JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2002, **78,** 597–607 NUMBER 3 (NOVEMBER)

PSYCHOLOGICAL ESSENTIALISM: A REVIEW OF E. MARGOLIS AND S. LAURENCE (EDS.), CONCEPTS: CORE READINGS

DAVID C. PALMER

SMITH COLLEGE

Concepts: Core Readings (1999) is an edited volume of articles representing five major theories of concepts within cognitive psychology. Of the five, only the prototype theory can be interpreted in terms arising from a behavioral analysis. Two conceptual problems pervade nearly all of the papers in the volume. The first is an implicit commitment to essentialism; concepts are reified and are commonly discussed as though they have essential properties. Second, the theories suffer from a defective theory of verbal behavior. No distinction is made between verbal operants and words drawn from a hat.

Key words: concepts, essentialism, prototype theory, stimulus classes, units of analysis, verbal behavior

Concepts: Core Readings (Margolis & Laurence, 1999) is a collection of 26 articles representing the several dominant theories on the topic within the field of cognitive psychology. I was invited to review the book for this special issue of the *Journal of the Experimental Analysis of Behavior* to provide a contrasting viewpoint to the other articles and perhaps, more usefully, to suggest common themes or to identify possible sources of empirical support for work in our field. When I accepted the assignment, I intended to offer an entirely constructive review that extracted what was of value, ignored the rest, and refrained from polemics, but I found little in the book to nourish that ambition. I have usually profited from forays into the cognitive literature, particularly in psycholinguistics, but the papers in this collection are mainly conceptual rather than empirical, and at that level, there is little common ground. The editors of the volume seem to agree; in a discussion of various schools of thought (Laurence & Margolis, 1999), the behavioral position appears only in a footnote: ''Another alternative is the view that concepts [are] behavioral or psychological abilities. We take it that behavioral abilities are ruled out for the same reasons that argue against behaviorism in general (see, e.g., Chomsky, 1959)'' (p. 6). The irrelevance of the behavioral position is regarded as an axiom and is discussed no further.

Nevertheless, I have tried to resist the urge simply to criticize the book, although I was continually exasperated by undefined or meaningless terms and by the innocent cheerfulness with which they were invoked, but our field does not need another diatribe against cognitive psychology. Rather, I have tried to use the book as a background for an analysis of assumptions and definitions. Consequently, in the first part of this article, I summarize, with relatively little editorial comment, the main themes of the book, leaving most of the evaluation to the reader. In this section, I have elected to adopt the terminology of my sources, but the reader should not infer that I regard either the terminology or the claims to be satisfactory as stated; objections to them are familiar and need not be repeated here. In the second part of the article, I identify several facts about behavior and biology that any theory of concepts must accommodate. In particular, I point out that variability in nature is fundamental, not accidental, and that essentialistic theories are inconsistent with that fact. I argue further that, with regard to concepts, a distinction must be made between two kinds of definition: those we must discover and those we can prescribe. Products of contingencies of selection must be defined by discovery, that is, by empirical investigation; the tools of formal analysis are inappropriate. The methodology of behavior analysis, shaped by its subject

Margolis, E., & Laurence, S. (1999). *Concepts: Core readings.* Cambridge, MA: MIT Press.

Correspondence concerning this article should be addressed to the author at Department of Psychology, Smith College, Northampton, Massachusetts 01063 (e-mail: dcpalmer@smith.edu).

matter, is well suited to interpret concepts arising from such contingencies.

OVERVIEW

Concepts opens with an excellent introductory chapter by the editors that summarizes, critiques, and integrates most of the papers that follow. In fact, this long chapter is less an introduction than a centerpiece; the other articles seem to be mere satellites selected to illustrate and expand points made here. The chapter is indispensable for the nonspecialist, because most of the papers in the volume are ponderous, and common threads are difficult to identify. I found the highlighted summaries of theories and their shortcomings to be particularly helpful. The remainder of the book is divided into sections that discuss one or another theoretical school. These include the classical theory, the neoclassical theory, prototype theory, conceptual atomism, and one whimsically dubbed the ''theory-theory.'' The authors argue that the different theories have different goals, and that cooperation across disciplinary boundaries within cognitive psychology will be necessary to integrate the various points of view into an adequate hybrid.

