This document is subject to the following notice:

**NOTICE:** This material may be protected by copyright law.

The copyright law of the United States (Title 17, U.S. Code) governs the making of photocopies or other reproductions of copyrighted material; the person using this equipment is liable for any infringement.
Moreover, if Winston claims to "shed some light on [the question:] How do we recognize examples of various concepts?" (1975, p. 157), his theory of concepts as definitions must, like any psychological theory, be subject to empirical test. It so happens that, contrary to Winston's claims, recent evidence collected and analyzed by Eleanor Rosch on just this subject shows that human beings are not aware of classifying objects as instances of abstract rules, but rather group objects as more or less distant from an imagined paradigm. This does not exclude the possibility of unconscious processing, but it does highlight the fact that there is no empirical evidence at all for Winston's formal model. As Rosch puts it:

Many experiments have shown that categories appear to be coded in the mind neither by means of lists of each individual member of the category, nor by means of a list of formal criteria necessary and sufficient for category membership, but, rather, in terms of a prototype of a typical category member. The most cognitively economical code for a category is, in fact, a concrete image of an average category member. (1977, p. 30)

One paradigm, it seems, is worth a thousand rules. As we shall soon see, one of the characteristics of the next phase of work in AI is to try to take account of the implications of Rosch's research.

Meanwhile, what can we conclude concerning AI's contribution to the science of psychology? No one can deny Minsky and Papert's claim that "computer science has brought a flood of ideas, well-defined and experimentally implemented, for thinking about thinking" (1973, p. 25). But all of these ideas can be boiled down to ways of constructing and manipulating symbolic descriptions, and, as we have seen, the notion that human cognition can be explained in terms of formal representations does not seem at all obvious in the face of actual research on perception, and everyday concept formation. Even Minsky and Papert show a commendable new modesty. They as much as admit that AI is still at the stage of astrology, and that the much heralded games does not seem at all distant from an imagined paradigm. This does not exclude the possibility of unconscious processing, but it does highlight the fact that there is no empirical evidence at all for Winston's formal model. As Rosch puts it:

Many experiments have shown that categories appear to be coded in the mind neither by means of lists of each individual member of the category, nor by means of a list of formal criteria necessary and sufficient for category membership, but, rather, in terms of a prototype of a typical category member. The most cognitively economical code for a category is, in fact, a concrete image of an average category member. (1977, p. 30)

Just as astronomy succeeded astrology, following Kepler's discovery of planetary regularities, the discoveries of these many principles in empirical explorations on intellectual processes in machines should lead to a science, eventually. (1973, p. 25)

Happily, "should" has replaced "will" in their predictions. Indeed, this period's contribution to psychology suggests an even more modest

hope: As more psychologists like Goldmeier are frustrated by the limitations of formal computer models, and others turn to investigating the function of images as opposed to symbolic representations, the strikingly limited success of AI may come to be seen as an important disconfirmation of the information-processing approach.

Before concluding our discussion of this research phase, it should be noted that some problem domains are (nearly enough) micro-worlds already; so they lend themselves to AI techniques without the need for artificial restrictions, and, by the same token, nongeneralizability is not the same kind of Waterloo. Game playing, particularly chess, is the most conspicuous example. Though some extravagant early predictions were not fulfilled, large computers now play fairly high caliber chess, and small machines that play creditable amateur games are being marketed as toys. But game players are not the only examples; excellent programs have been written for analyzing certain kinds of mass-spectroscopy data (Feigenbaum 1977), and for assisting in the diagnosis and treatment of some diseases (Shortliffe 1976). Such work is both impressive and important; but it shouldn't give the wrong impression. In each case, it succeeds because (and to the extent that) the relevant domain is well circumscribed in advance, with all the significant facts, questions, and/or options already laid out, and related by a comparatively small set of explicit rules—in short, because it's a micro-world. This is not to belittle either the difficulty or the value of spelling out such domains, or designing programs which perform well in them. But we should not see them as any closer to the achievement of genuine artificial intelligence than we do the "blocks-world" programs. In principle, interpreting mass spectrograms or batteries of specific symptoms has as little to do with the general intelligence of physicists and physicians, as disentangling vertices in projections of polyhedra does with vision. The real, theoretical problems for AI lie elsewhere.

