment to make the problem easier. For example, in the parts-sorting problem, the conveyor belt was painted white in order to exploit a silhouette-imaging technique. It kept getting greasy and had to be wiped off frequently by the experimenters lest the machine become confused. The parts-bin problem was solved using some esoteric "structured" light sources, and the bicycle parts had to be laid out just so or else the robot would quit in midstream After a few of these experiences, the professor realized that he could continue indefinitely giving out Ph.D's for laboratory demonstrations of single-task behavior, but that without the additional requirement that the task be performable with robustness in the face of environmental variety, not much was being learned about intelligent robotics.

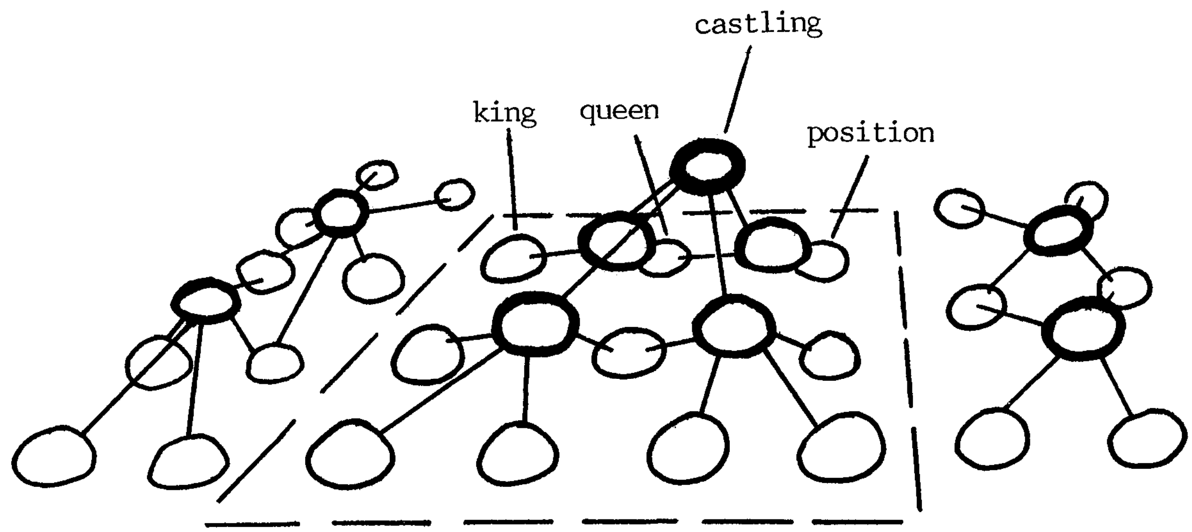
Thus, the challenge given to the ambitious student takes the robot out of the laboratory and puts it on the street. Passing a driving test for a California license denies the experimenter the luxury of painting a special stripe on the road in advance. The route will remain unknown until the last minute or seconds during the test, and the robot will have to communicate naturally with the DMV inspector to receive directions. You cannot seriously entertain the notion of patching together the best speech recognition, natural language, voice synthesis, image understanding, and rule-based planning AI programs to achieve this amazing feat. Something much more integrated is called for. Something much more elegant.

It was on this gut feeling that the professor gave out the versatility problem. By making the project so outrageously ambitious, he hoped to force his student to quickly abandon existing methods and to invent some new ones better suited to a more holistic, robust machine intelligence. And he knew that, should these new methods emerge, they would certainly be worth more than a Ph.D and a well-paying job in industry. More likely than not, they would earn their creator(s) a Nobel Prize in Artificial Intelligence.

The Problem With Problem Solving

The idea of studying isolated instances of behavior, perception, and language, as an alternative to the more intimidating and complex issue of intelligent behavior in its totality, is firmly entrenched. The overwhelming adherence to this reductionistic strategy throughout experimental psychology and AI accounts for the fact that more and more papers can be published, Ph.D's awarded, and conferences held with each passing year without the allied cognitive science community progressing at a concomitant frenzy towards a concensus understanding of how intelligence works. Despite this effort to break the complex problem down into simpler pieces, and despite notable success with the smaller pieces (*i.e.*, expert systems), fundamental aspects of intelligence remain unsolved. One of the more persistent of these, ironically, is the question of how the brain chops a continuous stream of input sensation and output behavior apart into discrete phenomena. This is the segmentation problem. In order for these reductionistically oriented researchers to be able to discern subproblems of intelligence and isolate variables, they more or less have to take their own internal segmentation process for granted It makes them see the pieces.

The legacy of reductionism for AI has been a methodology that concentrates on constrained problem "spaces." While impressive accomplishments have come out of this research paradigm, problem-solving methods seem to be strapped when it comes to the challenge of open-ended, versatile intelligence such as the professor has posed. In lieu of being able to explain clearly why this is so at this stage of the game, some clues can be offered.



Chess World.

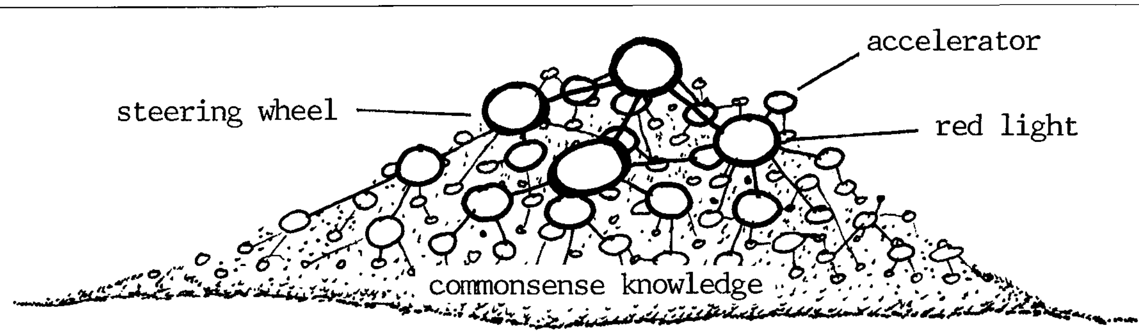
Figure 1.