The term *concept* is never clearly defined. Laurence and Margolis suggest that concepts are ''subpropositional mental representations.'' They hold that propositions are too big to be concepts; thus *the chair with brocade upholstery needs to be repaired* is a sentence, not a concept. But *chair with brocade upholstery* might be a concept, albeit a complex one. ''Lexical concepts,'' such as *chair* and *brocade*, correspond roughly to content words in a language, and it is these that provide most of the examples in the book. But concepts are not simply words; the correspondence is said to be ''rough'': Carey (1991/1999, chap. 20) claims that ''concepts are units of mental representation roughly the grain of single lexical items, such as *object*, *matter*, and *weight*'' (p. 460). Jackendoff (1989/1999, chap. 13) argues that the term concept resists simple definition, because

> The term plays a certain role in a larger world view that includes the nature of language, of meaning, and of mind. Hence the notion of concept cannot be explicated without at the same time sketching the background against which it is set; and the ''correctness'' of a particular notion of concept cannot be evaluated without at the same time evaluating the world view in which it plays a role. (p. 305)

These attempts to wrestle with a definition reveal that the term *concept* is seen by some of these authors as a natural phenomenon whose properties must be discovered rather than a formal category whose properties can be prescribed by definition. Thus the term *concept* is analogous to *eggplant*, which can be understood only by studying nature, but is unlike *hexagon*, which can simply be assigned a definition by a geometer as part of a formal system. In this view, then, concepts are ''things,'' things that are the proper subject of study. Jackendoff, on the other hand, seems to regard concepts as an abstract part of a theory of mind and not something with fixed properties that will reveal itself to an investigator. Most of the other authors simply skirt the problem of definition, perhaps regarding the term as too familiar to need definition. The uncertainty of the nature of the topic under discussion haunts the volume and is particularly troublesome as one moves through the book from one author to the next.

THE CLASSICAL THEORY

Laurence and Margolis offer this statement of the classical theory: ''Most concepts (especially lexical concepts) are structured mental representations that encode a set of necessary and sufficient conditions for their application, if possible, in sensory or perceptual terms'' (1999, p. 10). The qualifiers ''most'' and ''if possible'' reveal that this is offered, not as a prescriptive definition, but as an empirical claim. The theory presupposes that one can identify examples of concepts, that one can determine whether they are structured mental representations and whether they encode sets of sensory or perceptual conditions. According to this theory, concepts are composed of simpler elements, which can ultimately be reduced to primary sensory data. We identify a concept by listing necessary properties, functions, and conditions. Katz (1972/1999) offers the following properties for *chair*: ''Object, physical, nonliving, artifact, furniture, portable, something with legs, something with a back, something

with a seat, seat for one'' (p. 135). This theory suggests that the required features can be used to sort events into examples and nonexamples, and that concepts can be used to draw inferences, because the features are implied by the whole. (If Rover is a dog, Rover must have four legs.) But concepts are not just words; they are ''structured mental representations,'' and words derive their meanings from their associated concepts.

The classical theory ran into trouble before it was ever formulated as a theory. A passage from Plato was chosen to lend classical authority to the problem (1999, chap. 2).[1] In this selection, Socrates seeks to learn from Euthyphro the meaning of *piety* and is told that it is that which is loved by the gods. Socrates wonders what makes the gods love something and presses Euthyphro for a list of the essential properties of the concept. But having brought the gods into the definition, he can find no way to get them out; Euthyphro talks in circles and departs, leaving Socrates smugly unenlightened.