2 The later seventies: knowledge representation

In roughly the latter half of the decade, the problem of how to structure and retrieve information, in situations where anything might be relevant, has come to the fore as the "knowledge-representation problem". Of course, the representation of knowledge was always a central problem for work in AI, but earlier periods were characterized by an attempt to repress it by seeing how much could be done with as little
knowledge as possible. Now, the difficulties are being faced. As Roger Schank of Yale recently remarked:

Researchers are starting to understand that *tour de force* in programming is interesting but non-extendable ... the AI people recognize that how people use and represent knowledge is the key issue in the field. (1977, pp. 1007–1008)

Papert and Goldstein explain the problem:

It is worthwhile to observe here that the goals of a knowledge-based approach to AI are closely akin to those which motivated Piaget to call himself an “epistemologist” rather than a psychologist. The common theme is the view that the process of intelligence is determined by the knowledge held by the subject. The deep and primary questions are to understand the operations and data structures involved. (1975/76, p. 7)

Another memorandum illustrates how ignoring the background knowledge can come back to haunt one of AI’s greatest tricks in the form of nongeneralizability.

Many problems arise in experiments on machine intelligence because things obvious to any person are not represented in any program. One can pull with a string, but one cannot push with one. One cannot push with a thin wire, either. A taut inextensible cord will break under a very small lateral force. Pushing something affects first its speed, only indirectly its position! Simple facts like these caused serious problems when Charniak attempted to extend Bobrow’s STUDENT program to more realistic applications, and they have not been faced up to until now. (Papert and Minsky 1973, p. 77)

The most interesting current research is directed toward the underlying problem of developing new, flexible, complex data types which will allow the representation of background knowledge in larger, more structured units.

In 1972, drawing on Husserl’s phenomenological analysis, it pointed out that it was a major weakness of AI that no programs made use of expectations (1972/92, pp. 153f/241f and 162/250). Instead of modeling intelligence as a passive receiving of context-free facts into a structure of already stored data, Husserl thinks of intelligence as context-determined, goal-directed activity—as a search for anticipated facts. For him the *noema*, or mental representation of any type of object, provides a context or “inner horizon” of expectations or predefinitions for structuring the incoming data: a “rule governing the other consciousness of [the object] as identical—possible as amplifying essentially predefined types” (Husserl 1960, p. 53). As explained in chapter 7:

We perceive a house, for example, as more than a façade—as having some sort of back—some inner horizon. We respond to this whole object first and then, as we get to know the object better, fill in the details as to inside and back. (p. 153/241)

The *noema* is thus a symbolic description of all the features which can be expected with certainty in exploring a certain type of object—features which remain, as Husserl puts it, “inviolably the same: as long as objectivity remains intended as *this* one and of this kind” (p. 51)—“predefinitions” of those properties which are possible but not necessary features of this type of object.

Frames and knowledge representation

In 1974, Minsky proposed a new data structure remarkably similar to Husserl’s for representing everyday knowledge.

A frame is a data structure for representing a stereotyped situation, like being in a certain kind of living room, or going to a child’s birthday party ...

We can think of a frame as a network of nodes and relations. The “top levels” of a frame are fixed, and represent things that are always true about the supposed situation. The lower levels have *many terminals*—“slots” that must be filled by specific instances or data. Each terminal can specify conditions its assignments must meet ...

Much of the phenomenological power of the theory hinges on the inclusion of expectations and other kinds of presumptions. A frame’s terminals are normally already filled with “default” assignments. (1974 [chapter 5 of this volume], pp. 1f [111f])

Minsky’s model of a frame, the “top level” is a developed version of in Husserl’s terminology “remains inviolably the same” in the representation, and Husserl’s predefinitions have been made precise with “default assignments”—additional features that can normally be left. The result is a step forward in AI techniques from a passive information processing to one which tries to take account of the context of the interactions between a knower and his world.
constitution—that is, "explicating" the noemats for all types of objects—as the beginning of progress toward philosophy as a rigorous science; and Patrick Winston has hailed Minsky's proposal as "the ancestor of a wave of progress in AI" (1975, p. 16). But Husserl's project ran into serious trouble and there are signs that Minsky's may too.

