The Journal of Mind and Behavior
Spring 1987, Volume 8, Number 2
Pages 261-268
ISSN 0271-0137

Can We Construct Kantian Mental Machines?

Colin Martindale

University of Maine

The authors of this series of articles argue that cognitive psychology is not really a new paradigm but merely a disguised version of behaviorism. In large part, I agree with them on this point. A good recent example of the interchangeability of cognitive and behavioristic explanations is Rachlin, Logue, Gibbon, and Frankel's (1986) treatment of choice behavior. One, incidently, comes away from their article with the impression that the cognitive "disguise" really sets back rather than furthers our understanding. The cognitive interpretation is easier to understand on a subjective level, but it is less precise and does not account for as much as the behavioristic explanation. An intuitive sense of understanding is substituted for a more rigorous and scientific understanding. However, cognitive psychology is not simply behaviorism with mind stuck in between stimulus and response. Unfortunately, most cognitive psychologists forget to include in their theories a few "details"—e.g., motivation, incentive—that the behaviorists had covertly inserted between stimulus and response. It could just as well be argued that at least some types of cognitive psychology are partial reinstatements of the Wundtian or structuralist paradigm (Blumenthal, 1975). In this view, cognitive psychology has recovered much of the historical subject matter of psychology. Unfortunately, again, most cognitive psychologists forgot to include phenomena—e.g., affect, intention—that the structuralists knew to be important.

I disagree with the authors as to what should be done in order to instate a new and more useful paradigm. Teleological causes and dialectical reasoning have been ignored but can be handled by cognitive psychology. If this be the case, then there is no point in rejecting mechanistic explanations in favor of the humanistic explanations proposed by the authors. In this comment, I follow Rychlak's (1977) definition of humanistic psychology: a psychology that does not attempt to reduce Aristotelian final causes (intentions, etc.) to efficient causes (prior forces inducing posterior effects). The problem with cognitive psychology is not really so much with its theoretical structure. I think that the

Requests for reprints should be sent to Colin Martindale, Ph.D., Department of Psychology, University of Maine, Orono, Maine 04469.

problem is more with the topics most cognitive psychologists tend to focus upon. They are often a bit dry and far removed from what many of us consider to be focal human concerns. I have elsewhere tried to show that a cognitive approach can offer us useful insights into such topics (Martindale, 1981). The authors are perfectly correct in their criticisms of the old-fashioned "boxes in the head" cognitive psychology, (e.g., Shiffrin and Atkinson, 1969) but it should be noted that homunculi, CPUs, and the like have been banished in more recent cognitive theories (e.g., Minsky, 1985; Rumelhart and McClelland, 1986).

Lockean vs. Kantian Views of Understanding

The philosophical basis for most British and American psychology can be traced back to Locke's (1690/1894) *An Essay Concerning Human Understanding.* Locke, of course, argued that human mind is a *tabula rasa* upon which experiences are "written." Contiguity leads us to associate certain experiences with others. Kant (1781/1963), in his *Critique of Pure Reason,* proved that the opinions of the British associationists were, to put it mildly, complete nonsense. Locke argued that we learn to perceive in spatial terms by watching objects move around (in space). Kant pointed out that objects could not be perceived in the first place unless mind had the *a priori* capacity to structure experience in spatial terms. We can be certain that space is an *a priori* category of human understanding. We can never know if space is a characteristic of physical reality. Kant established the same for other processes such as time, causality, etc. Thus, our experience and understanding are a joint product of the innate structuring properties of mind and of a completely unknowable external reality. The *Critique* is certainly one of the most important books ever written. Whereas most other philsophers argue and speculate, Kant *proves.*

In many respects, a new Kantian psychology would not be notably different from present-day psychology. This is because, for many small-scale problems, it does not really matter if one is a Lockean or a Kantian. Were this not the case, psychology could not possibly continue to base itself upon an associationistic framework demonstrated over 200 years ago to be completely untenable. To give one simplified example, an experiment on psychophysics done by a Lockean and a Kantian would yield the same results and be written in exactly the same way. The Lockean would believe that he or she knew fairly exactly what the physical stimulus was, whereas the Kantian would not hold this belief.

On a deeper level, however, the new psychology would be fundamentally different. At the basis of a Kantian approach is the belief that the mind must be "pre-wired" to a certain extent. It must have the innate capacity to detect certain features and to perform certain operations. Kant (1781/1963) specified what the most basic of these must be. However, he left for us a large "gray area" between what must logically be innate and what could possibly be learned. If mind were a tabula rasa "camera" that took a composite photograph of every experience,

the result would be a complete blur. On the other hand, if mind had the innate ability to keep different composites for different objects (e.g., faces vs. other objects), meaningful composite photographs of faces would result (compare Galton, 1883; Martindale, 1981). In other words, one does not learn to discriminate faces from other objects. The capacity is innate. To be more precise, *some* capacity allowing the discrimination must be innate—it might be a capacity to discriminate among different combinations of spatial frequencies rather than among features defining face vs. non-face per se. Similar considerations apply to other classes of objects as well.

