Computers in Human Behavior, Vol. 2, pp. 215–234, 1986
Printed in the U.S.A.
0747-5632/86 $3.00 + .00

Anthropomorphism and Mechanomorphism: Two Faces of the Human Machine

L.R. Caporael

Rensselaer Polytechnic Institute

Abstract — *This paper explores the ambiguity of the "human machine." It suggests that anthropomorphism results from a "default schema" applied to phenomena, including machines, that a perceiver finds otherwise inexplicable. Mechanomorphism, the attribution of machine characteristics to humans, is a culturally derived metaphor that presently dominates cognitive science. The relationships between anthropomorphism and mechanomorphism pose a special difficulty for the question, "Can machines think?" Does a positive response reflect a cognitive bias on the part of the perceiver or a genuine attribute of the computer? The problem is illustrated for Turing's "imitation game" for thinking machines, and a strategy for constraining anthropomorphic attributions is proposed.*

Anthropomorphism is the ascription of human characteristics to nonhuman entities. We think of anthropomorphism in the context of primitive peoples who attribute human forms, emotions, motivations, and activities to the wind, sun, moon, trees, rivers and animals, use them as causal explanations for otherwise inexplicable events, and attempt to manipulate them with social mechanisms such as entreaty or threat. Anthropomorphized, nonhuman entities became social entities.

Anthropologists hold that anthropomorphic thought develops from animism (the belief that all things have a spirit or soul), legend, and the need to have visual images of gods. Thus, anthropomorphism is commonly considered to be the quaint and archaic activity of people far removed in time and place from the complex functioning of contemporary technological society. But anthropomorphism is widespread in modern life, so common that we take it for granted and fail to react to its peculiarity. Many people have entreated, coaxed, and threatened a recalcitrant car, have inferred human feelings and motivations as causal explanations for otherwise inexplicable malfunctioning, and in short, entered (briefly or extensively) into social relations with their automobiles. Anthropomorphized, even unintelligent machines may become social entities. This assertion is not invalidated by the claim that primitive peoples held a conviction that nonhuman entities really do have

---

Preparation of this article was supported in part by National Science Foundation Grant IST 8110625.

I am indebted to Glen Culbertson, Robyn Dawes, C.M. Heyes, and Warren Thorngate for their comments and criticisms. I also thank the Department of Anthropology at the American Museum of Natural History, New York City, for their hospitality during the completion of this work.

Requests for reprints should be sent to L.R. Caporael, Department of Science and Technology Studies, Rensselaer Polytechnic Institute, Troy, NY 12180.

human attributes whereas contemporary people are rationally aware that cars, ships and the like do not. Rather, the claim draws attention to the peculiarity of contemporary anthropomorphism, which occurs despite objective knowledge.

Anthropomorphizing "intelligent" machines (e.g., computers and robots) is most often grounds for amusement, certainly nothing meriting mulling as a scientific problem.[1] Hence when scientists do mention anthropomorphism, their treatment is sketchy and destitute of data. Sometimes the issue is its desirability (e.g., Boden, 1981). Occasionally, antecedents are proposed. Branscomb (1979) asserts that a computer that is easy to use "creates in us a warm feeling, and may be described as 'friendly,' 'faithful,' or 'obedient.'" According to him, these are desirable anthropomorphic qualities that can be obtained through careful design, planning, and testing; that is, they are a function of machine characteristics. In contrast, Weizenbaum (1976) writes that anthropomorphism is a psychological consequence of an emotional tie to the machine as an extension of the body. He further claims that naive users are more likely to anthropomorphize than are sophisticated users. But Minsky's (1967) description of a programmer watching the output of a poorly documented program "as if the program were an individual whose range of behavior is uncertain" (p. 120), suggests that there are conditions under which even sophisticated users may respond to a machine as if it were human.

The human machine has a delightful ambiguity—what better way to capture the Cartesian duality of our age? It points to computers and robots as "evocative objects" (Turkle, 1980, 1984), and distinguishes them from other social transformational technologies such as the automobile, air conditioner, or television. These we may be tempted to anthropomorphize when they malfunction, but there is no temptation to think of ourselves or others as being like automobiles, air conditioners, or televisions. Not so with computers and robots. If anthropomorphism, referring to machine-as-human, is one face of the human machine, we might allow "mechanomorphism," the human-as-machine, to be its other face. Mechanomorphism is the attribution of characteristics of machines to humans. It is a widely used cultural metaphor and a dominating scientific one.

Scientists of all stripes might be willing, indeed eager, to abandon the human machine to the disciplines concerned with broad interpretive frameworks of human experience (e.g., Lewis, 1963; Warwick, 1980) as if they could keep for themselves just the objective properties of the computer. But even while scientists and philosophers feverishly debate whether machines can think, the conditions under which informal attributions of "thought" or "intelligence" are made remain unclear. The latter issues differentiate the philosopher's ontological question—can machines think?—from the issues considered here—what makes people think machines can think?

For the sake of argument, I will make a bald statement here: machines do not think. In the following sections, I suggest anthropomorphism can be conceived as a schema used by the general public, or conventional community, and that mechanomorphism can be conceived as a schema (albeit an elaboration of anthropomorphism) used by the scientific community, especially by researchers

[1] A notable exception is the untranslated work of Masahiro Mori of the Robotics Department, Tokyo Institute of Technology. He proposes that the closer the physical resemblance of a robot to a human, the more affection it will arouse—up to a point. At that point begins the "uncanny valley," a psychological space filled with objects that are *too* human-like, yet lacking some essential human attributes. They arouse fear and distrust (see Reichardt, 1978).

in artificial intelligence and cognitive science. The machine-as-human schema can be encompassed within mainstream research areas in social psychology, and it can also be used as a vehicle for expanding our understanding of some interesting psychological phenomena. Its most significant contribution, however, may be in elucidating critical issues in social cognition, including the distinctions between social and nonsocial objects: the computer can be programmed to be more or less "social." While the computer overcomes certain technical obstacles in psychological research, mechanomorphism poses deeper conceptual ones. Once these are cleared away, it is possible to re-examine anthropomorphism. I suggest that certain anthropomorphic usages reflect an inherent cognitive heuristic.

If there exists some fundamental anthropomorphic bias, those who would attempt to answer the question, "Can machines think?" face a conundrum. Does a positive response reflect a cognitive bias on the part of the perceiver or a genuine attribute of the machine? The only question that we might truly decide is the circumstances in which it is a benefit or a liability. Behavioral research should guide the construction of "traps" to constrain anthropomorphic attributions when they are counterproductive.

I hope to illustrate that between research on the practical exigencies of computer use (the domain of many social, organizational, and human factors researchers) and research on information processing (the domain of cognitive science and artificial intelligence) lies a rich area concerned with affective or social aspects of human-machine interaction. This "social psychology of computing" (Caporael & Thorngate, 1984) not only enlarges the scope of applied and basic research on human-computer interaction, but also engages behavioral scientists for a question both old and new: What makes people think that *other people* (and machines) think?

SKETCHES OF TWO SCHEMATA

The scientific community, especially workers in artificial intelligence (AI) and cognitive psychology, can be broadly distinguished from the conventional community by what each perceives as the least understood or most problematic component of the human-machine relation. For the scientific community, the human is viewed as the least understood. At some level, be it in programming or theoretical abstraction, the activities of computers and automata are defined. Psychologically, the human is prone to unpredictability and unreliability far beyond that realized in intelligent machines. For the conventional community, the machine is seen as the least understood. At some level, illusory or otherwise, people have their theoretical abstractions that explain human behavior (Heider, 1958). It is the machine that is perceived to be too unpredictable and unreliable—the computer loses paychecks, makes billing errors, or, ubiquitously, "goes down."