During twenty years of trying to spell out the components of the noema of everyday objects, Husserl found that he had to include more and more of what he called the "outer horizon", a subject's total knowledge of the world:

To be sure, even the tasks that present themselves when we take single types of objects as restricted clues prove to be extremely complicated and always lead to extensive disciplines when we penetrate more deeply. That is the case, for example, with a transcendental theory of the constitution of a spatial object (to say nothing of nature) as such, of psycho-physical being and humanity as such, cultures as such.

(1960, pp. 54-55)

He sadly concluded at the age of seventy-five that he was "a perpetual beginner" and that phenomenology was an "infinite task"—and even that may be too optimistic. His successor, Heidegger, pointed out that since the outer horizon or background of cultural practices was the condition of the possibility of determining relevant facts and features and thus prerequisite for structuring the inner horizon, as long as the cultural context had not been clarified, the proposed analysis of the inner horizon of the noema could not even claim progress.

There are hints in the frame paper that Minsky has embarked on the same misguided "infinite task", that eventually overwhelmed Husserl.

Just constructing a knowledge base is a major intellectual research problem ... We still know far too little about the contents and structure of common-sense knowledge. A "minimal" common-sense system must "know" something about cause and effect, time, purpose, locality, process, and types of knowledge ... We need a serious epistemological research effort in this area. (p. 74 [138]).

Minsky's naiveté and faith are astonishing. Philosophers from Plato to Husserl, who uncovered all these problems and more, have carried on serious epistemological research in this area for two thousand years without notable success. Moreover, the list Minsky includes in this passage deals only with natural objects, and their positions and interactions. As Husserl saw, intelligent behavior also presupposes a background of cultural practices and institutions. Observations in the frame paper such as: "Trading normally occurs in a social context of law, trust, and convention. Unless we also represent these other facts, most trade transactions will be almost meaningless" (p. 34 [117]) show that Minsky has understood this too. But Minsky seems oblivious to the hand-waving optimism of his proposal that programmers rush in where philosophers such as Heidegger fear to tread, and simply make explicit the totality of human practices which pervade our lives as water encompasses the life of a fish.

To make this essential point clear, it helps to take an example used by Minsky and look at what is involved in understanding a piece of everyday equipment as simple as a chair. No piece of equipment makes sense by itself. The physical object which is a chair can be defined in isolation as a collection of atoms, or of wood or metal components, but such a description will not enable us to pick out chairs. What makes an object a chair is its function, and what makes possible its role as equipment for sitting is its place in a total practical context. This presupposes certain facts about human beings (fatigue, the ways the body bends), and a network of other culturally determined equipment (tables, floors, lamps) and skills (eating, writing, going to conferences, giving lectures). Chairs would not be equipment for sitting if our knees bent backwards like those of flamingos, or if we had no tables, as in traditional Japan or the Australian bush.

Anyone in our culture understands such things as how to sit on kitchen chairs, swivel chairs, folding chairs; and in arm chairs, rocking chairs, deck chairs, barbers' chairs, sedan chairs, dentists' chairs, basket chairs, reclining chairs, wheel chairs, slung chairs, and beanbag chairs—as well as how to get off/out of them again. This ability presupposes a repertoire of bodily skills which may well be indefinitely large, since there seems to be an indefinitely large variety of chairs and of successful (graceful, comfortable, secure, poised) ways to sit on/in them. Moreover, understanding chairs also includes social skills such as being able to sit appropriately (sedately, demurely, naturally, casually, sloppily, provocatively) at dinners, interviews, desk jobs, lectures, auditions, concerts (intimate enough for there to be chairs rather than seats), and in waiting rooms, living rooms, bedrooms, courts, libraries, and bars (of the sort sporting chairs, not stools).

In light of this amazing capacity, Minsky's remarks on chairs in his frame paper seem more like a review of the difficulties than even a hint
of how AI could begin to deal with our common sense understanding in this area.

There are many forms of chairs, for example, and one should choose carefully the chair-description frames that are to be the major capitals of chair-land. These are used for rapid matching and assigning priorities to the various differences. The lower priority features of the cluster center then serve ... as properties of the chair types ...