Clue #1:
Prerequisite Knowledge and Time Order of Knowledge Acquisition.

The knowledge representation problem concerns the question of how to come up with mapping from the external world to a discrete symbol system inside the computer. It is relatively easy to define all possible states of the world, for example, when the world consists of a game of chess, by coining abstract symbols for all the "pieces" of this world *i.e.*, the players, the chessmen, board positions, legal moves (expressed in terms of the preceding symbols), and the object of the game. It is hard to carry this same approach over into the world of automobile driving. The difficulty isn't so much in coming up with discrete symbols that apply (*e.g.*, steering wheel, accelerator, intersection, and red light). The difficulty is in circumscribing all of them. The chess world has nice clean boundaries at the edge of its symbol system, as shown in Figure 1, whereas in the driving problem, it seems that the symbol system peters out fuzzily into an enormous underlying pyramid of commonsense knowledge (see Figure 2).

The problem with trying to isolate driving-world knowledge can be explained as a violation of the principle of prerequisite knowledge, or the notion that the time order best suited for acquiring symbols into a knowledge base requires symbols written "on top of" other symbols to come later than them. It seems that driving knowledge would be a lot easier to "write" after the commonsense knowledge is already in place, because the driving symbols want to be written borrowing symbols from the commonsense knowledge.

Let's go back to the chess world for a moment. What order makes sense for putting new symbols into the chess knowledge? Let's suppose you choose to begin your chess knowledge base by creating a symbol standing for castling. Some of the elements you wish to associate under castling are when it's legal, when it's advisable, and which pieces move where. You decide to tackle the when-castling-is-advisable



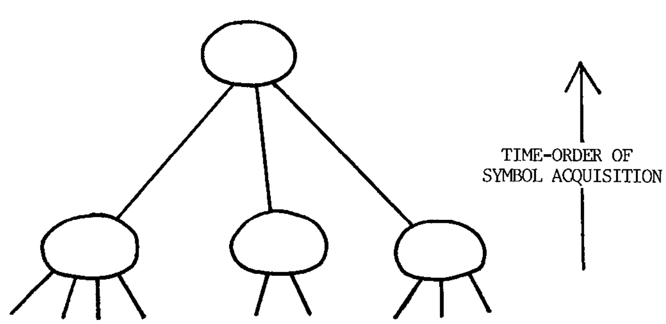
Driving World.

Figure 2.

branch first, inserting placeholding stubs for the other two symbol branches. The problem is that castling is not possible to express until symbols for such basic elements as king, queen, and board position are available for use in constructing its definition. It makes more sense to start with symbols for board positions and the chessmen, as these symbols are general purpose with respect to symbols that will come later. King will be referenced in defining castling, check, and checkmate, for example. In knowledge acquisition, it is easiest to acquire a symbol when all the referents needed to tie down its meaning are already in place.

A heuristic programmer working on the driving problem would probably start with obvious symbols like accelerator, red light, steering wheel, and only after digging very deeply into the subject would come across obscure but necessary symbols for elements like ball-rolling-into-street-in-front-of-car and gasoline-spills-over-while-filling-tank. When confronted with adding these latter symbols to the knowledge base, the programmer is in the same fix as he'd be in with the chess world if he tried to define castling before king, queen, and position. The necessary prerequisite symbols just aren't there.

The alternative time order of knowledge acquisition alleviates the problem. If prerequisite symbols for kids-playing and ball-rolls-into-street already exist, a symbol relating such as a driving hazard to the apply-brake response is easy to incorporate. The knowledge about kids playing with a ball is general-purpose: it can be used as an ingredient for defining any symbol that comes into the knowledge base after it is in place. Figure 3 summarizes the critical time order of symbol acquisition due to the principle of prerequisite knowledge.



Prerequisite Knowledge.

Figure 3.

Simply stated, heuristic programming won't work for building a car-driving robot because it approaches knowledge acquisition in the wrong time order. Merely to think of "driving knowledge" as a domain starts the researcher off on the wrong foot. Driving knowledge needs to be written on top of, and after, commonsense knowledge gathered in childhood, and this prerequisite knowledge is good for learning many tasks besides driving. It is just sitting there for subsequent use in acquiring any new symbols. Similarly, once driving symbols have been incorporated, they too become available for general-purpose use in subsequent knowledge acquisition. From this vantage, it doesn't make sense to divide knowledge into categories such as chess, or driving, or commonsense—better to look at it as a holistic fabric that serves as raw material for its ongoing expansion.

So, the professor's student already has a solid clue about machine versatility: Forget about knowledge domains, and concentrate instead on knowledge acquisition into a holistic symbol system, being careful to swim with the current of prerequisite knowledge rather than against it. In other words, collect knowledge so you can collect some more!

Clue #2:
Problems Whose Goals Are Ongoing.

Unfortunately, we cannot approach the problem of general purpose experiential knowledge acquisition as a "problem" to be "solved" in the classical sense of formal problem-solving methods (*i.e.*, theorem proving). It is more to be "worked on"—the goal is to maintain an ongoing process, not to attain a final state.

Several intriguing problems are extremely pertinent to intelligence with this flavor. An example of such a process problem:

> Organize a symbolic recording of an ongoing stream of fly-by sensory data, on the fly, such that at any given time as much as possible can be quickly remembered of the entire stream.

This is the sensory-learning problem. In the formulation of this research goal, note that no distinction exists between visual, auditory, or tactile data—they are all funneled into the same stream before processing. An extension of the sensory-learning problem is the sensory-motor learning problem, where the stream of output effector data is poured in with the sensor stream to form the thing that must be committed to a compressed, symbolic recording. The traditional viewpoint would ask, "What's the purpose of organizing all this sensory-motor data—*i.e.*, what do you want to accomplish using it?" There are two answers, neither very satisfying to the researcher acclimated to expert-systems thinking: You need this symbolic heap to support a whole host of worldly tasks too numerous to mention (and too versatile to know in advance), and you need it to plow back into the process of recording future sensory-motor data symbols.