The point of this selection is that many concepts resist satisfactory definition in terms that are any simpler than the term to be defined. The concept of *piety* seems irreducible to elements of sensory data; likewise, *seat for one* seems to be no simpler than *chair*. One might add that *subpropositional mental representation* is no advance over *concept*. Quine (1953/1999, chap. 5), Wittgenstein (1953/1999, chap. 6), and Putnam (1970/1999, chap. 7), in papers with behavioristic flavor, point to other difficulties with the theory. In a well-known passage, Wittgenstein argues that many concepts (e.g., *game*) do not have defining properties that are sufficient to determine the scope of the word; exemplars seem to be related by family resemblance rather than fixed characteristics. Quine's paper was included in the volume because the classical theory assumes that concepts are analytic, that is, that they can be translated into equivalent expressions or broken down into component elements with no loss, without reference to experience. Quine holds that the assumption that statements can be divided into the analytic and the synthetic, between those that must be true logically and those that conform to experience, is just an article of faith. My own loose translation and extension of his rather formal argument is that language as a set of symbols is meaningless; it can be understood only in relation to a web of verbal contingencies. From this perspective, the analysis of concepts is not a job for the logician but for the behavior scientist. Putnam offers related arguments and concludes that an adequate theory of concepts ''is not to be expected until one has a general and precise model of a language-user'' (p. 187).

[1] The appearance of Plato in this volume reminds one of Skinner's lament about the primitive state of psychology relative to other sciences: ''Greek physics and biology are now of historical interest only (no modern physicist or biologist would turn to Aristotle for help), but the dialogues of Plato are still assigned to students and cited as if they threw light on human behavior'' (1971, p. 5).

THE NEOCLASSICAL THEORY

The neoclassical theory is not a clearly articulated position; it is just the name given by the editors to a family of modern papers that share some features of the classical theory, and I can make few general statements about it. Laurence and Margolis describe it thus: ''Most concepts (especially lexical concepts) are structured mental representations that encode partial definitions, i.e., necessary conditions for their application'' (1999, p. 54). Of the four papers included in the volume under this heading, Jackendoff's is the most explicit. His purpose is to integrate a theory of concepts into a general theory of generative linguistics. He argues that concepts must be combinatorial according to certain innate rules and principles in order to permit the expression of an infinite number of meanings. The units of conceptual structure are categories such as thing, event, state, place, path, property, and amount (Jackendoff, 1989/1999, p. 315). These categories are more than just bins; they provide a framework of relations within which individual concepts find a place. For example, the verbs *kill*, *lift*, *give*, *persuade,* and many others are all examples of causatives, verbs that indicate that one variable causes another to change its state. A concept in the *path* category entails a relation to a place of origin and a destination. From this perspective, concepts can be understood only in relation to the mental structures that determine the combinatorial rules of concepts.

Like Chomsky's theory of syntax, to which Jackendoff's semantic theory is explicitly wedded, the theory is far removed from behavioral principles. It is taken for granted that there are structures in the mind, presumably innate, that impose order on one's conceptual development. I am inclined to agree with Jackendoff that such a theory can be evaluated only in relation to the paradigm in which it is embedded. Nevertheless, even nativists consider Jackendoff's theory incomplete. One of the virtues of the classical theory was that it defined the extension of a concept; that is, members could be sorted from nonmembers, something not possible in the neoclassical theory. Furthermore, in Jackendoff's theory some concepts, such as *cause*, *thing*, and *event*, appear to have special status that other concepts do not. Some concepts, by hypothesis, are part of the architecture of the mind; others become meaningful only relative to that architecture. There is a whiff of circularity to this scheme.

THE PROTOTYPE THEORY

The prototype theory was formulated in the 1970s as a response to the difficulties with the classical theory. Laurence and Margolis describe the position thus: ''Most concepts—including most lexical concepts—are complex representations whose structure encodes a statistical analysis of the properties their members tend to have'' (p. 27). Concepts can be thought of, not as a list, but as a distribution of properties, some more central or typical than others. The prototype is an abstraction of the central properties and need not correspond to any example. A robin is a better example of a bird than a penguin or a duck, if not logically, at least psychologically. That is, people more readily identify a robin as a bird, more often cite it as an example, and call it more typical of the concept.

Whereas the classical theory can be viewed as a formal theory about the nature of the world, the prototype theory is a psychological theory about our relationship to the world. What one regards as a central exemplar of a concept clearly depends upon one's experience. Today the prototypical computer sits on a desk; 30 years ago, it would have filled a large room. In New England, the prototypical egg is brown; in New York, it is white. The prototype theory points away from the classical theory's antiseptic world of logical categories and set theory, toward the messy, idiosyncratic world of stimulus and response classes, emerging or blurring with one's interactions with the environment.

