Prof. Hubert & Stuart Dreyfus
University of California, Berkeley

From Socrates to Expert Systems

I. What is Cognitivism?

Cognitivism is not simply a theory of cognition but, as the name, cognitiv*ism*, suggests, it is the strong view that *all* mental activity is cognitive -- that perception, understanding, learning and action are all to be understood on the model of fact gathering, hypothesis formation, inference making, and problem solving. Such a view is taken for granted by current cognitive psychologists, especially those seeking to program computers to behave intelligently. They think of themselves as pioneers of a new discipline, cognitive science, the true science of the mind, which has done for the mind what Galileo did for the universe and Darwin did for biology.

The cognitivist assumption is so self-evident to cognitive scientists that they seldom bother to state it explicitly, let alone argue for it. For example, Jerry Fodor says in passing, as if it were obvious, that mental life is essentially thought and thought involves inference.

> Conan Doyle was a far deeper psychologist -- far closer to what is essential about mental life -- than, say, James Joyce (or William James, for that matter).[1]

Lest my interpretation of this off-hand remark seems forced, note that Fodor's account of perception, which is least like thought, turns even it into detective work: "Perception typically involves hypothesis formation and confirmation."[2] In general all cognitivist theories make two essential assumptions:

1. Our ability to deal with things intelligently is due to our capacity to think about them reasonably (including subconscious thinking).
2. Our capacity to think about things reasonably amounts to a faculty for internal "automatic" symbol manipulation.[3].

This view is the culmination of the rationalist philosophical tradition. Descartes already assumed that all understanding consisted in forming and manipulating appropriate representations, that these representations could be analyzed into primitive elements, and that all phenomena could be understood as a complex combination of these simple elements. At the same time, Hobbes implicitly assumed that the elements were formal elements related by purely syntactic operations, so that reasoning could be reduced to calculation. "When a man *reasons*,

---

[1] J. A. Fodor, "Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum," Mind, 1985 p. 91.

[2] J.A. Fodor, *The Language of Thought*, Thomas Y. Crowell Company, 1975,p. 44.

[3] John Haugeland, Artificial Intelligence, The Very Idea, Bradford/MIT Press, 1985, p. 113.

he does nothing else but conceive a sum total from addition of parcels," Hobbes wrote, "for REASON ...is nothing but reckoning ..."[4] Finally, Leibniz, working out the classical idea of *mathesis*-- the formalization of everything -- sought grant money to develop a universal symbol system, so that "we can assign to every object its determined characteristic number."[5] According to Leibniz, in understanding we analyze concepts into more simple elements. In order to avoid a regress of simpler and simpler elements, there must be ultimate simples in terms of which all complex concepts can be understood. Moreover, if concepts are to apply to the world, there must be logical simples in the world, which the simple symbols represent. Leibniz envisaged "a kind of alphabet of human thoughts"[6] whose "characters must show, when they are used in demonstrations, some kind of connection, grouping and order which are also found in the objects."[7]

Cognitivism is rationalism turned into a research program. It proposes to use the computer to show how, on the rationalist account, the mind actually works. In a seminal statement of this project, Allen Newell and Herbert Simon claim that the human brain and the digital computer, while totally different in structure and mechanism, have, at the appropriate level of abstraction, a common functional description. At this level, both the human brain and the appropriately programmed computer can be seen as two different instantiations of a single species of device -- one which generates intelligent behavior by manipulating symbols according to formal rules. Newell and Simon state their view as a hypothesis:

> The Physical Symbol System Hypothesis. A physical symbol system has the necessary and sufficient means for general intelligent action. .
>
> By "necessary" we mean that any system that exhibits general intelligence will prove upon analysis to be a physical symbol system. By "sufficient" we mean that any physical symbol system of sufficient size can be organized further to exhibit general intelligence.[8]

The cognitivist viewpoint can be seen starkly by looking at the assumptions underlying the building of expert systems, since such systems involve a consistent application of the cognitivist perspective. Building expert systems is the work of so-called *knowledge engineers*. What the knowledge engineers claim to have discovered is that in areas that are cut off from everyday common sense and social intercourse, all a machine needs in order to behave like an

---

[4] Hobbes, Leviathan, Library of Liberal Arts, 1958, p. 45.

[5] Leibniz, Selections, ed. Philip Wiener, Scribner, 1951, p. 18.

[6] Ibid. p. 20.

[7] Ibid. p. 10

[8] Allen Newell and Herbert Simon, "Computer Science as Empirical Inquiry: Symbols and Search", in Mind Design, John Haugeland, ed., (Cambridge: Bradford/MIT Press, 1981), p. 41.

expert are some general rules and lots of very specific knowledge. This specialized knowledge is of two types:

> The first type is the *facts* of the domain -- the widely shared knowledge ... that is written in textbooks and journals of the field ... Equally important to the practice of the field is the second type of knowledge called *heuristic knowledge,* which is the knowledge of good practice and good judgment in a field ... that a human expert acquires over years of work.[9].