It is widely agreed that natural categories are defined by prototypes rather than by lists of shared features (Rosch, 1975). Are these prototypes innate or could they emerge from "composite photographs"? Consider color categories. If all color experiences were "averaged" together, we could never learn to discriminate among different colors. Indeed, it does seem to be the case that color categories are innate (compare Berlin and Kay, 1969). By this point, the reader may be getting the sinking feeling that he or she is being led into a Platonic trap, that my next point will be that one cannot learn anything he or she did not already know, that one cannot discriminate a duck from a chicken unless he or she already knew the difference in the first place, etc. Nothing in the Kantian approach requires us to go nearly so far as this. The point is that some primitive and innate features are necessary. By the time we get to the fine points of ducks vs. chickens or one face vs. another one, Kantian and Lockean explanations are not especially different.

Consider Lamiell and Durbeck's question of how one recognizes one's first extravert. Is this a question in the Kantian gray area? I think that it is. Certainly, in order to make sense of other people's behavior, we need some innate prototypes, features, or guidelines. The question is what these are. My first reaction to Lamiell and Durbeck's question was that the innate features would be much more primitive than introvert vs. extravert, that the first extravert would be recognized by a comparison process with previously encountered individuals. Then I realized that I was rather at a loss as to what the more primitive features might be. The clusters of traits grouped under the four temperaments would be one possibility. After all, some version of temperament theory has always been and still is at the basis of both implicit and scientific theories of individual differences (Hogan, 1983). But, of course, extraversion vs. introversion fits temperament theory quite nicely (Eysenck, 1953). Is this the set of innate personality-trait primitives that organizes our experience? The question is an open and interesting one for a Kantian psychologist. Of course, the question verges on heresy for a Lockean psychologist. Lockean philosophy is simply wrong. Kant established that 200 years ago. Even worse, Lockean philosophy forbids us to ask such basic questions and forces us to ask more trivial ones: Once we already know what an extravert is, how do we tell the difference between two of them? Lamiell and Durbeck also ask if we judge absolutely or

relatively. This is a Kantian question. As such, it must be answered. The Lockean assertion that we always judge relatively is not an answer. It is a possibility that, in fact, seems less likely than the Kantian alternative.

Clearly, I agree with the authors that psychology should be Kantian rather than Lockean. However, I am certainly not the only cognitive psychologist holding this belief. Weimer (1973) made an even stronger case than I have presented. Chomsky (1972), of course, makes a thoroughly Kantian argument in the case of language. The notion that learning could not occur without a set of innate primitive features and operations is actually quite common (compare, e.g., Hintzman, 1986; Pylyshyn, 1985). The new discipline of cognitive science (Norman, 1981) composed not only of psychologists but also of computer scientists, linguists, and philosophers gives hope of attacking Lockean tendencies in cognitive psychology. I hope that the philosophers will do a bit of remedial education on their colleagues in psychology.

Is a New Paradigm Necessary?

Considerations of Parsimony

As well as being Kantians, the authors argue that teleology and dialectics must be incorporated into our explanations and that this can only be done within a non-mechanistic humanistic paradigm. I fail to see why it cannot be done in a cognitive paradigm that remains thoroughly mechanistic. Rychlak (1977) argues that mechanistic explanations are not consistent with his introspections. Williams, in his article in this issue, argues that mechanistic explanations rob human life of meaning. These arguments are not cogent. From an evolutionary standpoint, once human beings gained the capacity for self-consciousness, it is fairly clear that belief that one's life had meaning would be tremendously adaptive. Existential Neanderthals who decided that human life is meaningless would not have had their whole hearts in the game of evolution— the "meaningless" cycle of reproduction. If we are meaningless automatons, evolutionary forces would have disguised this fact quite thoroughly. I think that Williams and I may be using the term "meaningful" in quite different ways: for him, gravity has no meaning, it just *is*. For me, gravity is full of meaning, since it explains all sorts of interesting things. On the level of psychology, I want to disallow ultimate meaningfulness to individual human lives. Unless we do so, I do not see how we can construct a meaningful science of psychology. However, parsimony, rather than desire, must decide the issue. If we can integrate teleology and dialectic thought into mechanistic explanations, then we have no need of more complicated humanistic explanations. If we can show that computers behave in a telic manner and "reason" dialectically, then we can conclude either that a humanistic explanation of computers is required or— more reasonably—that teleology and dialectics can be explained mechanistically.