One might have expected the formulation of the prototype theory to lead to an inquiry into the variables that affect ''concept formation,'' an inquiry that would necessarily touch, at least obliquely, on basic behavioral processes. Indeed, early proponents of the theory (e.g., Rosch & Mervis, 1975) were inclined to just such an inquiry. For example Rosch (1978/1999, chap. 8) summarizes evidence showing that ''prototypicality'' is systematically related to response rate, rate of acquisition, priming effects, and other measures of response strength. Furthermore, she states, ''to speak of a *prototype* at all is simply a convenient grammatical fiction. . . . To speak of a single entity that is the prototype is either a gross misunderstanding of the empirical data or a covert theory of mental representation'' (p. 200). The prototype theory seems promising, but if the present volume is representative of the field, it is regarded as inadequate.

For example, Armstrong, Gleitman, and Gleitman (1983/1999, chap. 10) argued that the data supporting the prototype theory were irrelevant to the real structure of concepts. They presented subjects with exemplars of categories with well-defined formal definitions, such as *odd number* or *plane geometry figure*. Latencies to identify *3* and *square* were shorter than *447* and *trapezoid*, respectively. That these data conflict with the logical necessity that all odd numbers are equally odd was taken as evidence that such data are irrelevant to the structure of concepts in the mind. Here logic trumps the data, and a reified notion of concept overrides behavior.

Others criticized the prototype theory for differentiating poorly between members and nonmembers. ''A plucked chicken is still a chicken,'' goes the argument, ''but it is nothing like the prototype.'' In this example, how a plucked chicken is classified is regarded as a fact, not a tact. Some argued that many concepts, particularly complex concepts, do not have central tendencies at all (e.g., *purple cow, George Washington's cousins*). Fodor, a champion of nativism and essentialism, believes

that concepts must be compositional in order to explain why they are ''productive,'' that is, why it is that new concepts can be created by stringing words together, irrespective of context. As an example of a compositional concept that cannot be understood as the set of prototypes of individual elements, Fodor and Lepore (1996) offer *pet fish who live in Armenia and have recently swallowed their owners*, but they do not explain why such a concatenation of words should be considered a concept at all. A second illuminating example of this viewpoint is the discussion of *apple that is not an apple*. This phrase is soberly advanced as a concept that poses a problem for prototype theory (Osherson & Smith, 1981/1999, chap. 11).

All of these critiques suffer from a defective theory of verbal behavior. Strings of words drawn from a hat are taken as seriously as tacts and mands. This error flits like a poltergeist through the book, causing trouble at every turn of the page. The term *concept* has lost all its moorings to psychology when *apple that is not an apple* is discussed as a test case (cf. Moore, 2000). Nevertheless, consideration of such examples led to dissatisfaction with the prototype theory and to the development of alternatives. Thus these alternatives were born in confusion.

THE THEORY-THEORY

According to the ''theory-theory,'' concepts play a role in our theories about the world and can be understood only with respect to those theories. For example, after Newton, the concept of *weight*, once absolute, became relative. That is, because weight was determined by gravitational attraction, it would vary according to one's position relative to other bodies. Modern physicists modified their concepts of time and space in the course of overhauling their theories about relative motion. In less exalted domains, children's theories of the world are presumed to change as they age. A young child may call all furry animals *cats*, for example, suggesting a different taxonomy from that of adults.

Most people go about their lives without reflecting much on Newton, Einstein, and Darwin, and therefore subscribe to theories about nature and natural categories only implicitly. According to Medin and Ortony (1989), a feature of most such theories is ''psychological essentialism,'' the position that everything has a special defining property that determines what it is. Most people think that animals retain their identity despite any change they might suffer in their phenotype. They erroneously think of the genotype as specifying the essential nature of organisms. Gelman and Wellman (1991/1999, chap. 26) found that children as young as 3 years often distinguish between internal and external changes to something. Most of them thought that a dog whose insides had been removed was no longer a dog, but that a dog without fur was still a dog.