The scientific and conventional communities have each developed a general schema, or model, that permits accommodation and assimilation of these problematic components so that the respective activities of the two communities can be maintained. Within the scientific community, the general schema is that the human is like a computer or a serial information processor. One of the more extreme formulations of the human-as-machine schema is Newell and Simon's (1972): "men and computers are merely two different species of a more abstract genus called 'information processing systems.' The concepts that describe abstract information processing systems must, perforce, describe any particular examples of such systems" (p. 230). Within the conventional community, it appears that the general schema is that the machine is like a human. For example, *Newsweek* describes an industrial robot, named "Clyde" by the employees, that malfunctioned and was described successively in the company newsletter as "sick," then "cured," and finally when it was returned to operation, was greeted by employees with a welcome back party (Nicholson, 1979). A later *Time* article, quite infamous now, substituted the computer for the usual "Man of the Year." An article on "robots that think," in *Popular Science*, reports a malfunctioning Unimation PUMA with two human figures painted on its arm. It had injured some workers that had gotten too close to it, and was named "Killer" (Schefter, 1980).

Of course, no scientist really believes the human is a machine (in the conventional sense) and no lay person really believes a machine is human. Rather, these are convenient fictions that permit "business as usual," and, as all such casual fictions, they are rarely opened for re-reading.

The Conventional Community

At first glance, the conventional community's human machine might seem the celluloid fallout of *Forbidden Planet*'s Robbie the Robot or *2001: A Space Odyssey*'s HAL. But fascination with the human machine extends back at least to 18th century tinkerers whose elaborate automata puzzle today's historians who wonder why so much time and detail would be devoted to these "toys" (Gideon, 1948). Strauss (1986) proposes an alternative interpretation that would place the fascination with early automata as precursors to both contemporary AI and cognitive science on the one hand, and to modern movies and stories on the other: the automata were a means of understanding by comparison what it means to be human at a time when divine designation seemed not quite enough.

The early artificers looked at their creations and asked, "What makes it humanlike?" Although it is natural to look to the object for an answer—as it there were some definitive attributes of "human" to which the essential features of the object need only be matched—it is fruitful to look to the perceiver. In the following paragraphs, I will consider some "perceiver-centered" areas in psychology to illustrate (a) how traditional topics may contribute to a better understanding of anthropomorphism, and (b) how anthropomorphism might be a vehicle for research. But anthropomorphism is not a property of perceivers, nor of objects: it lies in the *interaction* between the two.

Perceptions of Control: Machines. We might gain some insight into the motivational processes and consequences of anthropomorphism from studies concerned with perceptions of control when in fact there is no opportunity for control (e.g., Henslin, 1967; Langer, 1975; Langer and Roth, 1976; Wortman, 1976) and from studies where there is a failure to respond adaptively where control possibilities exist (e.g., Dweck, 1975; Wortman and Brehm, 1975; Wortman, 1976). The impetus to anthropomorphize may be related to the perception of one's inability to predict and control the environment. This perception is influenced by the availability of informal causal explanations of the observed behavior of the machine system. An individual's explanation may be veridical, or it may be inaccurate but still adaptive (i.e., within the actual capabilities and limitations of the machine).

When the individual cannot generate a workable explanation, then he or she is likely to generate an apparently nonadaptive social explanation (i.e., some personality characteristics or emotions) for observed machine behavior.

Naive users are more likely to be prone to anthropomorphism than sophisticated users because the former would be expected to have greater difficulty generating veridical and workable explanations for machine behavior. We would also expect different frequencies of oscillation between anthropomorphism and veridical attribution for users with different experience. In contrast to the apparently stable evaluation of one's control, the "machine-as-human" schema is likely to be fleeting, which suggests it could be a fruitful domain for investigating the dynamic processes of invoking and replacing active schemata, admixtures of how people work and how machines work.

Anthropomorphic explanations are maladaptive, at least for controlling the computer. They may, however, serve other functions, in which case the evaluation of the adaptiveness of anthropomorphism would be with respect to an explicit criteria. For example, the shift away from "real" control (i.e., from taking action that may reasonably produce desirable outcomes) might have two outcomes. One might be described as the illusion of control, once removed – a pretense in control. When all else fails, if one only says nice things about the machine to its face, it will continue to function smoothly. If it ceases to function smoothly, a combination of verbal abuse or exhortation will resolve the problem. Such an illusion or pretense may nevertheless help a person stay on task, or be an expression of being on task. The other outcome is learned helplessness, where one might abandon the task altogether on the grounds that "computers hate me," an anthropomorphic justification for avoiding computers. People may continue to use anthropomorphism when its "task instrumentality" is past if it still functions for "ego instrumentality" – that is, it makes the user feel better.

On the practical side, there are a variety of questions that link perceptions of the computer with performance of the user. One might ask whether (and what kind of) anthropomorphic attributions affect human-machine interaction? What characteristics of software will evoke attributions of human characteristics? Given a single program, under what circumstances do these and other attributions vary from moment to moment? Does the "warm feeling" Branscomb (1979) claimed for the user-friendly machine make people want to compute more, improve their performance or play at computing more? Does "the machine hates me" inspire us to conquer the truculent thing, or submit to it? Does "I must be terrible at this" (failure to anthropomorphize) promote giving up the task?

Perception of Control: Relationships. Whereas attributing (usually) negative characteristics to the machine seems to derive from its unpredictability, the unpredictability of human relationships seems to be associated with a positive anthropomorphic affirmation of the machine. The human machine fulfills a long standing fantasy of the robot or computer as companion. One of the best known is Asimov's creation of Robbie, a domestic robot, in 1940. In 1980, a computer hobbyist offered a program for a home computer equipped with various speech boards and proclaimed it a butler, babysitter, and companion (Hawkins, 1980). At the extreme, the companionable relationship may be sexual. Toffler (1970) attributed to a Cornell computer scientist the claim that man-machine sexual relationships were in the not-to-distant future. More frequently, the machine may be "someone" with whom to talk. Stockton (1980) describes a foreign scientist, exhausted after a long and trying trip, earnestly describing his feelings to a computer program (despite knowing better) in the host laboratory. Turkle (1984) describes members of the the MIT computer culture as "loving the machine for itself." Not quite a substitute for human relationships, the computer is "the human that never was"—challenging, compliant, all-forgiving, never rejecting, ultimately knowable and completely controllable, incapable of knowing or causing the emotional pain that makes human relationships risky (Turkle, 1984). But even before the era of "user-friendly" software, Jules Feiffer (1963) created a fantasy of the companionable computer. His hero, betrayed, rejected, ignored, and disappointed by other people, invented the "lonely machine."

The desirability of machines for companions probably tells us little about machines. Rather it appears to address, or obscure, deficits in interpersonal relationships (Turkle, 1980). Greist, Gustafson, Stauss, Rowse, Laughren, and Chiles (1973) reported that, compared with non-suicidal patients, twice as many suicidal patients preferred being interviewed by a computer rather than a psychiatrist. According to Taylor (1980), patients favor the use of computer interviews more than does the medical staff. Unfortunately, we get very little description of the extent to which people may engage in anthropomorphic behavior while they are interacting with various interview or counseling programs (e.g., Wagman, 1980; Wagman & Kerber, 1980), or the extent to which the program elicits anthropomorphic responding. It is not clear if the preference is associated with non-anthropomorphic considerations of time, efficiency, and privacy, or if people really do place the computer in the role of companion.

