

Human Intelligence

Hubert L. Dreyfus

Hubert L. Dreyfus is a contemporary philosopher who has written extensively in phenomenology and existential philosophy in addition to his philosophical criticisms of the field of artificial intelligence. He is the editor of Husserl, Intentionality and Cognitive Science (with H. Hall, 1982) and the author of What Computers Can't Do (1979), Michel Foucault: Beyond Structuralism and Hermeneutics (1983), Mind over Machine (with S. Dreyfus, 1986), and Being-in-the-World: A Commentary on Heidegger's Being and Time (1 986).

In the essay below Dreyfus argues that digital computers will never display ordinary commonsense understanding or intuitive expertise, and that these are the hallmarks of intelligent human behavior. Digital computers can only possess information in the form of representations, all of whose possibly relevant features must be explicit. And the only processes of which the computer is capable are explicit feature-determined and rule-guided operations on these representations. Dreyfus argues that ordinary human knowledge does not consist of such information, and than intuitive expertise in all areas, from the exercise of common sense to the employment of highly technical skills, is not guided by explicit rules of this sort at all. If these claims are correct, it means that the attempt by AI researchers to represent human knowledge and duplicate intelligent human behavior with digital computers is fundamentally misguided.

Misrepresenting Human Intelligence by Hubert L. Dreyfus

In What Computers Can't Do I argued that research in artificial intelligence (AI for short) was based upon mistaken assumptions about the nature of human knowledge and understanding. In the first part of this essay I will review that argument briefly. In spite of the noisy protestations from AI researchers, most of my critical claims and negative predictions have not only been borne out by subsequent research and developments in the field, but have even come to be acknowledged as accurate indications of major problems by AI workers themselves. In the years since the revised edition of my book was published, what I have come to see is not only that my

early pessimism was well-founded, but also that some of my assessments for the future of AI were overly optimistic. In the second part of this essay I will explain why I now believe that even the cautious and guarded optimism which I once had with respect to certain isolated areas of AI research was unjustified and, ultimately, mistaken.

I.

The early stages of AI research were characterized by overly ambitious goals, wishful rhetoric and outlandish predictions. The goal, in general, was to equal or exceed the capacities of human beings in every area of intelligent behavior. The rhetoric turned failure after failure into partial and promising success. And the predictions had computers doing everything an intelligent human being could do within a decade or so at most. The terms on these predictions have all expired with none of the miraculous feats accomplished, and most researchers have begun to face the hard facts about the real limits of artificial intelligence.

The basic project of AI research is to produce genuine intelligence by means of a programmed digital computer. This requires, in effect, that human knowledge and understanding be reconstructed out of bits of isolated and meaningless data and sequences of rule-governed operations. The problems facing this approach can be put quite simply. Human knowledge and understanding do not consist of such data, rules and operations; and nothing which does consist essentially of these things will ever duplicate any interesting range of intelligent human behavior.

Early research projects in artificial intelligence tried to meet head on the task of duplicating human mentality. Major areas of emphasis included natural language understanding, pattern recognition and general problem solving. Problems in each of these areas were seen initially as problems of size-organizing and using a very large quantity of data. In order to understand even a small and ordinary sample of natural language, for example, a very large mass of background facts seemed to be required, and this massive collection of facts seemed in turn to require some kind of organization in terms of relevance so that not every fact required explicit consideration in every exercise of linguistic understanding. Pattern recognition research ran into similar problems. The number of possibly relevant features was immense and rules for separating those features actually relevant to recognition of a given shape or figure from all the others proved incredibly difficult to formulate. More general problem solving faced exactly the same difficulties,

only several orders of magnitude larger due to the increased generality of the task.