(p. 52 [132]; emphasis added)

There is no argument why we should expect to find elementary context-free features characterizing a chair type, nor any suggestion as to what these features might be. They certainly cannot be legs, back, seat, and so on, since these are not context-free characteristics defined apart from chairs which then "cluster" in a chair representation; rather, legs, back, and the rest, come in all shapes and variety and can only be recognized as aspects of already recognized chairs. Minsky concludes:

Difference pointers could be "functional" as well as geometric. Thus, after rejecting a first try at "chair" one might try the functional idea of "something one can sit on" to explain an unconventional form.

But, as we already saw in our discussion of Winston's concept-learning program, a function so defined is not abstractable from human embodied know-how and cultural practices. A functional description such as "something one can sit on" treated merely as an additional context-free descriptor cannot even distinguish conventional chairs from saddles, thrones, and toilets. Minsky concludes:

Of course, that analysis would fail to capture toy chairs, or chairs of such ornamental delicacy that their actual use would be unthinkable. These would be better handled by the method of excuses, in which one would bypass the usual geometrical or functional explanation in favor of responding to contexts involving art or play.

(emphasis added)

This is what is required all right; but by what elementary features are these contexts to be recognized? There is no reason at all to suppose that one can avoid the difficulty of formally representing our knowledge of chairs by abstractly representing even more holistic, concrete, culturally determined, and loosely organized human practices such as art and play.

Minsky in his frame article claims that "the frame idea ... is in the tradition of ... the 'paradigms' of Kuhn" (p. 3 [113]); so it's appropri-
Indeed, Kuhn sees his book as raising just those questions which Minsky refuses to face.

Why is the concrete scientific achievement, as a locus of professional commitment, prior to the various concepts, laws, theories, and points of view that may be abstracted from it? In what sense is the shared paradigm a fundamental unit for the student of scientific development, a unit that cannot be fully reduced to logically atomic components which might function in its stead?

(p. 11; emphasis added)

Although research based on frames cannot deal with this question, and so cannot account for common-sense or scientific knowledge, the frame idea did bring the problem of how to represent our everyday knowledge into the open in AI. Moreover, it provided a model so vague and suggestive that it could be developed in several different directions. Two alternatives immediately presented themselves: either to use frames as part of a special-purpose micro-world analysis dealing with common-sense knowledge, as if everyday activity took place in preanalyzed specific domains, or else to try to use frame structures in “a no-tricks basic study” of the open-ended character of everyday know-how. Of the two most influential current schools in AI, Roger Schank and his students at Yale have tried the first approach. Winograd, Bobrow, and their group at Stanford and Xerox, the second.

2.2 Scripts and primitive actions

Schank’s version of frames are called “scripts”. Scripts encode the essential steps involved in stereotypical social activities. Schank uses them to enable a computer to “understand” simple stories. Like the micro-world builders, Schank believes he can start with isolated stereotypical situations described in terms of primitive actions and gradually work up from there to all of human life.

To carry out this project, Schank invented an event description language consisting of eleven primitive acts such as: ATRANS—the transfer of an abstract relationship such as possession, ownership, or control; PTRANS—the transfer of physical location of an object; INGEST—the taking of an object by an animal into the inner workings of that animal; and so forth. (1975a, p 39); and from these primitives he builds gamelike scenarios which enable his program to fill in gaps and pronoun reference in stories.

Such primitive acts, of course, make sense only when the context is already interpreted in a specific piece of discourse. Their artificiality can easily be seen if we compare one of Schank’s context-free primitive acts to real-life actions. Take PTRANS, the transfer of physical location of an object. At first it seems an interpretation-free fact if ever there were one. After all, either an object moves or it doesn’t. But in real life things are not so simple; even what counts as physical motion depends on our purposes. If someone is standing still in a moving elevator on a moving ocean liner, is his going from A to B deck a PTRANS? What about when he is just sitting on B deck? Are we all PTRANS-ing around the sun? Clearly the answer depends on the situation in which the question is asked.

Such primitives can be used, however, to describe fixed situations or scripts, once the relevant purposes have already been agreed upon. Schank’s definition of a script emphasizes its predetermined, bounded, gamelike character.

We define a script as a predetermined causal chain of conceptualizations that describe the normal sequence of things in a familiar situation. Thus there is a restaurant script, a birthday-party script, a football-game script, a classroom script, and so on. Each script has a minimum number of players and objects that assume certain roles within the script ... (Each primitive action given stands for the most important element in a standard set of actions.