Social cognition. Although loss of control over the machine or relationships may be a motive for anthropomorphizing, it does not address the cognitive processes—mental structures and mechanisms—individuals might use to attribute human characteristics to machines. Anthropomorphized, computers and robots have a mixed status as social and nonsocial objects. The machine is animate, but not sentient. It may be attributed sentience, causal agency, and feelings, but never the capacity for smell or taste. It may arouse but never return affection. It reacts but never acts. It is like a "decomposed" person, its ambiguous personhood lying somewhere between the eye of the beholder and the attributes of the machine.

At least four considerations might help us better to locate anthropomorphism—constructs held by the person, individuals' linguistic usage, properties of the object, and the interaction of person and object. Anthropomorphism could be the result of the application of implicit personality theories (Hakel, 1969; Passini and Norman, 1966) to the limited and ambiguous information available from a machine. If so, a variety of machine "dispositions" would be similarly described, and differences between individuals would reflect differences in their implicit personality theories. Although different individuals have different theories, there is also evidence for a common set of underlying factors, probably derived from shared cultural experience, that influence the attribution of traits and dispositions. Because even very simple programs and robots appear to evoke an anthropomorphic reaction, they can open the door to a better formulation of just what cognitive structures and mechanisms a perceiver might bring into an interaction.

Unfortunately for our current purposes, implicit personality theories may not distinguish between personal constructs and linguistic usage. They may represent less a summary encoding of a person's organized, empirical information derived from past experience than they may be propositions about language—measures of the similarities of the meanings of words that function "for the sake of telling the world how it ought to behave" (Shweder & D'Andrade 1980, p. 53). From this perspective, anthropomorphism is not so much a psychological phenomenon as it is a linguistic device. Holland and Rohrman (1979) might dispute even this interpretation. Based on their studies of children, they argue that there is no such phenomenon as animistic thought in children; simply linguistic confusion about the use of the word "alive." Adults may be deemed unlikely to have this confusion (at least given the caveat, machines do not think), and we might propose that their anthropomorphic usage is a linguistic device for quickly communicating ideas about machines much as sociobiologists (e.g., Dawkins, 1976) anthropomorphize genes for efficient communication.

In contrast to emphasizing factors within the individual, researchers might investigate social inference processes that focus attention on how the perceiver represents properties of the object (Jones & McGillis, 1976; Reeder & Brewer, 1979). Behavior is presumed to be classified by an observer with respect to a limited number of observable attributes (e.g., "intelligence") and then a set of implicit rules or "schematic representations" are used to infer a trait or disposition. Variations in these rules would apply to different classes of attributes. Obviously, we would expect the domain of attributes applicable to machine intelligence to be more limited, and hopefully more tractable to description, than the domain of attributes applicable to human behavior. For empirical study, attributes can be programmed into a computer's operation; for example, first person reference to the activity of the machine. Comparisons between attributes and resulting social inferences about persons and about machines might be used to study how cognitive processes are brought to bear in making social inferences.

The humanized machine is directly relevant to debates on the distinction between social and non-social objects and on whether different cognitive processes are involved in responses to them. Traditional classification schemes (reviewed by Ostrum, 1984) focus on properties of the perceiver and properties of the object. Properties of the perceiver that have different implications for social and non-social objects include the recognition of similarity between the perceiver and the object (intersubjectivity); egocentric appraisal, which engages the self in interacting with objects; different possibilities exerted by social and nonsocial objects for control of the perceiver's outcomes; and the arousal of more sensory systems by social than nonsocial objects. Classification schemes for properties of objects generally focus on a static-dynamic dimension in basic physical appearance and composition, and the locus of causal agency (external or internal to the object). Both classification schemes view the perceiver and object in static isolation, as if some critical mass of features could determine the sociality of an entity. Neither approach comfortably accommodates the ambiguities of the human machine.

Ostrum's (1984) alternative is to claim that knowledge of objects is inextricably bound to knowledge of the self, and the links between the two are the interdependencies of perceiver and object through action. Anthropomorphism makes sense from this perspective because it seems to emerge primarily from action. Equally important, computers (and robots) can be used to study the varieties of action that contribute to the construal of an object as social. For example, it appears that some quality of "quasi-predictability" invokes a machine-as-human schema. Scheibe and Erwin (1979) found that among research participants playing a binary guessing game, intermediate difficulty produced the highest level of personification of the computer. It also appeared to engage the most self-involvement.

Different categories of action may have different impacts on the construction of objects as social. The children Turkle (1984) studied tended to count movement as a criterion for judgments of aliveness for traditional objects (e.g., clock, cloud, dog), but used psychological criteria (e.g., intention, emotion, and intellect) to make judgments of computers. But whereas the percentage of children over eight years old using movement and similar criteria for aliveness of traditional objects was lower than the percentage for the eight or under group, the percentage of older children using psychological criteria to judge aliveness of computers *was greater* relative to the younger group, partially as a consequence of the increasing psychological sophistication of older children. This is not to say that older children claimed the computer was alive, but rather different criteria of "aliveness" entered into their thinking about distinctions between computers and people. The most critical of these were distinctions between cognition and affect: "Computers are smart, but they don't have feelings."

But is it a Schema? The focus on attributes of machines and constructs held by a person places anthropomorphism at the heart of what Fiedler (1982) calls "the most interesting and the most challenging problem in cognitive psychology: the interaction between permanent memory structures and actual stimulus information" (p. 1001). The schema concept is a theoretical attempt to solve this problem, and if, as Fiedler claims, "Deviation from the stimulus is the evidence for a second source of information, the schema" (1982, p. 1012), then the tendency to attribute human characteristics to machines invites study by those interested in investigating schemata.

Despite using the term "schema" throughout this paper, I admit to doubts about its suitability as a description of, or even a shorthand for, what people do when they talk to machines or describe the activity of machines in human terms. Although schemata may be erroneously evoked and erroneous in content, certain kinds of errors are not part of the implicit notion of schema. Knowing it was time to feed my plants would never evoke a schema representing planting them in hamburger meat. In attributing human characteristics to machines, the gap between observable interpretation of the stimulus properties of the machine and the evocation of an anthropomorphic schema is so great as to verge on the irrational—more than a mere deviation from the stimulus information. Also, the schema concept implies some stability in cognitive structure and some stability in "actual stimulus information"; but anthropomorphism is frequently fleeting and dynamic beyond what may be implicit in the contextual setting for applying schemata to novel or incomplete information. Finally, the schema framework tends to view affect as post-cognitive, either based on the evaluation of the schema's component attributes or on category membership (Fiske, 1982). But the human machine is not sufficiently "well behaved" in its component attributes or category membership for the post-cognitive position.

By its very ambiguity, anthropomorphism suggests it is not driven by an analytical engine, but by an emotional engine. It might be most promising to consider the human machine (in its dual ambiguous sense) as exhibiting what Zajonc (1980b) tentatively labeled *preferenda*, the interaction of perhaps vague, gross or configural global features of a stimulus with changing internal states of the individual.

Zajonc was particularly concerned with demonstrating the primacy of evaluative affective judgments of the binary like-dislike sort, but the broader thrust of his argument is relevant to the challenge posed by anthropomorphism. According to Zajonc, the judgment of an object in relation to the state of the judge is an affective judgment. This is in contrast to a cognitive judgment, which would be based solely on qualities of the stimulus object (e.g., the cat is black). Self-other comparisons inevitably are affective judgments, but they can parade as if they were solely cognitive judgments. Affect strongly participates in information processing and implicates the self[2] in relation to the object.