In the first edition of What Computers Can't Do I argued that the problems encountered by AI workers in these research areas were not just a matter of size, and would not succumb to more efficient programs and programming languages or to dramatic increases in computing speed and storage capacity. None of the available empirical evidence suggested that human beings function in the manner required by then current AI models, and much of the evidence suggested an entirely different and incompatible view of human mentality-but this view emerges only after some long-standing psychological and philosophical assumptions are discarded. Those assumptions lie at the very heart of the information processing model of the mind. Put generally, there is only one assumption-that human mental processes are essentially identical to those of a digital computer. Put more specifically, the crucial assumptions are these: 1) that mental processes are sequences of rule-governed operations and 2) that these operations are carried out on determinate bits of data (symbols) which represent features of or facts about the world (information, but only in a technical sense of that term). I can best explain why those specific assumptions are implausible by looking at the problems encountered in each area of AI research .

The main problem for programs whose aim was to understand natural language was the need to do so either without any context to determine or disambiguate meaning, or else with a context completely spelled out in terms of explicit facts, features, and rules for relating them. Even the meanings of individual terms are context- or situation-determined. Whether the word "pen" refers to a writing implement, a place for infants to play, a place in which pigs or other animals are kept, the area of a baseball field in which pitchers warm up or a place for confining criminals is determined by the context in which the term appears within a story or conversation. It is even clearer that context determines the meaning of whole sentences. "The book is in the pen" could mean that the child's storybook is in the playpen, that the pigs' enclosure is where the paperback fell out of the farmer's pocket, that the microfilm of the diary is cleverly concealed in the compartment containing the ink cartridge, and so on. We human beings don't seem to need to explicitly consider all the alternatives; and, in fact, it's not even clear that there could be an exhaustive list of all the alternative meanings for typical samples of natural language. We are always involved in a situation or context which seems to restrict the

range of possible meanings without requiring explicit or exhaustive consideration of the range of context-free alternatives.

The obvious solution from the standpoint of AI would be to give the computer the situation. The attempt to do this in a general way has been unsuccessful, and begs some important questions. The most important question begged is whether or not our command of the situation is just a matter of a number of facts which we accept, that is, a system of beliefs which could be made explicit. If it is, then it could, at least in principle, be given to a computer, and the only problems would be practical, problems of size and structure for the belief system. If our command of the situation cannot be represented as such a belief system, however, then there will be no way to get the computer into the situation or the situation into the computer so as to duplicate general human understanding. I believe that this is the real impasse which AI faces.

Influenced by philosophers such as Heidegger and Merleau-Ponty, I believe the evidence points toward the following picture of the relation between facts and situations. Our sense of the situation we are in determines how we interpret things, what significance we place on the facts, and even what counts as facts for us at any given time. But our sense of the situation we are in is not just our belief in a set of facts, nor is it a product of independent facts or context-free features of our environment. The aspects of our surroundings which somehow give us our sense of our current situation are themselves products of a situation we were already in, so that situations grow out of situations without recourse to situation-neutral facts and features at any point. We never get into a situation from outside any situation whatsoever, nor do we do so by means of context-free data. But the computer has only such data to work with and must start completely unsituated. From the standpoint of the programmer, our natural situatedness consists of an indefinite regress of situations with no way to break in from the outside, no way to start from nothing. And this is only part of the problem for AI. Not only can the situation not be constructed out of context-free facts and features, but in fact the situation as it figures in intelligent human behavior is not primarily a matter of facts and features of any kind. It is much more like an implicit and very general sense of appropriateness, and seems to be triggered by global similarities to previously experienced situations rather than by any number of individual facts and features. I will try to make clear just how I think this works in the second part of this essay. Here I will simply observe that, lacking access to anything very like the human situation, it is not surprising that digital computers also lack access to anything very like human understanding.

Pattern recognition programs encountered similarly instructive difficulties. Whether it is a matter of recognizing perceived figures and shapes or the similarity of board positions or sequences of moves in a game of chess, the context seems to guide human pattern recognition in ways that cannot be duplicated by the computer using context-free features and precise rules for relating them. Our sense of our situation seems to allow us to zero in on just those features that are relevant to the task at hand and to virtually ignore an indefinite number of further features. Moreover, a great deal of human pattern recognition seems to be based on the perception of global similarity and not to involve any feature by feature comparison at all.