If our analysis is correct that the car-driving problem will be easier if built on top of a base of commonsense knowledge good for other tasks besides driving, then this suggests that we might make progress in AI by backing away from a preoccupation with domain-specific problem solving, and concentrate on the problem of acquiring and organizing knowledge for its own sake. This is the thrust of sensory learning research.

Some other problems that share the same, nonformal flavor as the sensory learning problem are as follows:

> Construct a robot whose repertoire of skills is continually expandable in the direction desired by its human master, and that communicates naturally through its sensory-motor peripherals.
>
> Design a user-machine interface that consistently makes good use of its user's past experience in communicating information.

These are open-ended problems, never completed, which nonetheless must solve a multitude of diverse worldly problems, as innocuous side effects of the overall mandate, along the way. The difference between this, and artificial intelligence conceived to solve problems as we've used in the past, is mostly one of higher expectations. A final example of a process-problem: Imitate a Human.

This is Alan Turing's well known formulation of the goal of artificial intelligence research (Turing, 1950). Turing's test calls for the kind of versatility captured in the professor's challenge, as opposed to the isolated islands of problem-solving ability we now have in medical diagnosis and geological survey programs. The goal of the Turing Test is ongoing, and it cannot be neatly expressed in a closed formalism.

The second clue, then, is to begin looking at intelligence not as the ability to solve problems, but as the ability to learn a continuing thread of new thoughts.

Clue #3: The Chicken and the Egg Problem in Vision.

Which came first, the world or the image? This question pops up often in computer vision research. If we read a viewpoint into such titles as "Recovering Depth Information from Illumination" and "Shape from Shading," we find the more prevalent viewpoint, which I call

> The CHICKEN: The physical world behaves in accordance with established optical laws, such that the light field collected on the retina can be explained in terms of external considerations such as an object's position, dimensions, shape, surface reflectance, and the colorimetric makeup of light sources and relative positions of lights, objects and camera. So, in terms of cause and effect, the world precedes the image it casts.

From the point of view of this belief system, the goal of vision is to work backwards from the image, computationally speaking, to such a physical description of the world. One of the embarrassing questions one can ask the CHICKEN worshipper is "Do you mean to say that, no matter what I point the camera toward, physics can completely explain why the image is exactly the way it is? What if I point it toward an ocean scene with billowing cumulus clouds hovering over giant wave crests that glisten in the sun and break into whitecaps near the shoreline? What physical parameters and equations determine the image intensities in that scene?"

Clearly, physics models of image formation apply only to idealized worlds. For example, the Lambertian reflectance model was initially worked out for a point source of monochromatic light shining on a white spherical object of uniform surface reflectance against a backdrop of dark, empty space. This is a far cry from the busy street scenes that will greet the professor's student's robot when it goes out for a drive or a bike ride. Why haven't physicists attacked the problem of trying to explain the light levels in a busy street scene in terms of physical parameters? It doesn't make much sense to assume that there is a physical description for an arbitrary street scene. Thus, when you get right down to it, Chicken-think is founded on lessons bounded by extreme environmental simplicity. One senses that these lessons have been the victim of overzealous generalization in the name of religion.

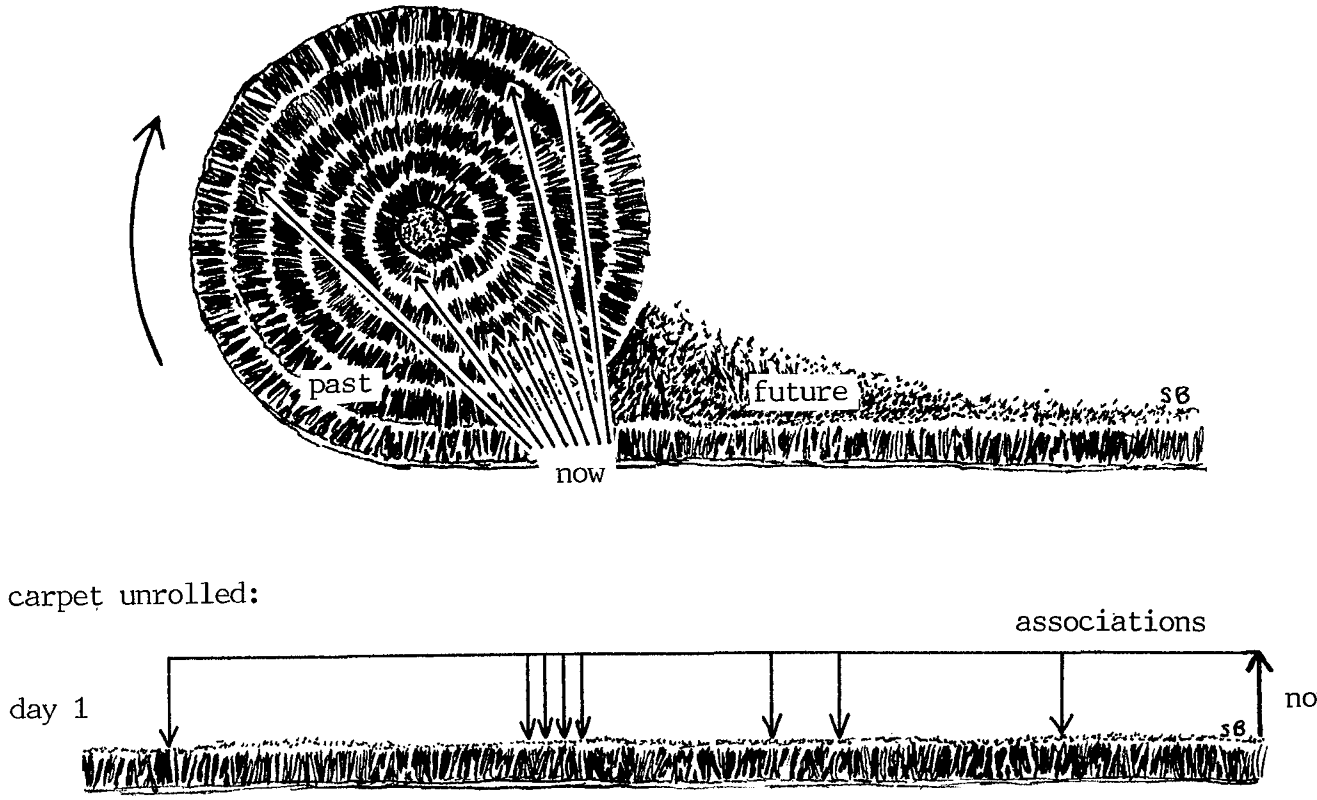
Let's look at the other side of the coin, the EGG position:

> The EGG: No ontological, physical world exists independent of the observer The observer collects a sensory data stream and commits it to memory on the fly, and the result is an expanding symbol system that constitutes the world of the observer. In terms of cause and effect, the sensory data stream is the cause, and world model formation is the effect.

The best part about this belief system is that it applies universally to whatever you put in front of the camera, whether ocean landscapes, busy street scenes, or white balls hanging in a black void. This is a distinct advantage if the objective is vision versatility.

I suspect that the difference between "isolated world" machine vision such as we have now and the versatile kind the professor has asked for is intimately hinged to the direction we lean on the chicken and the egg matter. That is, it's a matter of deep religious conviction. I took a B.S. in physics, and was properly inculcated as a young man in physics-world religion. I took ten years to convert to the other viewpoint. We're fortunate to live in an age of great freedom. I'm basically pragmatic: CHICKEN has taken us through this far, and EGG can take us much further in the future. If you've been wanting to try your hand at tolerating ambiguity, here's a chance to straddle two complementary, deep-seated belief systems.

I don't feel enough has been done with the EGG religion yet to really compare it with the hard physics viewpoint. It is encouraging to see awakening interest in discovery of structure and "nonaccidentalness" as a unifying theme in vision (Witkin and Tenenbaum, 1983; Lowe and Binford, 1983). From an early vantage, it seems that Egg-think offers a world view better able to explain knowledge acquisition than physics. In Egg-think, the world is only the totality of what has become known so far by each individual. The physics world exists quite independently of either the sensory-gathering or knowing process, and fails to account for most of the richness, complexity, and change we know exist.



Rolling-up-a-Carpet-on-a-Spool Metaphor of Intelligence.

Figure 4.

Equipped with the third clue, the professor's protégé will be skeptical of processing ideas that work well with certain inputs but fall short of full generality. When somebody tries to pass off a physical world model on him, he will ask for a physical account of why an individual who visits the exact same place twice will learn more about it on his second visit.

Clue #4:
Context in Natural Language.

Just like their counterparts in robotics and vision research, speech recognition and natural language [NL] investigators tend to focus on specific "problems." Ask these researchers if what they're interested in is an understanding of how all instances of language communication work, and they will say the problem is too hard—come back in 25 to 50 years. For now, we're interested in limited-domain discourse.

The parallels between the stance in NL, robotics, vision, and speech are striking. In each an overwhelming tendency has been to treat the input as objective, or having basically the same meaning regardless of whom it is input to. Since long before Noam Chomsky, I think people interested in understanding how language communication works have secretly hoped that the meaning of words and speech utterances could be reliably assigned without having to drag into the picture the entire experience of the receiver from day one But the utterance between two old buddies:

"Do you remember the time we poured vodka down Jimmy's drawers?"

and its immediate evoking of recall suggest that language can connect with potentially anything in the experience stream of the listener. Thus, the context that must be considered to assign meaning to this utterance, far from being something that can be found by backing up a few sentences, is everything that has ever happened to the listener. Moreover, the above utterance, fed into another listener, will probably evoke a confused "Huh?" That is, the information value is entirely subjective, depending on the individual listener's previous experience. Nils Nilsson's proposal to work on a computer individual with continuing existence (Nilsson, 1983) signals a move in the right direction. Figure 4 shows a metaphor of subjective, experiential intelligence as a carpet being rolled up on a spool.

Limited-domain discourse tends to reinforce the opposite intuition—namely that utterances have objective meanings, and the job of the NL system is to figure them out. If you recall from the discussion of how difficult it would be to program a car-driving robot heuristically, the problem was traced to lack of prerequisite knowledge. The natural language researcher who tries to isolate the subset of language knowledge needed to discuss medical diagnosis only has wandered into the same forbidding territory. If doctors are to be allowed to say things to the computer like "His liver is the size of a football," or "Order me about a mile of suture, this guy's going to need it," then somewhere along the line NL researchers are going to have to give up the idea that medical knowledge and medical language knowledge can be cleanly dissected from any other knowledge. Marv Minsky's Webs of Meaning concept (Minsky, 1983), the idea that the meaning of something lies in the way it connects to all other things we know, represents a shift toward a more holistic view of knowledge.

The fourth clue, then, is that context is not to be feared, but a vital resource that bestows intelligence. Efforts to limit context to the immediate past, or to a particular subject matter, are entirely counterproductive.

Clue #5: AI Science Needs a Totality of Knowledge Concept.

Another aspect of NL research that fits a "pattern" with vision, robotics, and speech research is that no suitable definition is given to exhaustively delimit the scope of language communication. It is easy to point out examples of purposive communication not covered by natural language studies, such as body language, nonspellable vocal sounds, vocal intonation, showing a child how to tie his shoes, and interpersonal touching. There is no attempt in NL research to throw a lasso around the subject of investigation. Until Dmitri Mendelyeev came along with his chart of the elements, chemistry was in a similar position. As a science, AI would be on a much sounder footing if for instance, language researchers had a way of circumscribing, in a nutshell, all that is communicable, or if vision scientists had a statement of all that is viewable. But, a considerable overlap region exists between these two; what would be better is an umbrella covering everything, a Totality of Knowledge concept. Working from such a unifying platform, the efforts of individual researchers and laboratories would be more likely to have wide application elsewhere.