The most global evaluation of an object is self vis-à-vis object—is it like me or not like me?—followed by action (and not necessarily self-aware thought) toward the object. In the case of computers or robots, we might expect that when the judgment is "not-like-me," affect would be low and action would deal with properties of the machine. But when the judgment is "like-me," affect would be high and action would implicate the self, resulting in anthropomorphic behavior. The search for the actual attributes of a machine is not generated until the respondent attempts to explain to others, or to him or herself, how the machine is the same or different from a human. But the question is one that automatically invites self-other comparison, thereby implicating the self.

It was here the early artificers of automata began. They wanted to compare how the machine is the same and different from human. They started by incorporating into their constructions first human form, then human motions. Their descendents in artificial intelligence, like Turkle's (1984) children, went beyond form and motion to thought. And like the early automata builders, they did so without realizing that their question automatically implicates the self.

The Scientific Community

Three analogies, loosely associated with changing historical emphases, influenced the use of the human-as-machine schema in the scientific community. Early cyberneticists asserted a *structural* analogy, comparing the components of a machine system and the neuronal circuitry of the central nervous system (e.g., McCulloch, 1949; von Neumann, 1958). The *cognitive function* analogy drew on a comparison between the human information processor and computer problem-solving (e.g. Minsky, 1968; Newell & Simon, 1972). Research problems were primarily restricted to mathematical or logical analysis. The *social cognitive function* analogy, involving pattern recognition and natural language processing, may be characterized by "story understanding" and "question answering" programs. The analogy is close to that of social interaction (e.g., Lehnert, 1978; Wilensky, 1983).

The human-as-machine schema has been enormously valuable for expanding our understanding of information processing, whether it is done by human or computer. Programs simulating human cognitive processes have provided a new class of theories. Research in related areas in artificial intelligence and cognitive psy-

[2] In its current incarnation, the self "may be viewed as a system of nodes in a memory network with associative links connecting various aspects or concepts comprising the self" (Linville, 1982, p. 93). This is a mechanomorphic description *par excellence*, but Linville's elaboration of what she means is consistent with ordinary intuitions about the term (e.g., the self includes past experience, affect, membership in different groups, physical traits, talent, behavior, etc., and it varies over time and circumstance). The ordinary intuitions about self are adequate to my purposes.

chology has greatly extended our understanding of problem solving, pattern recognition and concept formation (Simon, 1980). But were the analogy clearly and merely an analogy (cf. McCulloch, 1950), discourse on humans and machines would have the same rhetorical quality as does discourse in aerodynamics on the analogy between birds and airplanes. Few would claim, to paraphrase Newell and Simon's (1972) quote given earlier, that birds and airplanes are different species of a more abstract genus of flying systems and that the concepts describing one perforce describe the other. The *generality* of the analogy is unjustified. So what underlies the assumption of generality for the human machine analogy?

There is a crucial difference between flying systems and information processing systems. For the former, the assessment of any analogy between birds and airplanes relies on comparing observable properties associated with the dimension of interest (flying) of two "species" both of which are external to the observer. The comparison between "men and computers" is vastly more complicated because it is not clear what observable properties are associated with the dimension of interest—thought. The observable properties of an external system are compared to unobservable properties of an internal system. One solution I will discuss in greater detail later was originally proposed by Turing (1950/1964). It argues that the only basis we have for inferring that others have thought, consciousness, minds, or feelings is by comparing their behavior with what we expect or know to be our own in similar circumstances. The comparison unavoidably implicates the self.

From the first, that question, "Can machines think?" (Turing, 1950/1964) pointed to a particular depiction of the self as the crucial grounds for comparison. Thinking—good rational thought devoid of feeling, the kind of which Bacon would have approved and that Norman (1980) describes as the Pure Cognitive System—was the important dimension. Similarly, the attribution of thinking was to be a cool, rational assessment of the observable "facts." By fiat, the self would not be implicated in thought. It was a conception of human uniqueness that can be traced to two related illusions born in the zeitgeist of the Enlightenment. One was the illusion of unbounded human rationality. The other was the illusion of detachment, that is, that thought could be separated from feeling—I think therefore, I am; not, I feel (and think), therefore, I am. These illusions have been highly useful in scientific practice. But evidence that humans are not so rational (Kahneman, Slovic & Tversky, 1982) nor so detached (Zajonc, 1980b) is still too new, too tentative to have eliminated the tinges of Enlightenment hubris.

So what are we to make of PARRY, ELIZA, SAM, PAM, PHRAN, and PHRED ("PHRED is PHRAN's sibling program," says Wilensky, 1983), the software characters that populate scientific laboratories? Researchers have made fantastic claims for them and assorted relatives that have tantalized both popular and scientific imaginations, the most incredible claims originating in the era of punch cards and batch processing (Caporael, 1984). The ambiguous human machine illustrates the action of *back metaphor*, a concept originally proposed by Burian (1978) to describe the consequences of anthropomorphism in biological thinking. In "information processing thinking," the machine is described in terms of a human metaphor constrained to the possible behavior of the machine, and then the constrained metaphor refers back, frequently as an explanation, to human behavior. The "thought" of the machine-as-human, a computer, has no developmental history, no multiplicity of motives, no bodily extension in space (Neisser, 1963). Nevertheless, for the human-as-machine, motivation and emotion has been described in terms of two mechanisms, goal-terminating subroutines and interrupt mechanisms that allow the serial information processor to respond in real time to urgent needs without allowing any one subgoal to dominate the processor (Simon, 1967). Mechanism, Turkle (1984) observes, replaces meaning. Under the shroud of detachment, affect is discarded and anthropomorphism is transformed in the scientific community into mechanomorphism.

Two objections can be raised to the scientific transformation of anthropomorphism to mechanomorphism. First, haphazard as it may be, people do look to science to tell them about human nature. Distinctions scientists make among scientific facts, simplifications, metaphors, and speculations frequently disappear by the time they are interpreted for the conventional community. Mechanomorphic claims about the nature of mind and self will be erroneous and limiting because such claims are back metaphors. Turkle's (1984) book is rife with descriptions of people who think of themselves as "collections of programs," or, unable to abandon the vividness of their emotional life, propose their minds are rational programs and their feelings are murky. The concept of people as machines produces its own back metaphor in the conventional community. Because humans can do so much and they are collections of programs, machines can be programmed to do anything humans do. (Last month one of my students confidently announced and fiercely argued that he had no doubt that in the next 10 years there would be domestic robots that could pull clothes out of the dryer and fold them. Only "sophisticated programming" would be necessary to distinguish the children's underwear.) Mine is no Luddite objection to the further displacement, this time by machines, of the human spirit from the center of the universe. To the contrary, the high value our culture places on reason, logic, efficiency and productivity suggests the caricature of the merged human machine at the center of the universe. A more brutal displacement is likely to be based on showing that adults in industrial cultures have more in common with children and nonindustrialized peoples than they do with machines (cf. Norman, 1980).

The second objection to mechanomorphism is how all the myriad implications of human-as-machine qualify the "search space" of human cognition. The objection is more than a concern over neglect of issues such as affect and the significance of the pervasive social context of human mental activity. Rather, the identity between the information structure and processes describing humans and machines prevents serious consideration of alternatives to mechanomorphism. Wimsatt (1984) describes two sources of the identity problem. First, even though there may be a variety of models describing "information processing," the commonality of generic approaches results in models that may produce "pseudo-robustness" by failing to include or consider environmental variables of the reductionistic system being modeled. Second, when a similarity (e.g., in processes or in outputs) is used to transform an old hard problem (how do people think?) into a new problem (how do information processors work?) and the redefined problem is more analytically tractable, there is a tendency to claim the redefinition encapsulates the core features of the old problem and thus solves it. The temptation is very strong to claim the new formulation was really the old problem all along—the latter has merely been "clarified."