(1975b, p. 131; emphasis added)

His illustration of the restaurant script spells out in terms of primitive actions the rules of the restaurant game:

**Script:** restaurant.

**Roles:** customer; waitress; chef; cashier.

**Reason:** to get food so as to go down in hunger and up in pleasure.

**Scene 1**, entering:

PTRANS—go into restaurant
MBUILD—find table
PTRANS—go to table
Move—sit down

**Scene 2**, ordering:

ATrans—receive menu
ATTEND—look at it
MBUILD—decide on order
MTRANS—tell order to waitress
Scene 3, eating:
ATrans—receive food
INGEST—eat food

Scene 4, exiting:
MTTrans—ask for check
ATrans—give tip to waitress
PTTrans—go to cashier
ATrans—give money to cashier
PTTrans—go out of restaurant

(1975b, p. 131)

No doubt many of our social activities are stereotyped, and there is nothing in principle misguided in trying to work out primitives and rules for a restaurant game, the way the rules of Monopoly are meant to capture a simplified version of the typical moves in the real estate business. But Schank claims that he can use this approach to understand stories about actual restaurant-going—that, in effect, he can treat the sub-world of restaurant going as if it were an isolated micro-world. To do this, however, he must artificially limit the possibilities; for, as one might suspect, no matter how stereotyped, going to the restaurant is not a self-contained game but a highly variable set of behaviors which open out into the rest of human activity. What “normally” happens when one goes to a restaurant can be preselected and formalized by the programmer as default assignments; but the background has been left out, so that a program using such a script cannot be said to understand going to a restaurant at all.

This can easily be seen by imagining a situation that deviates from the norm. What if, when one tries to order, one finds that the item in question is not available, or before paying one finds that the bill is added up wrongly? Of course, Schank would answer that he could build these normal ways restaurant going breaks down into his script. But there are always abnormal ways everyday activities can break down: the juke box might be too noisy, there might be too many flies on the counter, or, as in the film Annie Hall, in a New York delicatessen one’s girl friend might order a pastrami sandwich on white bread with mayonnaise. When we understand going to a restaurant we understand how to cope with even these abnormal possibilities because going to a restaurant is part of our everyday activities of going into buildings, getting things we want, interacting with people, and so on.

To deal with this sort of objection, Schank has added some general rules for coping with unexpected disruptions. The general idea is that in a story “it is usual for nonstandard occurrences to be explicitly mentioned” (Schank and Abelson 1977, p. 51); so the program can spot the abnormal events and understand the subsequent events as ways of coping with them. But here we can see that dealing with stories allows Schank to bypass the basic problem, since it is the author’s understanding of the situation which enables him to decide which events are disruptive enough to mention.

This ad hoc way of dealing with the abnormal can always be revealed by asking further questions; for the program has not understood a restaurant story the way people in our culture do, until it can answer such simple questions as: When the waitress came to the table, did she wear clothes? Did she walk forward or backward? Did the customer eat his food with his mouth or his ear? If the program answers, “I don’t know”, we feel that all of its right answers were tricks or lucky guesses and that it has not understood anything of our everyday restaurant behavior. The point here, and throughout, is not that there are subtle things human beings can do and recognize which are beyond the low-level understanding of present programs, but that in any area there are simple taken-for-granted responses central to human understanding, lacking which a computer program cannot be said to have any understanding at all. Schank’s claim, then, that “the paths of a script are the possibilities that are extant in a situation” (1975b, p. 132) is insidiously misleading. Either it means that the script accounts for the possibilities in the restaurant game defined by Schank, in which case it is true but uninteresting; or he is claiming that he can account for the possibilities in an everyday restaurant situation which is impressive but, by Schank’s own admission, false.

Real short stories pose a further problem for Schank’s approach. In a script what the primitive actions and facts are is determined beforehand, but in a short story what counts as the relevant facts depends on the story itself. For example, a story that describes a bus trip contains in its script that the passenger thanks the driver (a Schank example). But the fact that the passenger thanked the driver would not be important in a story in which the passenger simply took the bus as a part of a longer journey, while it might be crucially important if the story concerned a misanthrope who had never thanked anyone before, or a very law-abiding young man who had courageously broken the prohibition against speaking to drivers in order to speak to the attractive woman driving the bus. Overlooking this point, Schank claimed at a recent meeting that his program, which can extract death statistics from