The expert chess player sees the board in terms of fields of force rather than precise positions of each of the individual pieces, recognizes intuitively the similarity to situations encountered in previous games even though none of the individual positions of pieces and few of the objective relations among individual pieces are the same, and selects a move after explicitly examining relatively few alternatives. The computer, on the other hand, analyzes the board in terms of the position of each of the pieces and then either has recourse to heuristic rules which connect that information to precise moves or else uses brute computing force to examine every possible course of action to as great a depth as time will allow. The latter strategy has been more successful, but the moves selected by this technique are still inferior to those chosen by the human expert in spite of the fact that the computer examines thousands of times as many future positions in its selection process. I will explain exactly why this is so in the second part of the essay.

Humans also recognize people they know or familiar surroundings without noticing, much less carefully comparing, the individual features of the persons or things recognized. Duplicating this kind of ordinary recognition has proven impossible for AI. The reason, I think, is the same as in the case of the expert chess player, but more on this later.

Programs designed to duplicate the general human ability to solve problems achieved results only by restricting the task in such a way that general human problem solving was never at issue. Programs typically solved word puzzles which were restricted so as to include only relevant information and which contained explicit cues to invoke the correct heuristic rule as needed. The human ability to identify the kind of problem faced, to sort information in terms of relevance, and to find the correct method of solution on the basis of similarity to previously solved problems-that is, full-fledged human

problem solving-was simply bypassed, supplied, in effect, by the human 'processing' of the problems to be solved'.

In the late sixties and early seventies, the difficulties described above were taken seriously by workers in AI. Instead of trying to duplicate in one giant step general human understanding of the world, attention turned to producing understanding in very restricted 'worlds'. These artificially restricted domains were called 'microworlds'. Impressive micro-world successes included Terry Winograd's SHRDLU, Thomas Evans's Analogy Problem Program, David Waltz's Scene Analysis Program and Patrick Winston's program for learning concepts from examples. The micro-worlds were constrained in such a way that the problems of context-restricted relevance and context-determined meaning seemed to be manageable. The hope was that micro-world techniques could be extended to more general domains, the micro-worlds made increasingly more realistic and combined to eventually produce the everyday world, and the computer's capacity to cope with these micro- worlds thereby transformed into genuine artificial intelligence.

The subsequent failure of every attempt to generalize micro-world techniques beyond the artificially restricted domains for which they were invented has put an end to the hopes inspired by early micro-world successes and brought AI to a virtual standstill. Some researchers, including Winograd, have given up on AI entirely. The micro-world strategy failures [have been instructive, however, focusing attention in the direction I had argued was crucial for more than a decade, namely, toward the nature of everyday human understanding and knowhow. The problem encountered in the attempt to move from micro-worlds to any aspect of the everyday world is that micro-worlds aren't worlds at all, or, from the other side, domains within the everyday world aren't anything like microworlds. This insight emerged in the attempt to program children's story understanding. It was soon discovered that the 'world' of even a single child's story, unlike a micro-world, is not a selfcontained domain and cannot be treated independently of the larger everyday world onto which it opens. Everyday understanding is presupposed in every real domain, no matter how small. The everyday world is not composed of smaller independent worlds at all, is not like a building which can be built up of tiny bricks but is rather a whole somehow present in each of its parts. Once this was realized, micro-world research and its successes were recognized for what they really were, not small steps toward the programming of everyday or common-sense know-how and understanding, but clever evasions of the real need to program such general competence and

understanding. And the prospects for programming a digital computer to display our everyday understanding of the world were looking less bright all the time. Cognitive scientists were discovering the importance of images and prototypes in human understanding. Gradually most researchers were becoming convinced that humans form images and compare them by means of holistic processes very different from the logical operations which computers perform on symbolic descriptions.²

A recent Scientific American article echoed my earlier assessment of AI:

Probably the most telling criticism of current work in artificial intelligence is that it has not yet been successful in modeling what is called common sense.... [S]ubstantially better models of human cognition must be developed before systems can be designed that will carry out even simplified versions of common-sense tasks.³

II.