Our models of human behavior are in a tenuous position if they are based on the current technology and only advance under the auspices of a new invention. Our psychological theories must then lag behind our technology; historically they have, be it the technology of the hydraulic engine, telephone switchboard, computer, or holographic memory. But limitations in theories or choices of empirical directions can be overcome by discussion and by successful exploitation of alternatives. They are corrected with less difficulty (which is not to say easily corrected) than mechanomorphic claims let loose in the conventional community. There, they can acquire a life of their own impervious to corrective efforts.

Taking a very broad view of the cognitive revolution initiated with intelligent machines, we can discern some shifts in the direction of more viable conceptions of both humans and machines. The early vast claims exposed the vastness of simple problems. "Decision tools," expert systems that might represent hundreds of rules for making a narrowly defined decision (Duda & Shortliffe, 1983), are replacing anthropomorphized machines that could "think, learn, and create" (Simon & Newell, 1958). Understanding a simple story about love and marriage (Wilensky, 1983) is replacing chess playing and mathematical problem solving as the epitome of human intellectual activity (James & Partridge, 1972). Language and development, emotion and interaction, skill and performance, are rounding out the agendas of cognitive scientists (Norman, 1980). The human machine is a character that can enter on center stage in the next act.

THE ANTHROPOMORPHIC BIAS

Identifying the Beast

A reader of an earlier version of this paper complained that anthropomorphism did not seem to be a "fundamental tendency": rather, it is but one example of the tendency toward metaphoric understanding of the uncomprehended. I am inclined to believe this view is common. I also believe that it grows out of the description of sociality as secondary to cognition, of feeling as secondary to thought. That anthropomorphism is a "fundamental tendency" may not be unequivocably demonstrable, but epistemological and ontogenetic arguments might be mustered suggesting it to be a "default schema" applied to nonsocial objects, one that is abandoned or modified in the face of contradictory information.

The "realist" view of cognitive processes defended by Ostrum (1984) "derives from pragmatic analysis of the response demands placed on people by their social and nonsocial environment. Realists argue that the foundation of all cognitive processes derives from episodes involving social objects (as opposed to nonsocial objects)" (p. 25). Zajonc (1980a) proposes social cognition is the "general case" of human cognitive activity. Three crucial parameters are involved in cognition: (a) the observer-object interaction, (b) affect, and (c) involvement of self. Nonsocial perception is the special case where these parameters are set to zero. Cognitive processes, from birth onward, are initially developed to handle the complexities of social information, and are modified through interaction with the environment for application to nonsocial objects.

I have argued elsewhere (Caporael, in press) that there are evolutionary grounds for a realist view of cognitive processes. Humans were under strong selection pressures for evolving affective and cognitive mechanisms that would support the development and maintenance of membership in the band-sized group, which was characteristic in the evolutionary history of the species. Such mechanisms would be available for "reweaving" groups into societies on a larger scale, but would also introduce what might now be identified as cognitive biases or deviations from rational behavior. Anthropomorphism might be one such bias (or remnant of one) insofar as in our evolutionary history, the most important objects about which to make predictions would be other group members. Attributing human characteristics may be part of a psychological *Bauplan* (i.e., a default schema in the strongest sense, cf. Gould & Lewontin, 1978/1984) originating in human evolutionary history with its application constrained and directed by individual experience and cultural transmission of beliefs and attitudes (cf., Boyd & Richerson, 1985).

Anthropomorphism as a default schema is neither novel nor recent. From his work on animistic thought in childhood, Piaget (1929/1967) concluded that universal life is the primary assumption of child thought. The attributes of inert matter are gradually detached by thought and experience from the primitive continuum along which all things are regarded as living. Among anthropologists, animism is a hallmark of "primitive thought." Frazer (1922/1963) interpreted magic, with its anthropomorphic ascriptions to nature, as "pseudo-science"—attempts to control the course of nature for the practical ends of humans. For Malinowski (1925/1954), magic filled the gap where knowledge, technical skill and past experience fell short.

The apparent correspondence between ontogeny and history, a sort of cultural recapitulation notion, raises several interesting questions. Is there really such a correspondence? If so, what are the minimal requirements on the perceiver-entity interaction (a preferenda, perhaps?) that might induce the attribution of human characteristics? What kind of characteristics are attributed? Should there be special categories distinguishing the "as if" anthropomorphism of computer users from the "it is" anthropomorphism of young children and non-industrialized peoples? Initial directions lie in a large, confused and confusing literature extending back to the 19th century. Answers lie in experimental research manipulating anthropomorphic attributions.

The ontogeny/history correspondence for anthropomorphism might also be exploited as an epistemological methodology. Piaget was once singular in his use of this methodology. Nisbett, Krantz, Jepson, and Kunda (1983) have applied this approach to the analysis of the growth of statistical heuristics in everyday induction and demonstrated the significance of the *conjunction* of individual experience with randomizing devices and cultural beliefs, developed in the 17th century, of a mechanistically determined world. Wiser and Carey (1983), examining the evolution of thermal theories, observe that the misconceptions of novice science students are usefully conceived as conceptual changes that correspond to theory changes in the history of science.

Anthropomorphic attributions by modern adults tend to go unnoticed or to occasion brief amusement at best. Traditionally, anthropomorphic thinking has been ascribed to children and to "the lower races" (Tyler, 1871/1953) in contrast to the ascription of reason to adults and civilized societies. Ageism and racism are not barriers to contemporary research, but they did put the whole issue to bed a long time ago. Consequently anthropomorphism is not available, in the cognitive heuristics research sense of availability (Kahneman, Slovic & Tversky, 1982), for constructing alternative explanations of behavior. Wiser and Carey (1983) fail to see the anthropomorphic attribution in their example of Aristotle's physics where bodies fall because they seek a specific natural resting place. His physics is simply "foreign to the modern mind" (p. 296).

Heider and Simmel's 1944 study illustrates how nonavailability can influence interpretations and directions of research. They self-consciously resort to anthropomorphism to describe their stimulus, but barely attribute any significance to it in the interpretation of their study. Subjects in their study watched a two and a half minute film clip of a large triangle, a small triangle and a circle moving in the vicinity of a rectangle with a flap that opens and closes. The geometric forms move through the flap, which opens and closes in synchrony with the forms' movements so that the forms at various points are inside or outside the boundaries of the rectangle. Unless you are familiar with the study, this tells you literally nothing. So I will do what Heider and Simmel (1944) somewhat apologetically did: the little triangle and the big triangle were chasing each other and went into the house, the big triangle beat up the little triangle who nevertheless persevered and moved off with the little circle. The big triangle was trapped in the house. The consistency with which subjects give some paraphrase of this description when instructed just to "describe what you see" is astonishing. Only one subject described the film in terms of the movement of geometric shapes (and still slipped at the end, referring to a triangle as "he"), two subjects described the stimuli in terms of animals and the rest of the 34 subjects independently described the activity in terms of social events with high agreement on the gender and personality traits of the geometric forms.

Despite the wide citations of this study, the anthropomorphization of these nonsocial stimuli was not itself noticed nor declared a phenomenon meriting further investigation. Researchers used the triangle film and its variants as just another stimulus to study pre-information (Shor, 1957), or selective perception (Massad, Hubbard & Newtson, 1979), or multiple sufficient causes, augmentation and discounting (Kassin & Lowe, 1979). The research assumed either that the cognitive processes for dealing with nonsocial events are the simpler, conceptually fundamental processes that are the building blocks for processing social events or that there are no differences between social and nonsocial stimuli. Yet Heider and Simmel's (1944) original study suggests it may be easier to apply the principles of social knowledge to nonsocial objects than the converse.