Problem-solving AI does not strive for a totality concept. Indeed, in order to be judged as valid research, proposals must "adequately constrain" the range of effects, inputs, environments, behaviors, and goals that will be studied. If we look at the terms used to express limits of research interest, say in speech recognition research, a certain number of words, spoken by a number of speakers, in a limited subject category is typical. The question of what constitutes the totality of all words never enters in, nor does the question of the totality of subjects one can speak on. These troublespots are avoided.

The corresponding difficulty in computer vision surfaces as the problem of defining all the objects that can be viewed in the universe. Not only would it be impossible to exhaustively describe the totality of objects, but such an imaginary category "leaks" as a totality concept when it comes to nonphysical, viewable elements like rainbows, wrinkles in clothing, images projected on a screen, a person scratching his forehead, and the list goes on and on.

Natural language people have their words, sentences, and subject domains. Vision people have their objects and images. Signal interpretation people have their signals and emitting sources. Nobody in AI has put a finger on the whole ball of wax, nor does it seem necessary to do so as long as limited-world problem-solving is the stated objective of each AI project. That's how the professor's challenge blows past AI thinking out of the water! It forces you to look at the whole ball of wax! At the Totality of Knowledge! That's clue #5.

Clue #6: Segmenting Totality.

Natural language, speech recognition, and vision workers make the cognition problem inordinately easier (in the short run!) by seeing to it that the computer's input already comes "chopped up." For instance, NL researchers tend to conceptualize language in terms of sentences, and are willing to program this concept into the machine at a very low, mechanical level, by giving punctuation marks a terminator status, or by signaling sentence completion with a carriage return. Story understanding systems are "told" a new story is starting with a formal language token. In this manner, the experimenter pre-segments the AI program's lifetime input, in a sense dictating, a priori, which parts of the input have more to do with each other, and which parts have less to do with each other.

Similarly, in image understanding, the camera is ordinarily given a still "scene" to analyze or, in the case of motion vision, a short segment that brackets the action under study. In the lifetime of such systems, many input samples, or problems, are received, and it is not normally expected that the image understanding process decide for itself when an old problem ends and a new one begins. The experimenter makes this decision.

Segmentation is an area where a clear line separates existing machine intelligence and human intelligence. The human can do it, and he must do it in order to conduct limited-domain, problem-solving AI research. Otherwise he wouldn't be able to articulate the domain of specialization and the primitive objects considered essential to build knowledge on top of—the axiomization stage of knowledge engineering.

In contemplating the professor's challenge, I doubt if anyone would particularly care to axiomatize all the features and objects, parameters, and slots that the robot will need to go down to the DMV and pass a driving test. We need to offload the segmentation problem (What goes with what? And what is left unconnected after deciding what goes with what?) onto the computer. We need to work on machine segmentation of totality, and that's clue #6.

Before this challenge can be undertaken, we need the totality statement talked about earlier, so we can clearly define just what that totality is that needs to be segmented. I will propose such a definition shortly.

Clue #7: Intercommunicability and Knowledge Transmission.

Part of the professor's challenge rests on the notion that intelligence is much more about being richly connected with the outside world and all its myriad sources of intelligence, than with shuffling around symbols to solve a problem in a closed logical system. The proposition that we move beyond isolated islands of intelligence toward a more versatile, "open systems" brand of smarts can be interpreted as a direct mandate to incorporate stronger concepts for knowledge acquisiton and knowledge transmission (machine-machine communication) into the next "generation" AI paradigm.

It appears that logic and deductive inference as a theoretical foundation has reached an asymptote in terms of the degree of intelligence that can be achieved. By no coincidence, logic theory offers nothing in the way of insight about knowledge acquisition and transmission. The very foundation of logic, factual propositions such as

Man Is a Biped. John is a Man.

are tainted by the quality of being observerless, or independent of any knowledge-gathering process. They should remind us of Lambert's white ball, whose reflected intensities exist independently of whether or not they happen to be collected on a camera surface. Both are instances of what I have come to recognize over the years as an unspoken, unwritten, yet thoroughly institutionalized "style" of formalizing knowledge, where knowledge is stripped of its origin.

Back to the issue of knowledge representation—the job of mapping worldly things into coded symbols in the computer: If, as has been the case up until now, this job is performed by the AI experimenter, then to a certain extent the knowledge in the AI program is tied indirectly to the outside world through the eyes, ears, and grey matter of its programmer. To say that Bacon (Langley, Bradshaw and Simon, 1981) *discovered* Kepler's Law overlooks the simple verities that (1) Bacon could never have done it if Langley *et al.* had never heard of the planets, the Copernican model, orbits, time measurement, and the like, and (2) Bacon didn't announce its discovery to the world, Langley did. This reminds me a little bit of giving a difficult chemistry problem to a chemistry student with a pocket calculator, and then when he solves it, giving credit to the calculator for knowing chemistry. Admittedly, this analogy is a bit extreme, but it does contain the seeds of some truth. Lenat and Brown have acknowledged the role they as scientists played in attributing "meaning" to machine-discovered symbols, in their insightful and candid retrospective on AM (Lenat & Brown, 1983). So long as a human is needed as a go-between connecting the AI program's terminal symbols (*i.e.*, literal atoms) and the outside world *i.e.*, so long as representation is tied off manually, knowledge acquisition and transmission will resist scientific explanation, and the role of the AI experimenter's KA and KT ability in making his computer look smart will have to be subtracted from the credit given the machine. More importantly, if knowledge acquisition and transmission can be automated, it will be fast and inexpensive to put knowledge into computer systems, and to have computers share what they know with each other when desirable by their human masters.