For the reasons discussed in the preceding section, I concluded in 1979 that AI would remain at a standstill in areas that required common-sense understanding of the everyday world, that there would be no major breakthroughs in interpreting ordinary samples of natural language, in recognizing ordinary objects or patterns in every day contexts, or in everyday problem solving of any kind within a natural rather than artificially constrained setting. The evidence to date indicates that I was correct in my assessment of AI's prospects in these areas. However I also predicted success for AI in certain isolated tasks, cut off from the everyday world and seemingly selfcontained, tasks such as medical diagnosis and spectrograph analysis. It appeared to me at the time that ordinary common sense played no role in such tasks and that the computer, with its massive data storage capacity and ability to perform large numbers of inferences almost instantaneously and with unerring accuracy, might well equal or exceed the performance of human experts. It has turned out that I was mistaken about this. In a book that we have just finished,⁵ my brother, Stuart, and I attempt to explain this surprising result. Here I can give only a brief account of that explanation.

The attempt to give computers human expertise in these special domains has come to be referred to as "expert systems" research. It works as follows. Human experts in the domain are interviewed to ascertain the rules or principles which they employ. These are then programmed into the computer. The idea seems simple and

uncontroversial. Human experts and computers work from the same facts with the same inference rules. Since the computer can't forget or overlook any of the facts, can't make any faulty inferences, and can make correct inferences much more swiftly than the human expert, the expertise of the computer should be superior. And yet in study after study the computer proves inferior to the human experts who provide its working principles. To understand how this is possible, we need to look closely at the process by which humans acquire expertise.

The following model of the stages of skill acquisition emerged from our study of that process among airplane pilots, chess players, automobile drivers, and adult learners of a second language. We later found that our model fit almost perfectly data which had been gathered independently on the acquisition of nursing skills.⁶ The model consists of five stages of increasing skill which I will summarize briefly in terms of the chess players. For more mundane skills such as automobile driving, you may be able to check much of the model against your own past experience.

Stage 1-Novice

During this first stage of skill acquisition through instruction, the novice is taught to recognize various objective facts and features relevant to the skill, and acquires rules for determining what to do based upon these facts and features. Relevant elements of the situation are defined so clearly and objectively for the novice that recognition of them requires no reference to the overall situation in which they occur. Such elements are, in this sense, context-free. The novice's rules are also context-free in the sense that they are simply to be applied to these context-free elements regardless of anything else that may be going on in the overall situation. For example, the novice chess player is given a formula for assigning point values to pieces independent of their position, and the rule, "always exchange your pieces for the opponent's if the total value of pieces captured exceeds that of pieces lost." The novice is generally not taught that there are situations in which this rule should be violated.

The novice typically lacks a coherent sense of his overall task and judges his performance primarily in terms of how well he has followed the rules he has learned. After he acquires more than just a few such rules, the exercise of this skill requires such concentration that his capacity to talk or listen to advice becomes very limited.

The mental processes of the novice are easily imitated by the digital computer. Since it can use more rules and consider more context-free elements in a given amount of time, the computer typically outperforms the novice.

Stage 2-Advanced Beginner

Performance reaches a barely acceptable level only after the novice has considerable experience in coping with real situations. In addition to the ability to handle more context-free facts and more sophisticated rules for dealing with them, this experience has the more important effect of enlarging the learner's conception of the world of the skill. Through practical experience in concrete situations with meaningful elements which neither instructor nor learner can define in terms of objectively recognizable context-free features, the advanced beginner learns to recognize when these elements are present. This recognition is based entirely on perceived similarity to previously experienced examples. These new features are situational rather than context-free. Rules for acting may now refer to situational as well as context-free elements. For example, the advanced chess beginner learns to recognize and avoid over-extended positions, and to respond to such situational aspects of board positions as a weakened king's side or a strong pawn structure even though he lacks precise objective definitional rules for their identification.