Setting the Traps

Throughout this paper, I have taken the dogmatic position that machines do not think in order to clarify the relationship between person and machine. It *is not* anthropomorphic to attribute human characteristics to humans. It *is* anthropomorphic to attribute human characteristics to machines (or rivers or rocks) that do not have human characteristics. But it *is not* anthropomorphic to attribute human characteristics to machines that indeed do have the characteristics attributed to them. By temporarily closing that last door, it becomes possible to pose the question, "what makes people think that machines think?" without having to define thinking first and decide whether or not machines do it. At the same time, by hypothesizing anthropomorphism to be a default schema, I have painted myself (and readers who have concurred with me so far) into a solipsistic corner. If anthropomorphism is an inherent bias, how could we *know* if machines think? How could we distinguish its thought from our perhaps unwarranted ascriptions of thought?

We could not use the "imitation game." It is based on the argument that the only way we know other people are thinking or have consciousness is by the process of comparing them to ourselves, and there is no reason to treat machines any differently. If we cannot distinguish the written responses of a machine from those of a human, and we are willing to grant that one of these two entities thinks, than we should be willing to grant that the other one does, too (Turing, 1950/1964;

much of the ensuing discussion is based on Hodges' (1983, pp. 415–426) critical analysis of the Turing paper). But if there is an inherent anthropomorphic bias, the imitation principle clearly incorporates a confirmatory bias, and is therefore unreliable. Turing exacerbated the problem by arguing that if the machine could be attributed with thought, by extension it could be attributed with feelings. The extension to feelings was not made on the grounds of any particular argument. Rather it was a corollary of the self-other comparison process Turing recommended.

In a previous section, I wrote that Turing's question, "Can machines think?" focused on a particular depiction of the self embedded in Western culture. Despite the significance of the self-other comparison in determining if machines can think, Turing wrote as if only attributes of the machine were relevant to the defense of intelligent computers. That is, he wanted the computer both as a social object, implicating the self through comparison with the self, and as a nonsocial object, independent of the self. Hodges' (1983) biography of Alan Turing makes it possible to illustrate how the self-comparison implicates the self beyond just the cultural depiction. Were this merely a biographical exercise I would leave it to Hodges. But the Turing test, as the imitation game is also known, has become something of a standard for comparing humans and machines. As such, it requires further discussion.

Turing's model of thought, or intelligence, was rooted in playing chess and solving mathematical puzzles and problems. These topics were chosen in part because they involved "no contact with the outside world"; they were amenable to specification in the discrete state language of the computer. Turing believed it would be but a matter of time before "real" thought would be similarly specifiable in terms of rule-governed discrete states. Thus, Turing implied there was to be a model of intelligence, the essential feature of which could be mapped onto the computer, *independently* of the self in the self-other comparison by which we are to decide if machines think. At least superficially, Turing was content with a functionalist theory of mind. But he did not leave a list of the essential features. Rather he wrote a list of responses to objections that machines could think. Although the responses were construed in terms of the computer he was confident would be built in the future, they were also a personal statement of what Turing thought important to the characterization of thought. (This interpretation explains the puzzling response to "objections from evidence for extra-sensory perception." Turing was impressed by J.B. Rhine's claims on mental telepathy.)

Hodges (1983), cognizant of the intertwining of Turing's life and work, seems to accept that there can be independent criteria to judge thought, but objects to Turing's characterization on the grounds that it did not acknowledge the significance of human interaction and experience to thought or knowledge (the realist social cognition objection). Yet Turing could not escape that significance, even if he failed to acknowledge it. Another reason for choosing chess and problem solving as the basis for a model of mind is that these mattered to Turing. They had been of interest since childhood and the focus of his intellectual life. They were thus the relevant dimensions for the self-other comparison for determining if machines could think. "The discrete state machine, communicating by teleprinter alone, was like an ideal for his own life, in which he would be left alone in a room of his own, to deal with the outside world solely by rational argument" (Hodges, 1983, p. 425). Thus the "Turing test" is just that: a test of Alan Turing. It provides not a "fair, demanding but possible, and crisply objective" test (Hofstadter

& Dennet, 1981, p. 93), nor a rigorous set of test criteria, but rather an intensely personal process born of bias, desire, and past experience.

Nevertheless, chess playing and problem solving could have been operational criteria for answering the question, "do machines think?" We might say, "Machines think when they *are solving* mathematical or chess problems." The advantage of this functionalist theory is that if we can agree on the operationalization, we can gather around the computer and see if it is solving problems. There might still be the temptation, if the computer fails to solve the problem, to say the machine had the intention or goal to solve the problem. After all, thinking humans fail to solve problems on occasion. The bare operational definition, however, seems to distort what we mean by thought. More importantly, *any* operational definition for thought ultimately depends on a comparison with the self and is thus exposed to the anthropomorphic bias.

If the imitation game, a list of attributes, or an operational definition of thought cannot determine whether machines think, there are at least three remaining alternatives. One is dogmatically to declare that only a human brain (or mind) is capable of thought – a machine is a machine so it cannot think and that's that (Searle, 1980). Apparently this solution is sufficiently disagreeable to have sustained over 30 years of debate. Another alternative is to say that if we think it thinks, it thinks. If there is an anthropomorphic beast, we will just walk into its jaws (Dennett, 1981). But ultimately that tells us less about machines than what we think of ourselves. A third is to allow there to be an anthropomorphic beast and set traps for it.

Even if an anthopomorphic bias is inherent, we can still rationally decide when it might be useful to anthropomorphize and when it is not. Such decisions are made with people. In the military, for example, it is often convenient to act as if people do not think, but rather that their behavior is controlled by a set of instructions, much like a computer's. ("It's not for me to reason why, it's just for me to do or die.") Whether machines think is thus changed from an ontological question to a pragmatic question.

To address the pragmatic question, we would want to know the circumstances under which the propensity to anthropomorphize is strong and to have a relatively fixed point at which we decide it is counterproductive. That point is where we want to set traps. A trap is a strategy to subvert a propensity in the circumstances where the propensity is undesirable. Putting a scale in front of the refrigerator is a minor trap for a dieter. Some traps might be major. Consider the bias toward ingroup favoritism – the attribution that one's own group is more loyal, trustworthy and friendly than is the outgroup. Suppose that on the basis of research (e.g., Brewer, 1979, 1981; Brewer & Campbell, 1976; Campbell, 1982; Janis & Mann, 1977) it was determined that decision making groups operating in secrecy were especially prone to ingroup favoritism, and that eliminating secrecy extended or eliminated definitions of the group boundary (circumstances of the propensity). Further suppose that ingroup bias, productive in forming a sense of community, was counterproductive when the bias converged on national boundaries and the decisions involved control of nuclear weapons (counterproductive point). We might decide to subvert the bias by eliminating secrecy for military decision making (the trap).

The circumstances for anthropomorphizing and the point at which it is counterproductive (or enhances production) can be established by research. For example, it should be a relatively direct empirical issue to determine the circumstances under which novice programmers anthropomorphize the computer, and whether the propensity is associated with more time spent using a computer, or with abandoning the whole task altogether. A minor trap for anthropomorphism might be to eliminate the use of "I" and the user's name from the computer's responses (or to add them if it was desirable to enhance enthropomorphic tendencies). In artificial intelligence and cognitive science research, the circumstances are more subtle and complicated. Sometimes it may be useful to be anthropomorphic even if we are quite confident a system cannot think. Heider and Simmel (1944) anthropomorphized geometric forms for communicative economy. At other times, the point where we decide anthropomorphism is counterproductive may be the point where it is "correct" to attribute human characteristics to the machine because it does, in fact, think.