Using all three kinds of knowledge, Edward Feigenbaum, the father of expert systems, developed a program called DENDRAL. It takes the data generated by a mass spectrograph and deduces from this data the molecular structure of the compound being analyzed. Another program, MYCIN, takes the results of blood tests such as the number of red cells, white cells, sugar in the blood, etc. and comes up with a diagnosis of which blood disease is responsible for this condition. It even gives an estimate of the reliability of its own diagnosis. Such programs give impressive performances.

And, indeed, isn't the success of expert systems just what the tradition would lead one to expect? If we agree with Feigenbaum that: "almost all the thinking that professionals do is done by reasoning... "[10] we can see that once computers are used for reasoning and not just computation they should be as good or better than we are at following rules for deducing conclusions from a host of facts. So we would expect that, if the rules that an expert has acquired from years of experience could be extracted and programmed, the resulting program would exhibit expertise. Again Feigenbaum puts the point boldly:

> [T]he matters that set experts apart from beginners, are symbolic, [and] inferential ... Experts build up a repertory of working rules of thumb, or "heuristics," that, combined with book knowledge, make them expert practitioners.[11]

Since each expert already has a repertory of rules in his mind, all the expert system builder need do is get the rules out and program them into the computer.

II. The Sources of Cognitivism.

We have seen that it is easy to trace cognitivism back to seventeenth century rationalism. But the roots of the cognitivist account of intelligence reach much further back into our tradition.

---

[9] Edward Feigenbaum and Pamela McCorduck, The Fifth Generation, Addison-Wesley Pub. Co., 1983, pp. 76-77.

[10] Ibid., p. l8.

[11] Ibid., p. 64.

As far as I can tell, the cognitivist account of skill grows out of two observations made by Socrates. Both were good phenomenology, but, like a good philosopher, he may have over-generalized them. He saw that experts can often explain why they do what they do, and that these explanations reveal principles from which the behavior in question can be seen to follow rationally. Generalizing these observations, Socrates claims in the Gorgias that a craft must have "principles of action and reason"[12] and, in the Laches that "that which we know we must surely be able to tell."[13] Let us look carefully at each of these claims.

The claim that a craft or t*echné* must be based on principles which can be articulated by the practitioners leads Socrates to rule out of account all forms of intuitive expertise which do not seem to be based on any principles at all. Cooking, for example, is "unable to render any account of the nature of the methods it applies."[14] It "goes straight to its end, nor even considers or calculates anything."[15] Socrates holds that such intuitive abilities are not crafts at all but mere knacks based on trial and error.[16]

Socrates shares the assumption that experts in a craft know principles of action and reason and that what they know they must be able to tell, with modern knowledge engineers. In its weak form, as just stated, the claim has a basis in the experience of experts. Some experts can at least tell the maxims of their craft. A maxim is a rule that requires some knowledge of the domain in order to be understood and applied. Michael Polanyi has stressed that much can be said but that what can be said will never be the whole story:

> Analysis may bring subsidiary knowledge into focus and formulate it as a maxim ... but such specification is in general not exhaustive. Although the expert diagnostician, taxonomist and cotton-classer can indicate their clues and formulate their maxims, they know many more things that they can tell...[17]

---

[12] Socrates, Gorgias, 501 a.

[13] Socrates, Laches, 190

[14] Gorgias, 465a.

[15] Ibid., 501a. Aristotle, on the contrary, stays close to the normal everyday phenomenon and so sees the immediate, intuitive response precisely as characteristic of the expert craftsman. "Art (*techné*) does not deliberate," he says in Physics, Bk. II, Ch. 8.

[16] This would seem to rule out, for example, chicken sexors who, without being able to articulate any reasons and without, as far as anyone has been able to tell, using any features (cues) or rules, are able to sort male and female one day-old chicks at a rate of 1000 an hour with 99.5% accuracy [John H. Lunn, "Chick Sexing", American Scientist, 36, 1948].

[17] Michael Polanyi, Personal Knowledge, Routledge, 1962 p. 88.

According to Polanyi, most experts can state rules of thumb they once followed -- rules that can help explain how they achieve their goals, but the rules make sense only to those already familiar with the relevant skill domain.

So far, so good. But Socrates seems to want much more. He seems to want to elicit rules or principles from experts in each craft domain that would enable *anyone* to acquire expertise in that domain. If we assume for the sake of this discussion, as Socrates assumes in his dialogue, Euthyphro, that Euthyphro, a religious prophet, is an expert at recognizing piety, and Socrates is not, it looks like Socrates is asking for more than Euthyphro's piety recognizing *maxim* when he says: "I want to know what is characteristic of piety ... to use as a standard whereby to judge your actions and those of other men."[18] He seems rather to want a *strict rule* that could be used even by non-experts.