Because the advanced beginner has no context-free rules for identifying situational elements, he can communicate this ability to others only by the use of examples. Thus the capacity to identify such features, as well as the ability to use rules which refer to them, is beyond the reach of the computer. The use of concrete examples and the ability to learn context-determined features from them, easy for human beings but impossible for the computer, represents a severe limitation on computer intelligence.

Stage 3-Competence

As a result of increased experience, the number of recognizable elements present in concrete situations, both context-free and situational, eventually becomes overwhelming. To cope with this the competent performer learns or is taught to view the process of decision making in a hierarchical manner. By choosing a plan and examining only the relatively small number of facts and features which are most important, given the choice of plan, he can both simplify and improve his performance. A competent chess player, for example, may decide, after studying his position and weighing

alternatives, that he can attack his opponent's king. He would then ignore certain weaknesses in his own position and personal losses created by his attack, and the removal of pieces defending the enemy king would become salient.

The choice of a plan, although necessary, is no simple matter for the competent performer. It is not governed by an objective procedure like the context-free feature recognition of the novice. But performance at this level requires the choice of an organizing plan. And this choice radically alters the relation between the performer and his environment. For the novice and the advanced beginner, performance is entirely a matter of recognizing learned facts and features and then applying learned rules and procedures for dealing with them. Success and failure can be viewed as products of these learned elements and principles, of their adequacy or inadequacy. But the competent performer, after wrestling with the choice of a plan, feels personally responsible for, and thus emotionally involved in, the outcome of that choice. While he both understands his initial situation and decides upon a particular plan in a detached manner, he finds himself deeply involved in what transpires thereafter. A successful outcome will be very satisfying and leave a vivid memory of the chosen plan and the situation as organized in terms of that plan. Failure, also, will not be easily forgotten.

Stage 4-Proficiency

The novice and advanced beginner simply follow rules. The competent performer makes conscious choices of goals and plans for achieving them after reflecting upon various alternatives. This actual decision making is detached and deliberative in nature, even though the competent performer may agonize over the selection because of his involvement in its outcome.

The proficient performer is usually very involved in his task and experiences it from a particular perspective as a result of recent previous events. As a result of having this perspective, certain features of the situation will stand out as salient and others will recede into the background and be ignored. As further events modify these salient features, there will be a gradual change in plans, expectations, and even which features stand out as salient or important. No detached choice or deliberation is involved in this process. It seems to just happen, presumably because the proficient performer has been in similar situations in the past and memory of them triggers E similar to those which worked in the past and

expectations of further events similar to those which occurred previously.

The proficient performer's understanding and organizing of his task is intuitive, triggered naturally and without explicit thought by his prior experience. But he will still find himself thinking analytically about what to do. During this reasoning, elements that present themselves as salient due to the performer's intuitive understanding will be evaluated and combined by rule to yield decisions about the best way to manipulate the environment. The spell of involvement in the world of the skill is temporarily broken by this detached and rule-governed thinking. For example, the proficient chess player' can recognize a very large repertoire of types of positions. Recognizing almost immediately, and without conscious effort, the sense of a position, he sets about calculating a move that best achieves his intuitively recognized plan. He may, for example, know that he should attack, but he must deliberate about how best to do so.

Stage 5-Expertise

The expert performer knows how to proceed without any detached deliberation about his situation or actions, and without any conscious contemplation of alternatives. While deeply involved in coping with his environment, he does not see problems in a detached way, does not work at solving them, and does not worry about the future or devise plans. The expert's skill has become so much a part of him that he need be no more aware of it than he is of his own body in ordinary motor activity. In fact tools or instruments become like extensions of the expert's body. Chess grandmasters, for example, when engrossed in a game, can lose entirely the awareness that they are manipulating pieces on a board, and see themselves instead as involved participants in a world of opportunities, threats, strengths, weaknesses, hopes and fears. When playing rapidly they sidestep dangers in the same automatic way that a child, himself an expert, avoids missiles in a familiar video game. In general, experts neither solve problems nor make decisions; they simply do what works. The performance of the expert is fluid and his involvement in his task unbroken by detached deliberation or analysis.