Trickier traps would be more desirable where the goal is to duplicate or simulate human cognition, albeit the best we might be able to do are nonobvious forms of the imitation game. For example, many claims about machines thinking, understanding, or creating have been predicated on the assumption that its inferences or problem solution would be correct. But people show systematic deficits, that is, deviations from what would be expected on the basis of normative rationality or statistics (Kahneman, Slovic & Tversky, 1982). Presumably, human deficits are a consequence of mental "heuristics" that usually do produce appropriate inferences—instances of being right for the wrong reasons (Nisbett & Ross, 1980). One trap, which we might call the "heuristic challenge" for the sake of convenience, would consist of the requirement that an inferencing program would show the same patterns and kinds of *errors* that humans show using the *same* computer algorithms or heuristic that usually produces acceptably valid inferences. That is, could the set of instructions that generates the conclusion that there are more Mexican restaurants in San Diego than in Minneapolis also generate the conclusion that bright, outspoken Linda who majored in philosophy and was concerned about issues of discrimination and social injustice is more likely to be a bank teller and a feminist than just a bank teller (the conjunction fallacy based on representativeness, Tversky & Kahneman, 1983)?

The heuristic challenge could stand accused of involving a mechanomorphic bias, and it would be guilty to the degree that it extends interpretations of cognitive processes by drawing on the computer metaphor (as Turing would extend thought to feeling by drawing on the human metaphor). As it is, the contrasts and comparisons drawn in the mental heuristics literature are to logical, economic and statistical systems, which may be described by computer programs but are not rooted in that domain. There is a sense in which mental heuristics make especially pretty traps because, as a research area, they are at the frontiers of behavioral science (Simon, 1980). But at least some of the phenomena to which they refer were described by 19th century ethnographers (e.g., the putative diagnosis or cure for a disease would be something related to the symptom by resemblance). Thus, this use of mental heuristics should be considered "nonobvious" rather than "new."

CLOSING THOUGHTS

In summary, I have suggested that anthropomorphism merits attention as a psychological phenomena in its own right, especially as a "default schema" under conditions of quasi-predictability; that it can be used as a vehicle for basic and applied cognitive research; and that it is a foil against which to pose theoretical and metatheoretical questions, such as characterizations of objects in relation to observers and the conceptual independence of properties of objects that intersect in some fashion with properties of humans, even in scientific analyses. But we should also look at anthropomorphism "in reverse." People do not automatically attribute human characteristics to humans. What properties do certain people—stroke victims, for example, or our enemies—lack that we fail to attribute to them (at least some) human characteristics? We can undertake research on the attribution of human characteristics while remaining agnostics on the question of whether machines think, just as we can remain agnostics on whether people think, or the cat wants us to open the door so he or she can go out. In the end, the question of what makes people think machines can think is a question about what makes us think other people can think.

REFERENCES

Boden, M. (1981). *Artificial intelligence and natural man.* New York: Basic Books.

Boyd, R., & Richerson, P.J. (1985). *Culture and the evolutionary process.* Chicago: University of Chicago Press.

Branscomb, L.M. (1979). *The human side of the computer.* Paper presented at the Symposium on Computer, Man and Society, Haifa, Israel.

Brewer, M.B. (1979). Ingroup bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, **86**; 307–324.

Brewer, M.B. (1981). Ethnocentrism and its role in interpersonal trust. In M.B. Brewer & B.E. Collins (Eds.), *Scientific inquiry and the social sciences* (pp. 345–360). San Francisco: Jossey-Bass.

Brewer, M.B., & Campbell, D.T. (1976). *Ethnocentrism and intergroup attitudes: East African evidence.* New York: Halsted Press.

Burian, R.M. (1978). A methodological critique of sociobiology. In A.L. Caplan (Ed.), *The sociobiology debate* (pp. 376–395). New York: Harper & Row.

Campbell, D.T. (1982). Legal and primary-group social controls. *Journal of Social and Biological Structures*, **5**, 431–438.

Caporael, L.R. (1984). Computers, prophecy and experience: A historical perspective. *Journal of Social Issues*, **40**, 15–49.

Caporael, L.R. (in press). *Homo sapiens, Homo faber, Homo socians:* Technology and the social animal. In W. Callebaut and R. Pinxten (Eds.), *Evolutionary epistemology: A multiparadigm program.* Dordrecht: Reidel.

Caporael, L.R., & Thorngate, W. (1984). Introduction: Toward the social psychology of computing. *Journal of Social Issues*, **40**, 1–13.

Dawkins, R.A. (1976). *The selfish gene.* New York: Oxford University Press.

Dennett, D.C. (1981). True believers: The intentional strategy and why it works. In A.F. Heath (Ed.), *Scientific explanation* (pp. 53–75). New York: Oxford University Press.

Duda, R.O., & Shortliffe, E.H. (1983). Expert systems research. *Science*, **220**, 261–268.

Dweck, C.S. (1975). The role of expectations and attributions in the alleviation of learned helplessness. *Journal of Personality and Social Psychology*, **31**, 674–685.

Feiffer, J. (1963). The lonely machine. *Feiffer's Album.* New York: Random House.

Fiedler, K. (1982). Causal schemata: Review and criticism of research on a popular construct. *Journal of Personality and Social Psychology*, **42**, 1001–1013.

Fiske, S.T. (1982). Schema-triggered affect: Applications to social perception. In M.S. Clark & S.T. Fiske (Eds.), *Affect and cognition* (pp. 55–78). Hillsdale, NJ: Erlbaum.

Frazer, J.G. (1963). *The golden bough.* (Abridged edition). New York: Macmillan. (Original work published 1922).

Giedeon, S. (1948). *Mechanization takes command.* New York: Norton.

Gould, S.J., & Lewontin, R.C. (1984). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. In E. Sober (Ed.), *Conceptual issues in evolutionary biol-*