The theory-theory accommodates people's tendency to classify things as if they had essential properties, but unfortunately, the ''theory'' that is supposed to be reflected by this fact is just a fiction inferred from people's behavior. Most people no more have theories about the nature of organisms than bats catching insects have theories about acoustics. To say that a child overgeneralizes *cat* because of a defective theory reveals a taste for tautology and a puzzling disregard for basic processes of generalization and discrimination.

CONCEPTUAL ATOMISM

The main tenet of conceptual atomism is that although concepts can be combined into higher order concepts, most lexical concepts cannot be decomposed into smaller parts. *Bird* is not a structured concept with properties such as *flies*, *has wings*, *feathered*, and so on, but is an integral mental representation, an element of the language of thought. This theory is the most abstract of all of the cognitive theories and is the most carefully reasoned. It is a position forced by logical argument from a set of unspoken assumptions about the mind. The principal architect of the theory is Fodor, and his main achievement is a reductio-ad-absurdum argument against the prevailing assumptions of the cognitive position.

Fodor and his colleagues reject the classical theory, namely, that concepts are structured entities built up of primitive elements that themselves can be derived from sensory data. They observe that our definitions of things seldom show a tendency to reduce to sensory

data. To illustrate the point, they cite dictionary definitions of *grandmother* as *mother of one's mother or father*; *kill* as *cause to die*; *ketch* as *two-masted sailing vessel such that the aft (or "mizzen") mast is stepped forward of the helm* (Fodor, Garrett, Walker, & Parkes, 1980/1999, chap. 21). Raw sensations play little role in such definitions. In any case, we are driven, eventually, to a set of primitive concepts that cannot be analyzed further. Of the problem of concept acquisition, Fodor et al., conclude,

> It seems to us that there is only one possible answer: theories of concept learning *presuppose* the availability of the primitive conceptual basis; they don't explain it. If, however, the primitive basis is presupposed in concept learning, then it cannot itself be learned. If it is not learned, then, presumably, it is innate. (p. 504)

The scope of this claim is admittedly uncertain, but Fodor et al. make no secret of their preferences: "Whatever is not *definable* must be innate. . . . We've argued that morphemically simple expressions are typically undefined, that undefined expressions typically express primitive concepts; and that primitive concepts must be innate" (p. 510).

According to Laurence and Margolis, most cognitive scientists reject this conclusion, whatever they might think of the logic of the argument. For our purposes, it is enough to note that the notion of *concept* embraced by Fodor's arguments is that of a mental symbol, not that of a response class. The force of his conclusions is therefore restricted to such symbols; there is no reason to suppose that such arguments are relevant to behavior.

THE LURE OF ESSENTIALISM

The central and pervasive error in this book is a misunderstanding of the nature of its subject matter. The psychological essentialism attributed to folk psychology appears in the paradigm itself. Throughout the book concepts are discussed as reified abstractions; they are not stimulus classes, response classes, a relation between the two, or any other variable amenable to operational definition, nor are they open to or derived from direct observation. Rather, they are said to be structured mental representations. The danger with such abstractions is not that they are mental, or hypothetical, or cognitive, but that they invite one to slip unwittingly into essentialism, the view that classes of natural phenomena have a set of necessary and sufficient properties that determines class membership. This, of course, is the classical theory of concepts, a theory that most of the book is dedicated to refuting. *Concepts* presents one with the paradoxical spectacle of arguments favoring alternatives to the classical theory that are couched in essentialist locutions.

For example, it is held that a central task of a theory of concepts, among other things, is to determine reference. "What makes it the case," the editors ask, shortly after rubbishing the classical theory, "that DOG applies to all and only dogs?" (Laurence & Margolis, 1999, p. 55). This question illustrates how insidiously abstractions can become reified and how insensibly reifications become essentialistic. The concept *dog*, originally an abstraction, has become an agent, something that points to dogs. Worse, the object pointed to, "all and only dogs," is itself an essentialistic fiction. The class of "all and only dogs" is not to be found in nature; it can be found only in an erroneous theory. Darwin has persuaded most scientists not just that it is futile to look for the boundary between *dog* and *not dog* but that the distinction is meaningless. Boundaries between classes are not a feature of nature but of our models of nature. In our models, classes may have essential properties, but natural classes, particularly products of selection contingencies, do not.