Euthyphro's first response to this demand is like that of any expert. He gives Socrates examples from his field of expertise, in this case mythical situations in which men and gods have done things that everyone considers pious. Socrates then makes his usual demand that Euthyphro tell him the features and rule he uses in recognizing these cases as examples of piety. And Euthyphro, like any expert examined by Socrates, produces various maxims, which, if taken as definitions or strict rules, as Socrates takes them, do not hold up under cross-examination.[19] That Socrates demands context free features seems clear in the Laches where he asks Laches, presumably an expert on courage, "What is that common quality, which is the same in all cases, and which is called courage?"[20] But no expert can supply such common features and rules based on them. This leads Socrates to the famous conclusion that since no one can state consistent, context-free principles that provide the rationale for his actions, no one *knows* anything at all.

This is where Plato came to the aid of Socrates. Perhaps experts were operating on principles they could not easily articulate, Plato suggested. Experts, at least in areas involving non-empirical knowledge such as morality and mathematics, had, in another life, Plato claimed, learned the principles involved, but they had forgotten them. The role of the philosopher was to help such moral and mathematical experts recall the principles on which they were acting. These principles would ground the knowledge of the skill. Such knowledge must be "fastened by the reasoning of cause and effect" and "this is done by `recollection'."[21]

---

[18] Euthyphro, 6e3 - 6.

[19] Maxims cannot be taken as strict rules because they have *ceteris paribus* conditions, that is, that they apply “everything else being equal.” This need for judgment does not detract from their usefulness to competent performers, but makes them useless to the beginner who, having little or no experience in the domain, needs to follow a strict rule defined over context-free features accessible to any person or machine.

[20] Laches, 191e

[21] Meno, 98a.

Knowledge engineers, generalizing the doctrine of recollection, would now say that the rules experts -- even experts in empirical domains -- use, have been put in a part of their mental computers where they work automatically. Thus, Feigenbaum tells us:

> When we learned how to tie our shoes, we had to think very hard about the steps involved ... Now that we've tied many shoes over our lifetime, that knowledge is "compiled," to use the computing term for it; it no longer needs our conscious attention.[22]

On this modern version of Platonic recollection, the rules the expert followed consciously at first still functioning in the expert's mind whether he is conscious of them or not. After all, how else could one account for the fact that the expert can perform the task?

Even Polanyi, who as far as I know was the first to see the importance of the difference between strict rules and maxims, accepts the cognitivist account of unconscious rule-following. He tells us that:

> By watching the master ... the apprentice unconsciously picks up the rules of the art, including those that are not explicitly known to the master himself.[23]

This is no lapse but an essential part of Polanyi's crypto-cognitivist view concerning the relation of rules and skilled action:

> In performing a skill we are ... acting on certain *premisses* of which we are focally ignorant, but which we *know subsidiarily* as part of our mastery of that skill, and which we may get to know focally by analyzing the way we achieve success ... in the skill in question.[24] .

Polanyi is thus a phenomenologist concerning the awareness of skilled performers and a kind of cognitivist concerning the unconscious causes of skilled performance. He is not a full-fledged cognitivist, however, since he holds that the premisses from which the unconscious mind deduces actions are maxims not strict rules.

Such a cognitivist view gives the knowledge engineer a venerable task. As Feigenbaum explains: "[A]n expert's knowledge is often ill-specified or incomplete because the expert himself doesn't always know exactly what it is he knows about his domain."[25] So the knowledge engineer has to help him recollect what he once knew.

---

[22] Feigenbaum, op. cit., p. 55.

[23] Polanyi, op. cit., p. 53.

[24] Polanyi, op. cit., p. 162. (My italics.)

[25] Feigenbaum, op. cit., p. 85.

> [An expert's] knowledge is currently acquired in a very painstaking way; individual computer scientists work with individual experts to explicate the expert's heuristics -- to mine those jewels of knowledge out of their heads one by one ... [T]he problem of knowledge acquisition is the critical bottleneck in artificial intelligence.[26] .

Thanks to Feigenbaum and his colleagues, we have a new name for what Socrates and Plato considered the philosopher's task: *knowledge acquisition research.*[27]

But nothing has changed. When Feigenbaum suggests to an expert the rules the expert seems to be using, he gets a Euthyphro-like response. "That's true, but if you see enough patients/rocks/chip designs/instruments readings, you see that it isn't true after all,"[28] and Feigenbaum comments with Socratic annoyance: "At this point, knowledge threatens to become ten thousand special cases."[29] Still the doctrine of recollection assures the knowledge engineer that the principles must be there in the expert's mind, the expert is simply poor at remembering them.