I have come to the conclusion that the knowledge representation problem needs to be reframed so as to place the entire burden of worldly connectedness on the computer, using sensorimotor data streams as the connecting interface. The brain would consist of a hardware or firmware representation process that records and compresses the lifetime sensory-motor data stream as it is being collected. Symbols standing for patterns would be induced and recorded. At least then, the computer could maintain a traceable link between the outside world (*i.e.*, the peripheral I/O data stream) and its internal symbol system, having created this mapping itself. (If these ideas seem new and crazy to you, you're not alone—I feel the same way. It may take two or three years to see where this kind of thinking will lead, and then some. Right now, though, we should be satisfied that new ideas offer strengths in exactly the same places current AI beliefs are weakest—*i.e.*, in learning and totality.)

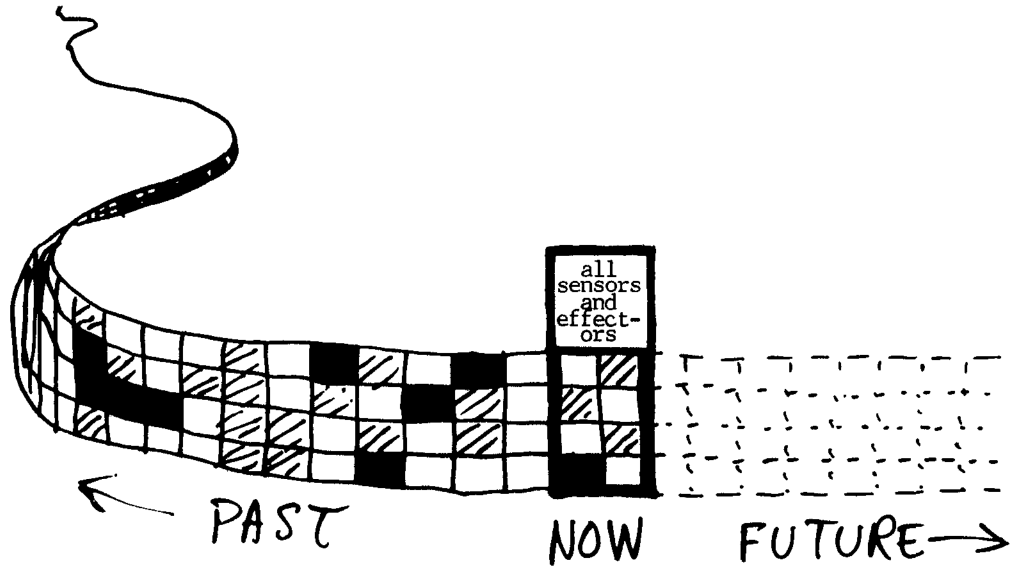
It would benefit AI to give the topic of knowledge representation a fresh, new hearing, this time working with a strong approach to knowledge acquisition, segmentation, lifelong data streams, and transmission of knowledge. We will, I am almost certain, need a fresh theory of knowledge strong in these areas in order to achieve the professor's goal of machine versatility. Once this theory is hammered out, I think it will become clear how the robot will be able to learn what it will need to pass the versatility demonstration, and people will believe it can be done before the robot has even been built. Such is my optimism.

But before this, we need to break out of the "mind set" of some previous conceptions. To review seven clues for shifting our thinking:

1. Be aware of prerequisite knowledge and the crucial role it plays in time-order of knowledge acquisition
2. Collect knowledge for its own sake in an ongoing process, and renew your allegiance to Turing's goal statement.
3. The world isn't really out there, it's all in your mind. Prepare for religious conversion to Egg-think
4. Enlarge context to the maximum, so that informative value of input depends on everything that has come before.
5. Get your hands on a good totality of knowledge concept before going further
6. Make the machine responsible for segmentation Give it only one very, very long, ongoing input "sentence".
7. Acquire and transmit knowledge through sensorimotor I/O streams; be able to trace meaning of all internal symbols formed.

Future Directions

From the preceding exploration of why existing methods are not up to the task of versatile intelligence, we can begin to see a rough outline of some new methods taking shape. The strong points of these new thinking tools will be not



The External World of the Individual.

Figure 5.

domain-restricted problem solving, but long-term knowledge aggregation for general purpose utilization.

Thinking Tool #1: Lifelong Data Streams

We need to begin thinking of intelligence, behavior, the external environment, and communication in terms of their totality over the lifetime of the *individual intelligent machine*. How can this be done? One way to get a handle on this totality is to define the external world separately for each individual as being the lifelong stream of data crossing the sensory-motor periphery. Notice that the external world, so defined, extends into the indefinite future (Figure 5). This leads to the following Totality of Knowledge claim: Can you think of any knowledge you either have now or can acquire in the future that is not expressed in this data stream?

If you are uncomfortable with the idea of equating the totality of knowledge with the totality of your personal knowledge, then throw in the lifelong, peripheral data streams of everyone else as a way to circumscribe the Totality of Knowledge (Figure 6).

I'm satisfied with the first formulation—it doesn't much matter to me if some thing or some knowledge exists that I can never know about. It's sort of like extraterrestrial life—it doesn't make a whole lot of difference whether or not it actually exists unless it begins to communicate, in which case