ogy (pp. 252–270). Cambridge, MA: MIT Press. (Reprinted from *Proceedings of the Royal Society of London*, 1978, **205**, 581–598).
Greist, J.H., Gustafson, D.H., Stauss, F.F., Rowse, G.L., Laughren, T.P., & Chiles, J.A. (1973). A computer interview for suicide-risk prediction. *American Journal of Psychiatry*, **130**, 1327–1332.
Hakel, M.D. (1969). Significance of implicit personality theories for personality research and theories. *Proceedings of the 77th Annual Convention of the American Psychological Association*, pp. 403–404.
Hawkins, B. (1980, January). Breslin—the home computer that runs my house. *Popular Science*, pp. 82–84, 154.
Heider, F. (1958). *The psychology of interpersonal relations.* New York: Wiley.
Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, **57**, 243–259.
Henslin, J.M. (1967). Craps and magic. *American Journal of Sociology*, **73**, 316–330.
Hodges, A. (1983). *Alan Turing: The enigma.* New York: Simon and Schuster.
Hofstadter, D.R., & Dennett, D.C. (1981). *The mind's I.* New York: Basic.
Holland, V., & Rohrman, N. (1979). Distribution of the feature [+ animate] in the lexicon of the child. *Journal of Psycholinguistic Research*, **8**, 367–378.
James, E.B. & Partridge, D. (1972). Machine intelligence: The best of both worlds? *International Journal of Man-Machine Studies*, **4**, 23–31.
Janis, I.L., & Mann, L. (1977). *Decision making.* New York: Free Press.
Jones, E.E., & McGillis, D. (1976). Correspondent inference and the attribution cube: A comparative reappraisal. In J. Harvey et al. (Eds.) *New directions in attribution research* (Vol. 1, pp. 389–420). Hillsdale, NJ: Erlbaum.
Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.
Kassin, S.M., & Lowe, C.A. (1979). On the development of the augmentation principle: A perceptual approach. *Child Development*, **50**, 728–734.
Langer, E.J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, **32**, 311–328.
Langer, E.J., & Roth, J. (1976). Heads I win, tails it's change: The illusion of control as a function of the sequence of outcomes in a purely chance task. *Journal of Personality and Social Psychology*, **33**, 951–955.
Lehnert, W.G. (1978). *The process of question answering.* Hillsdale, NJ: Erlbaum.
Lewis, A.O., Jr., (Ed.). (1963). *Of men and Machines.* New York: E.P. Dutton.
Linville, P.S. (1982). Affective consequences of complexity regarding the self and others. In M.S. Clark & S.T. Fiske (Eds.), *Affect and cognition* (pp. 79–109). Hillsdale, NJ: Erlbaum.
Malinowski, B. (1954). *Magic, science and religion.* New York: Doubleday. (Original work published 1925).
Massad, C.M., Hubbard, M., & Newtson, D. (1979). Selective perception of events, *Journal of Experimental Social Psychology*, **15**, 513–532.
McCulloch, W.S. (1949). The brain as a computing machine. *Electrical Engineering*, **68**, 492–497.
McCulloch, W.S. (1950). Machines that think and want. *Comparative Psychology Monographs*, No. 20, 39–50.
Minsky, M. (1967). Why programming is a good medium for expressing poorly understood and sloppily formulated ideas. In M. Krampen and P. Seeitz (Eds.), *Design and planning 2: Computers in design and communication* (pp. 117–127). New York: Hasting House.
Minsky, H. (Ed.). (1968). *Semantic information processing.* Cambridge, MA: MIT Press.
Neisser, U. (1963). The imitation of man by machine. *Science*, **139**, 193–197.
Newell, A. & Simon, H.A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.
Nicholson, T. (1979, April 23). Blue-collar robots. *Newsweek*, pp. 80–81.
Nisbett, R.E., Krantz, D.H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, **4**, 339–363.
Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings in human judgment.* Englewood Cliffs, NJ: Prentice-Hall.
Norman, D.A. (1980). Twelve issues for cognitive science. *Cognitive Science*, **4**, 1–32.
Ostrum, T.M. (1984). The sovereignty of social cognition. In R.S. Wyer, Jr. & T.K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 1–38). Hillsdale, NJ: Erlbaum.
Passini, F.T., & Norman, W.T. (1966). A universal conception of personality structure? *Journal of Personality and Social Psychology*, **4**, 44–49.
Piaget, J. (1967). *The child's conception of the world.* (J. Tomlinson & A. Tomlinson, Trans.). Totowa, NJ: Littlefield, Adams. (Original work published 1929).

Reeder, G.D., & Brewer, M.B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, **86**, 61–79.

Reichardt, J. (1978). *Robots: Facts, fiction and prediction.* London: Thames and Hudson.

Schefter, J. (June, 1980). New workers on the assembly line: Robots that think. *Popular Science*, pp. 46–57.

Scheibe, K.E., & Erwin, M. (1979). The computer as alter. *Journal of Social Psychology*, **108**, 103–109.

Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, **3**, 417–457.

Shor, R. (1957). Effect of preinformation upon human characteristics attributed to animated geometric figures. *Journal of Abnormal and Social Psychology*, **54**, 124–126.

Shweder, R.A. (1977). Likeness and likelihood in everyday thought: Magical thinking and everyday judgments about personality. *Current Anthropology*, **18**, 637–648.

Shweder, R.A. & D'Andrade, R.G. (1980). The systematic distortion hypothesis. In Shweder, R.A. (Ed.), *Fallible judgment in behavioral research.* San Francisco: Jossey-Bass.

Simon, H.A. (1967). Motivational and emotional controls of cognition. *Psychological Review*, **74**, 29–39.

Simon, H.A. (1980). The behavioral and social sciences. *Science*, **209**, 72–78.

Simon, H.A., & Newell, A. (1958). Heuristic problem solving: The next advance in operations research. *Operations Research*, **6**, 449–450.

Stevens, A.L., & Gentner, D. (1983). Introduction. In D. Gentner & A.L. Stevens (Eds.), *Mental models* (pp. 1–6). Hillsdale, NJ: Erlbaum.

Stockton, W. (1980, December 14). Creating computers that think. *The New York Times Magazine*, pp. 48–54, 60, 62, 64, 68, 70.

Strauss, L. (1986). *Automata: The history of a mechanical metaphor.* Unpublished manuscript, Department of Science and Technology Studies, Rensselaer Polytechnic Institute, Troy, NY.

Taylor, T.R. (1980). The role of computer systems in medical decision making. In H.T. Smith & T.G.G. Green (Eds.), *Human interaction with computers* (pp. 231–266). New York: Academic Press.

Toffler, A. (1970). *Future shock.* New York: Random House.

Turing, A.M. (1964). Computing machinery and intelligence. In A.R. Anderson (Ed.), *Minds and machines* (pp. 4–30). Englewood Cliffs, NJ: Prentice-Hall. (Reprinted from *Mind*, 1950, **59**).

Turkle, S. (1980, January/February). The computer as Rorschach. *Society*, pp. 15–24.

Turkle, S. (1984). *The second self: Computers and the human spirit.* New York: Simon and Schuster.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, **90**, 293–315.

Tyler, E.B. (1953). Animism. In M. Mead & N. Calas (Eds.), *Primitive heritage* (pp. 46–52). New York: Random House. (Original work published 1871).

von Neuman, J. (1958). *The computer and the brain.* New Haven: Yale University Press.

Warwick, P.S. (1980). *The cybernetic imagination in science fiction.* Cambridge, MA: MIT Press.

Wagman, M. (1980). PLATO DCS: An interactive computer system for personal counseling. *Journal of Counseling Psychology*, **27**, 16–30.

Wagman, M., & Kerber, K.W. (1980). PLATO DCS, An interactive computer system for personal counseling: Further development and evaluation. *Journal of Counseling Psychology*, **27**, 31–39.

Weizenbaum, J. (1976). *Computer power and human reason.* San Francisco: W.H. Freeman.

Wilensky, R. (1983). Story grammars versus story points. *Behavioral and Brain Sciences*, **6**, 579–623.

Wimsatt, W.C. (1984). *Heuristics and the study of man.* Unpublished manuscript, Department of Philosophy, University of Chicago.

Wiser, M., & Carey, S. (1983). When heat and temperature were one. In D. Gentner & A.L. Stevens (Eds.), *Mental models* (pp. 267–297). Hillsdale, NJ: Erlbaum.

Wortman, C.S. (1976). Causal attributions and personal control. In J.H. Harvey et al. (Eds.), *New directions in attribution research* (Vol. 1, pp. 23–52). Hillsdale, NJ: Erlbaum.

Wortman, C.B., & Brehm, J.W. (1975). Responses to uncontrollable outcomes: An integration of reactance theory and the learned helplessness model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8, pp. 278–336). New York: Academic Press.

Zajonc, R.B. (1980a). Cognition and social cognition: A historical perspective. In L. Festinger (Ed.), *Retrospections on social psychology* (pp. 180–204). New York: Oxford University Press.

Zajonc, R.B. (1980b). Feeling and thinking: Preferences need no inferences. *American Psychologist*, **35**, 151–175.