To illustrate the subtle fallacy of essentialist locutions consider a slight modification of the editors' question: "What makes it the case that *herring gull* applies to all and only herring gulls?" In Great Britain the herring gull, *Larus argentatus*, is a separate species from its cousin, the lesser black-backed gull, *Larus fuscus*, and the two species do not typically interbreed. In Canada, the herring gull has some black markings not typically found in British gulls. In Alaska and Siberia, the gulls show more black markings and are a bit smaller. As one continues westward, the gulls become more and more similar to the lesser black-backed gull. When one reaches Great Britain again, the gulls have differentiated enough that they no longer interbreed with the herring gull, and we call them a separate species (Dawkins, 1993; Ridley, 1993). Following the two species around the globe is rather

like tracing a Mobius strip; somehow you end up where you started, having seen both sides but without ever crossing from one to the other.

This is just one example of several ''ring species'' that might be adduced in this context. The common feature of such examples appears to be this: A population makes a living along the margin of some inhospitable region like the Arctic Circle or the Gobi Desert. As it spreads out along this margin, opposite ends of the population become far removed from one another geographically and can become differentiated. Eventually the population spreads to the point that the inhospitable region has been entirely surrounded. If differentiation has progressed enough, the ends of the population may find themselves in separate ecological niches and reproductively isolated from one another. Even if chance mating between members of the two groups were to produce offspring, such offspring might be poorly equipped to compete with their ''purebred'' cousins in either ecological niche, and the isolation of the two groups could continue indefinitely.

It is evident that the editors' question is ill conceived. ''All and only'' is a qualifier appropriate to formal models, not to natural categories like dogs. A species need not be a ring species to illustrate the point. Is a coydog a dog or a coyote? Are there two species or just one? There is no fact of the matter. It is only in a defective model of nature that such questions arise. To try to answer them is to misunderstand the nature of classes formed by contingencies of selection. Unfortunately, although an occasional author is innocent, tacit essentialist assumptions can be found on nearly every page of the text. Concepts are invoked to support categorization, analytic inference, reference determination, and epistemic justification (logical classification); these are exercises in formal analysis for which essentialist categories are required, but for which concepts are ill suited.

Why is psychological essentialism so pervasive, even among scientists who no doubt regard themselves as staunch Darwinians? I consider three reasons. First, the fundamental variability of nature is obscured when contingencies are stable; variability can be seen as error (and is usually treated as such in our statistical models). Second, behavior analysis is virtually alone among psychological schools in adopting a methodology that identifies units of analysis empirically rather than a priori; a priori definitions do not capture the generic nature of fundamental terms. Third, when behavior is defined topographically, in isolation from its controlling variables, entirely different response classes can be conflated. All instances of a given form of behavior are considered equivalent, an error that is particularly serious when the behavior is verbal. Words become things, and strings of nonsense are interpreted as though they were verbal operants.

Variability in Behavior Is Not Error

As Skinner (1953) and many others have observed, classes of behavior are the product of selection processes, be they phylogenetic (natural selection) or ontogenetic (selection by reinforcement). But a selection contingency is a sieve, not a blueprint. Unlike a blueprint, a sieve has no way to determine that everything that passes through it is of the same kind. When we rinse lettuce, both grit and water pass through the colander. Thus a selection contingency cannot produce something with essential properties; variability is inevitable.

Consider three contingencies of selection, from the domains of reinforcement, evolution, and selective breeding. A contingency of reinforcement in an experimental chamber may strengthen behavior that meets certain properties of force, latency, or location. A contingency of natural selection may differentially favor organisms that are able to endure harsh winters or those that are adept at escaping large cats. A nurseryman, selectively breeding raspberries, may use sweeter berries, more disease-resistant canes, or greater productivity as criteria for choosing the breeding stock for the next generation. In each case, selection looks backward, sorting among a pool of candidates according to some criteria; it does not specify how the candidates should meet those criteria. There are many ways to press a lever, endure a winter, escape a large cat, or satisfy a nurseryman; the selection criteria are blind to these differences. Thus, even when contingencies are stable, variability is tolerated by selection contingencies and is therefore inevitable.