The fluid performance of the expert is a natural extension of the skill of the proficient performer. The proficient performer, as a result of concrete experience, develops an intuitive understanding of a large number of situations. The expert recognizes an even larger number along with the associated successful tactic or decision.

When a situation is recognized, the associated course of action simultaneously presents itself to the mind of the expert performer. It has been estimated that a master chess player can distinguish roughly 50,000 types of positions. We doubtless store far more typical situations in our memories than words in our vocabularies. Consequently these reference situations, unlike the situational elements learned by the advanced beginner, bear no names and defy complete verbal description.

The grandmaster chess player recognizes a vast repertoire of types of positions for which the desirable tactic or move becomes immediately obvious. Excellent chess players can play at a rate of speed at which they must depend almost entirely on intuition and hardly at all upon analysis and the comparison of alternatives, without any serious degradation in their performance. In a recent experiment International Master Julio Kaplan was required rapidly to add numbers presented to him audibly at the rate of about one number per second, while at the same time playing five-second-a-move chess against a slightly weaker, but master level, player. Even with his analytical mind completely occupied with the addition, Kaplan more than held his own against the master in a series of games. Deprived of the time necessary to see problems or construct plans, Kaplan still produced fluid and coordinated play.

What emerges from this model of human skill acquisition is a progression from the analytic, rule-governed behavior of a detached subject who consciously breaks down his environment into recognizable elements, to the skilled behavior of an involved subject based on an accumulation of concrete experiences and the unconscious recognition of new situations as similar to remembered ones. The innate human ability to recognize whole current situations as similar to past ones facilitates our acquisition of high levels of skill and separates us dramatically from the artificially intelligent digital computer endowed only with context-free fact and feature recognition devices and with inference-making power.

This model provided Stuart and me with an explanation of the failure of the expert systems approach which also connects it with the failure of previous work in AI. When the interviewer elicits rules and principles from the human expert, he forces him, in effect, to revert to a much lower skill level at which rules were actually operative in determining his actions and decisions.

This is why experts frequently have a great deal of trouble 'recalling' the rules they use even when pressed by the interviewer. They seem more naturally to think of their field of expertise as a huge set of

special cases.⁷ It is no wonder that systems based on principles abstracted from experts do not capture those experts' expertise and I hence do not perform as well as the experts themselves.

In terms of skill level, the computer is stuck somewhere between the novice and advanced beginner level and if our model of skill acquisition is accurate, has no way of advancing beyond this stage. What has obscured this fact for so long is the tremendous memory of the computer, in terms of numbers of facts and features which can be stored, and the tremendous number of rules and principles which it can utilize with super human speed and accuracy. Although its skill is of a kind which would place it below the level of the advanced beginner, its computing power makes its performance vastly superior to that of a human being at the same skill level. But power of this kind alone is not sufficient to duplicate the ability, the intuitive expertise, of the human expert.

This model of human skill levels also explains the failure of AI researchers to duplicate human language understanding, pattern recognition, and problem solving. In each of these areas we are, for the most part, experts. We are expert perceivers, expert speakers, hearers and readers of our native language, and expert problem solvers in most areas of everyday life. That doesn't mean that we don't make mistakes, but it does mean that our performance is entirely different in kind from that of the programmed digital computer. In each of these areas the computer is, at best, a very powerful and sophisticated beginner, competent in artificial micro-worlds where situational understanding and intuitive expertise have no part to play, but incompetent in the real world of human expertise.

I still believe, as I did in 1965,* that computers may someday be intelligent. Real computer intelligence will be achieved, however, only after researchers abandon the idea of finding a symbolic representation of the everyday world and a rule-governed equivalent of common-sense know-how, and turn to something like a neural-net modeling of the brain instead. If such modeling turns out to be the direction that AI should follow, it will be aided by the massively parallel computing machines on the horizon--not because parallel machines can make millions of inferences per second--but because faster, more parallel architecture can better implement the kind of pattern processing that does not use representations of rules and features at all.