I am not arguing that human mind operates like a computer. The point is that a computer can mimic certain mental activities.

Note that everything Kant said about human understanding applies just as well to computer "understanding." Computers are in fact quintessential Kantian engines. Human beings cannot know reality in itself but only reality as it is structured by mind. Now, a computer cannot know human experience as we know it but only our world as structured by its "mind." For the computer, our reality exists only as a pattern of bits being either on or off in one or another of its registers. No matter how much vision you give a computer, it could never in principle know what handwriting, say, looks like to us. The handwriting is digitized and ultimately "known" to the computer only as patterns of voltages that are in no sense isomorphic to what handwriting looks like. The computer can be programmed to behave as if it could know our reality just as we are programmed to behave as if we know physical reality. Thus, nothing in Kant's analysis of human understanding requires a humanistic as opposed to a mechanistic psychology.

Dialectics

The authors feel, as did Kant, that we should pay more attention to dialectics: *A* is defined only in terms of *not-A*, etc. I agree that this is an important mode of human thought. I do not, however, think that it is as natural, useful, or important as other modes of reasoning. If it were, we should not have so much difficulty in understanding dialectical reasoning such as that of Hegel or Marx. The fact that cognitive psychologists have tended not to study dialectics may imply that they do not agree as to its importance, but it is unfair to imply that they deny its existence. It would be difficult to find anyone who would argue that "up," say, has any meaning unless it is contrasted with "down." It is certainly an important human ability to postulate *not-A*—even though one may never have encountered it and it may not even exist—when presented with *A*. If we are told to do *A*, we may certainly do *not-A* even though it is not in our "behavioral repertoire." This ability need have nothing to do with free will. In principle, the choice should be predictable from past experience: Why should I emit *not-A* unless performance of *A* has not been optimally profitable in the past? Of course, we give up the hope of a scientific psychology unless we equate free will with error variance.

Current digital computers are binary machines. One could say that "on" has meaning only when contrasted with "off." In other words, digital computers are dialectical engines. Such dialectical operations as NOT and NOR are among their most important primitive operations. Even if one disagrees that computers are implicitly dialectical machines, it cannot be denied that they can be programmed to perform dialectical operations.

The authors should recall that the last dialectical school in the human sciences—the structuralism of Lévi-Strauss (1958/1967) and his followers—led

naturally to the delirious hall of mirrors of deconstructionism (Derrida, 1967), where everything is defined in terms of everything else, and meaning of any sort is totally lost.

Associationism

In his article in this issue, Rychlak argues that we cannot hope to explain human behavior on the basis of past associations. I strongly agree with this. However, it may be because the strength of associations between mental elements may not have a very close relationship with how often these elements have co-occurred. Strength of associations may guide our behavior but not because of past learning. As Rychlak notes, elements coding dialectic opposites tend to be associated, and it is unlikely that this is due to frequency of past co-occurrence. On another level, typicality of an exemplar is not determined by its frequency of occurrence (Rosch, 1975). As I implied above, typicality – at least in many domains – may be laid down on an innate pattern. Note, for example, that it is a small step from Cantor and Mischel's (1979) supposedly learned person prototypes to Jung's (1953) innate archetypes. Can we construct a computer program where strength of association is partly predetermined and partly learned? In fact, this is an easier task than writing a program that has to learn everything. Even behaviorists agree that some associations are innately stronger or easier to learn than are others. Recall the Garcia, McGowan, and Green (1972) experiments showing that rats can learn to associate taste aversion with flavor on a single trial if the reinforcer is nausea but that they cannot learn the task if the reinforcer is electric shock.

Teleology

Human behavior is goal-directed. For a pure behaviorist, this does cause a problem of how to get the cause (the goal) occurring before the effect (approaching the goal). I have always admired the Hullian solution of fractional anticipatory goal responses (Hull, 1943). The explanation is rather like a surrealist painting in that it is interesting but not very closely related to any possible reality. As soon as we allow mind back into psychology, I fail to see that there is the slightest problem in getting the cause before the effect: we conceive a certain mental pattern (the goal) and perform operations until "reality" matches this mental pattern (compare Martindale, 1981; Miller, Galanter, and Pribram, 1960; Murray, 1938). Minsky (1985) has called anything – a person or a computer – that engages in such activity a "difference engine." I see no "intelligence loans" (Dennett, 1978) outstanding with such explanations: one mental event (cause) precedes another (effect). Cause and effect may be reversed on the behavioral level but not at the mental level.