One important question remains: Is Plato a partial cognitivist like Polanyi or a full-fledged computer-cognitivist like Feigenbaum? For Plato, are the unconscious principles that serve as premisses for deducing actions, maxims or strict rules? We have seen that Socrates seems to think that experts should be able to articulate strict rules that would enable *anyone* to share their expertise. That would make Socrates a full-fledged cognitivist, about skills at least. But the question of Socrates' intellectualism is subject to debate. Terrence Irwin, indeed, claims that Socrates is a pure intellectualist who holds that the rules experts use must fully capture their expertise. "Socrates demands an explicit understanding of the principles guiding (not necessarily explicitly) our application of `just' or `pious', so that we can see whether they are consistent and justifiable, and can appeal to them to project our judgments to new cases."[30] Alexander Nehamas, however, has argued that Socrates, the son of a stonemason, would surely know that the rules a craftsman can articulate are not sufficiently explicit and complete to convey the craft to an outsider, rather they would be Polanyian maxims. "Socrates, himself a statuary and a statuary's son[31], knew perfectly well that in ancient Athens the crafts were most often transmitted along with their `secrets within the family from generation to generation'."[32]

---

[26] Ibid., pp. 79-80.
[27] Ibid., p. 79.
[28] Ibid., p. 82.
[29] Ibid., p. 82.
[30] Terrence Irwin, Plato's Moral Theory, Oxford Press, 1977, p. 65.
[31] D.L. V.I.18
[32] Alexander Nehamas, "Socratic Intellectualism", Proceeding of the Boston Area Colloquium in Ancient Philosophy, Vol. II, ed. John Cleary, 1986, p. 299.

Who is right, Irwin or Nehamas? As we have seen, Socrates does seem to hold in the Euthyphro that if he had Euthyphro's rules he too would be an expert on piety. Still, since this example comes from the domain of morality where Socrates is surely to some extent already an initiate, it would be consistent with Nehamas's view that Socrates held that "only one virtuous person can recognize another."[33] The same holds for the definition of courage, since Socrates was acknowledged to have been a brave soldier. There is no evidence I know of that Socrates thought that an expert must be able to articulate rules that, in principle, would enable an outsider following these rules to behave like an expert.

Plato never explicitly holds this full-fledged cognitivist view either, but it seems to me that the doctrine of recollection as Plato uses it, i.e. not only as an account of why moral and mathematical experts cannot easily articulate their rules, but also as an account of how they learned these rules in the first place, commits him to pure cognitivism. That is, it commits him to the view that the principles underlying all expertise must be strict rules, not merely maxims. If this is, indeed, the case, Socrates and Plato would turn out to share responsibility for cognitivism.

The argument is simple. If a soul starting from scratch is supposed to have acquired expert knowledge in a specific domain -- in Plato's example in the Meno, mathematics -- by seeing in another world the principles underlying the skill domain, then the principles learned must have been the rank beginner's strict rules applicable to features rather than the advanced beginner's maxims requiring prior familiarity with the skill domain. Since all the expert has, and all the expert needs, according to Plato, are the principles he has acquired in the other world, before he had any familiarity with anything in this world, this comes to the claim that strict rules are sufficient to generate expertise, and this is precisely the pure cognitivist claim.

To sum up: What experts actually say suggests that they can at best formulate maxims that they presumably remember from when they learned them. These maxims can still be of use to others with some skill in the relevant domain. Socrates does not seem to have contradicted this observation, but he makes two moves toward cognitivism: (1) he rules out as trial and error skills such as cooking which are learned solely by apprenticeship; (2) he assumes that all true skills are based on principles; and (3) he claims that these principles must be expressible by those who exhibit the skill.

Plato, however, went further and provided an account of why experts were in fact so poor at articulating the principles underlying their performance, and in so doing developed the theory of recollection which committed him to the view that expertise is generated by strict rules.

Plato's account did not apply to everyday skills, like cooking. It took two thousand years before Leibniz, who was influenced by Descartes, boldly generalized the Platonic account to all forms of intelligent activity:

---

[33] Ibid., p. 308.

> [T]he most important observations and turns of skill in all sorts of trades and professions are as yet unwritten. This fact is proved by experience when, passing from theory to practice, we desire to accomplish something. *Of course, we can also write up this practice, since it is at bottom just another theory more complex and particular ...*[34]

This sets things up for the last move, which makes the rules and the elements to which they apply explicitly syntactic. Thus for modern cognitive scientists all mental activity is based upon rational calculations of the sort which can be implemented by a computer program.

[34] Leibniz, Selections, ed. Philip Wiener (New York: Scribner, 1951), p. 48. My italics.