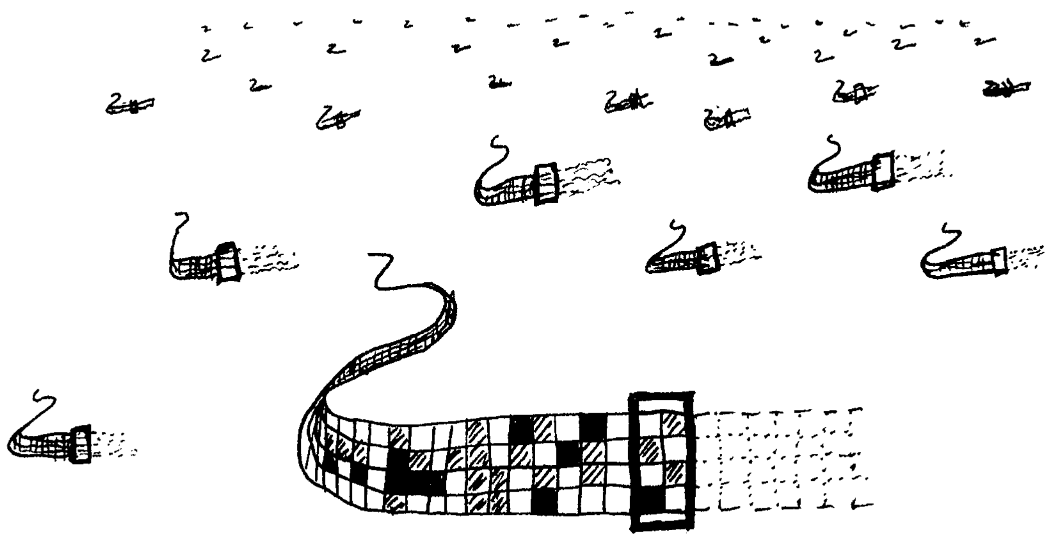

Totality of Knowledge Circumscribed by the External Worlds of All Beings.

Figure 6.

it becomes covered by the umbrella of the egocentric Totality of Knowledge formulation, future inputs department.[1]

In robotics, for example, the external environment will be conceptualized as the cumulative sensory data stream appearing at the remotest edges of the camera, microphone, and tactile sensors, combined with the cumulative effector data stream transduced at arm-wrist-hand, camera-adjustment, camera-aiming, locomotion, audio-output, and video-output transducers. The earliest suggestion of the data stream world model I have found is in Solomonoff (1960).

Thinking Tool #2: The Representation Process

The most significant subproblem to be addressed is how the computer can remember this totality as it has transpired up to the present. This memory should consist of a symbol base built automatically by the computer on the basis of input and memory combined. This representation process makes a symbolic recording a lifelong sensory-motor I/O, whereby a compressed version of it is kept in which patterns have been segmented. Existing symbols serve as prerequisite knowledge, or available building blocks for learning new symbols; as soon as a new symbol is acquired, it joins the pool of knowledge available for use as prerequisite knowledge. Larry Rendell's (1983) idea of "knowledge as information compression" and Paul Scott's (1983) view of intelligence as "the organization of experience" indicate mounting interest in experiential recording as a unifying AI research theme

Thinking Tool #3: Transmission Of Knowledge

Intelligent communication works on the principle of causing symbol formation to occur in the receiver, indirectly, by generating patterns in his sensory data stream. The representation process within the receiver fields communications the same way it handles knowledge acquisition from inanimate sources—there is no need for a separate theory to explain communication.

This last thinking tool is the most daring one. It leads to a communication theory that views language as a learned system of audio visual patterns that, through consistent presentation with previously symbolified patterns, leads to an "indexing" system for recalling knowledge and behavior.

These are the three most important themes that need to be kept in mind as we reshape our thinking from single-purpose, programmed, performing systems to multipurpose, automatic learning systems. The groundwork is just now beginning to be laid (Bierre, 1982) for what promises to be an exciting, new direction in AI. The challenge of versatility poses many difficult problems that we are just starting to learn how to articulate, and that offer opportunities for genuine breakthroughs to the scientists who possess the courage and determination to solve them As an example, we need people who can think about hardware, software, and memory in nontraditional ways, in order to devise a computer for doing representation in parallel, on multiple parallel data stream input, in real time. Associative memories (Foster, 1976; Weems, Levitan & Foster, 1983) and connection machines (Hillis, 1981) are important steps in this direction, but much more needs to be done. Enough work is here to keep dozens of AI professionals busy for a decade or more.

A few words need to be said about applied versus research AI. Currently, interest is keen in applying expert systems techniques to real world problems where there is economic leverage, and this interest will expectably command the attention of the majority of the AI community in the immediately ensuing years. It is important that domain-bounded intelligence continue to be developed toward useful applications. There is a danger, though, that new research initiatives, such as sensory learning, might be unfairly pitted against applied AI in the scramble for funding. I think it is important to maintain a clear distinction between research and applied projects, and to pursue a well-balanced approach that will continue to produce a steady flow of working systems, and progressive, far-ranging, new scientific ideas for future systems. According to John McCarthy (McCarthy, 1983), we have to push harder on the basic research front at this point to maintain that balance.

As I have indicated, the price we will likely have to pay for dramatic progress will be backtracking on some dearly held beliefs that skirt the fringes of unquestioned, religious belief. A defensive reaction to the call for radically new AI goals, beliefs, and methods is only natural, and we cannot expect the ensuing debate to remain on a strictly rational level. But if the history of science is any indication, such a period of intense struggle over conflicting fundamental beliefs is the most reliable predictor that a dramatic surge of scientific progress is about to be made (Kuhn, 1962).

It is precisely this surge of progress that the next ten years of AI research should be about. Already, stirrings of doubt indicate that we have gone about as far as we are going to go with formal logic and heuristic programming methods. Rather than ignore and repress these faint rumblings, we must allow them to surface into free expression, and at the same time begin aggressively to pursue alternative AI frameworks that show promise for allowing us to hurdle long-accepted barricades.

Let us keep in mind one final observation: In posing the versatility problem, the professor instinctively knew that his challenge would disqualify current AI methods, and in so doing would inject doubt, confusion, and other unsettling feelings among members of his research staff, uneasy feelings that could spill over into the entire AI community. You have to understand that the professor is getting on in years (as many of us are), and for whatever selfish reasons, he secretly wants to see the mystery of human-quality intelligence unraveled, not in 50 years, but in his own lifetime. I suspect he is feeling the very mortal craving to want to be there and see how it all turns out in the end.

I'm with the professor. How about you?