An intention clearly precedes the occurrence of the act that was intended. Thus, it would seem to qualify as an efficient cause even though it is, according to Rychlak (1977), by definition, a final cause. So long as we agree that humans have intentions and act in order to achieve goals, I am not sure that the terminology is crucial. However, if we treat the intention as analogous to a "stimulus," I can see how progress toward explanation can be made. If we refuse to treat it in such a way, I am confused as to how we can make any progress and as to what we have gained.

Note that computers are teleological or telic machines. A computer's operations are not "pushed" by whatever has happened in the past but "pulled" toward goals. Any DO WHILE statement is a simple example of this. Most computer programs can be thought of as "difference engines" that have a clear goal—e.g., minimizing a quantity, inverting a matrix, computing a correlation. One could even argue that computers are much more telic than human beings in that a typical computer program has no memory for and is not influenced at all by its own past behavior, whereas—though we are goal oriented creatures—human behavior is at least to some extent "pushed" by memories of prior experiences. A humanistic or teleological explanation of computers would obviously be unwarranted anthropomorphism. Such an explanation would be misleading at best. Though computers act in a telic fashion, this can be explained completely in terms of efficient causation. People are not computers, but I can see the benefits of explaining their behavior in terms of efficient causation. It is unclear how it would advance our understanding of mind if we insist on completely differentiating efficient causes (e.g., stimuli) from final causes (e.g., intentions).

References

Berlin, B., and Kay, P. (1969). *Basic color terms: Their universality and evolution.* Berkeley: University of California Press.

Blumenthal, A. (1975). A reappraisal of Wilhelm Wundt. *American Psychologist, 30,* 1081-1088.

Cantor, N., and Mischel, W. (1979). Prototypes in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology,* Vol. 12 (pp. 3-52). New York: Academic Press.

Chomsky, N. (1972). *Language and mind.* New York: Harcourt Brace Jovanovich.

Dennett, D.C. (1978). *Brainstorms.* Montgomery, Vermont: Bradford Books.

Derrida, J. (1967). *De la grammatologie.* Paris: Editions de Minuit.

Eysenck, H.J. (1953). *The structure of human personality.* New York: Wiley.

Galton, F. (1883). *Inquiries into human faculty and its development.* London: Macmillan.

Garcia, J., McGowan, B.R., and Green, K.F. (1972). Biological constraints on conditioning. In A.H. Black and W.F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 3-27). New York: Appleton-Century-Crofts.

Hintzman, D.L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review, 93,* 411-428.

Hogan, R. (1983). A socioanalytic theory of personality. In M. Page (Ed.), *Nebraska symposium on motivation 1982* (pp. 55-89). Lincoln, Nebraska: University of Nebraska Press.

Hull, C.L. (1943). *Principles of behavior: An introduction to behavior theory.* New York: Appleton-Century-Crofts.

Jung, C. (1953). *Two essays on analytical psychology.* New York: Bollingen.
Kant, I. (1963). *Critique of pure reason.* London: Macmillan & Co. (Originally published 1781)
Lévi-Strauss, C. (1967). *Structural anthropology.* Garden City, New York: Anchor. (Originally published 1958)
Locke, J. (1894). *An essay concerning human understanding.* Oxford: Clarenden, (Originally published 1690)
Martindale, C. (1981). *Cognition and consciousness.* Homewood, Illinois: Dorsey.
Miller, G.A., Galanter, E., and Pribram, K.H. (1960). *Plans and the structure of behavior.* New York: Holt, Rinehart & Winston.
Minsky, M. (1985). *The society of mind.* New York: Simon and Schuster.
Murray, H.A. (1938). *Explorations in personality: A clinical and experimental study of fifty men of college age.* Cambridge, Massachusetts: Harvard University Press.
Norman, D.A. (Ed.). (1981). *Perspectives on cognitive science.* Hillsdale, New Jersey: Erlbaum.
Pylyshyn, Z.W. (1985). *Computation and cognition.* Cambridge, Massachusetts: MIT Press.
Rachlin, H., Logue, A.W., Gibbon, J., and Frankel, M. (1986). Cognition and behavior in studies of choice. *Psychological Review, 93,* 33-45.
Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology,* 104, 192-233.
Rumelhart, D.E., and McClelland, J.L. (Eds.) (1986). *Parallel distributed processing: Explorations in the microstructure of cognition,* Vol. 1. Cambridge, Massachusetts: MIT Press.
Rychlak, J. (1977). *The psychology of rigorous humanism.* New York: Wiley.
Shiffrin, R.M., and Atkinson, R.C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review, 50,* 179-193.
Weimer, W.B. (1973). Psycholinguistics and Plato's paradoxes of the *Meno. American Psychologist, 28,* 15-33.