[1]For a more thorough treatment of knowledge totality, see Bierre, 1984

References

Bierre, P. (1982) The Representation of Knowledge. Tech. Rep. 002, Clairvoyant Systems, Mountain View, CA.

Bierre, P. (1984) The Totality of Knowledge. Tech. Rep. 004, Clairvoyant Systems, Mountain View, CA.

Foster, C. C (1976) *Content Addressable Parallel Processors.* New York: Van Nostrand Reinhold.

Hillis, W. D. (1981) The Connection Machine. Tech. Rep. 646, MIT AI Lab.

Kuhn, T. S. (1962) *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Langley, P., Bradshaw, G. L., & Simon, H. A. (1981) BACON.5: The Discovery of Conservation Laws. *IJCAI* 7:121-126.

Lenat, D. B. & Brown, J. S. (1984) Why AM and EURISKO Appear to Work. *Artificial Intelligence* Vol. 23, 3:269-294.

Lowe, D. G., & Binford, T. O. (1983) Perceptual Organization as a Basis for Visual Recognition. *AAAI-83:* 255-260.

McCarthy, J. (1983) President's Quarterly Message: AI Needs More Emphasis on Basic Research. *AI Magazine* Vol 4, 4:5

Minsky, M. (1983) Why People Think Computers Can't. *MIT Technology Review* Vol. 86, No 3: 65-70, 80-81

Nilsson, N J. (1983) Artificial Intelligence Prepares for 2001 *AI Magazine* Vol. 4, No 4: 7-13.

Rendell, L. A. (1983) Toward a Unified Approach for Conceptual Knowledge Acquisition. *AI Magazine* Vol. 4, No 4: 19-27.

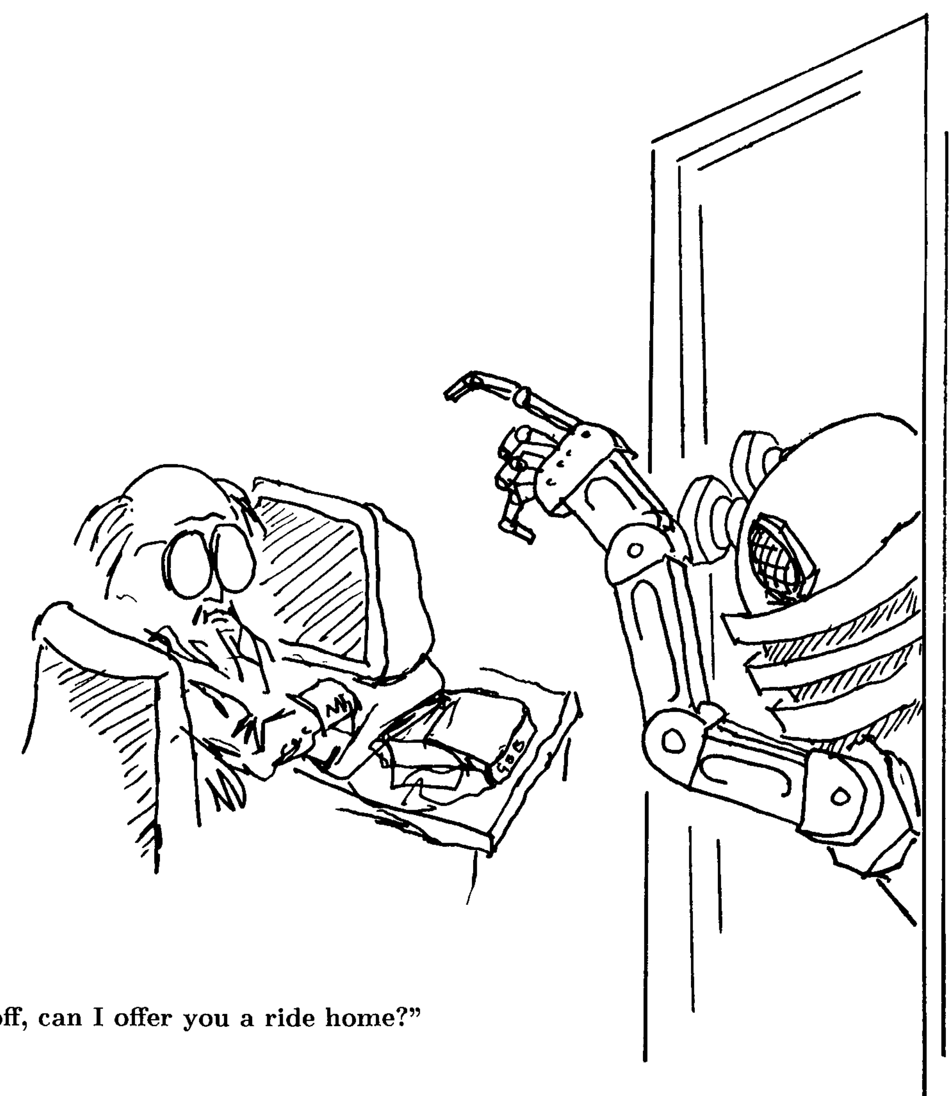
Scott, P. D. (1983) Learning: The Construction of A Posteriori Knowledge Structures *AAAI-83:* 359-363.

Solomonoff, R. J (1960) A Preliminary Report on a General Theory of Inductive Inference *Zator* ZTB-138, Cambridge, MA, November, 1960

Turing, A. M (1950) Computing Machinery and Intelligence *Mind* 59: 433-460

Weems, C., Levitan, S. & Foster, C. (1983) Titanic: A VLSI Based Content Addressable Parallel Array Processor. *COINS* Tech. Rep. 83-32, University of Massachusetts at Amherst,

Witkin, A. P , & Tenenbaum, J M (1983) On the Role of Structure in Vision. In Rosenfelt and Beck (*Eds*) *Human and Machine Vision* New York: Academic Press.



"Dr. Nokemoff, can I offer you a ride home?"