But an even stronger claim can be made:

When selection contingencies change along some trajectory—for example, when we require more and more forceful lever presses, when the weather gets colder gradually over many centuries, when large cats become more numerous or become better hunters, when the nurseryman becomes harder to please—lineages of behavior or organisms will often track the changing contingencies in the familiar processes of the shaping of behavior, the evolution of organisms, and the specialization of domestic plants and animals. To the extent that the event of interest is the product of such systematic processes, variability is not just tolerated, it is necessary. Shaping, evolution, and the selective breeding of new horticultural strains would be impossible without variability in the relevant substrate. Thus, variability in nature is not error, nor is it merely inevitable; rather, it is fundamental.

In a period of stasis, when contingencies are stable for a long time, variability, though inevitable, may be detrimental. A random mutation is more likely to impair functioning than improve it. Under such conditions, products of selection may seem to have essential properties, for variations will look like mistakes. The temporal window through which we look at the world is so narrow that this error is easy to make. This may explain the ''psychological essentialism'' that is said to be so pervasive in everyday discourse; more ominously, it may explain the tendency of psychologists to make the same error when formulating theories of concepts. When we consider the ephemeral nature of current conditions it becomes clear that variability is fundamental to living things and that our theories of concepts must accommodate that fact (see Palmer & Donahoe, 1992, for further discussion of these points).

Empirical Versus a Priori Definitions

The fundamental variability that arises from selection processes is respected by Skinner's methodological precept that we must define our units of behavior and its controlling variables empirically rather than a priori (Skinner, 1935). He observed that when one does so, one finds that the relation between behavior and other variables is most orderly when our definitions of units embrace a range of values. If we define our units narrowly, permitting little or no variability, the order in our observations diminishes or disappears.

As an illustration, consider the following classroom demonstration: A hungry pigeon is exposed to a differential-reinforcement-of-low-rate (DRL) contingency in which the responses that are reinforced are only those that follow other responses by 5 s or more. After several sessions of this schedule, behavior becomes orderly: A cumulative record of all key pecks shows a low but steady rate of responding. If, however, one counts only those pecks that follow the preceding peck by, for example, 5.0 to 5.1 s, the cumulative record would be erratic; such responses do not occur with the systematic regularity of responses defined more broadly. An analysis of interresponse times in DRL schedules typically reveals that most responses (neglecting very short-latency responses) fall in a distribution several seconds on either side of the DRL criterion. By Skinner's criterion, this distribution of values is an appropriate defining feature of behavior in this experiment. A similar demonstration can be offered for the generic nature of discriminative stimuli, as confirmed by generalization gradients observed following the conditioning of discriminated operants. Thus, the constructs of *stimulus class* and *response class,* empirically defined, are perfectly suited to natural phenomena that have emerged from contingencies of selection.

Empirical definitions are those that must be discovered. In contrast, a priori definitions are those that are prescribed in advance. One can set up a formal model of nature, complete with assumptions, definitions, and rules of inference. In such models, terms may have essential properties and terms can be given a priori definitions. For example, in geometry a hexagon can be given a formal definition, namely, *equilateral six-sided polygon.* One can perform formal operations on hexagons, so defined; one can calculate the area, the length of a bisecting line, the magnitude of interior angles, the ratio of the areas of two hexagons, and so on. These operations may be useful to us to the extent that our model is like nature, but such essentialist categories do not capture the variability found in nature. For example, if we look at the cells of a honeycomb carefully enough, we will notice

that the cells do not quite satisfy our a priori definition of hexagons. There are tiny variations from one cell to the next; that angle is slightly rounded; this side is slightly shorter than that one. An empirical definition is required, not a formal one, but it is a rare scientist who respects the distinction.

When the object of study is an abstraction of human categorization itself, the matter of definition is central. Most of the papers in *Concepts* pay no heed to the distinction between a priori and empirical definitions, either implicitly or explicitly, and this failure leads to costly confusion. Intrinsically variable phenomena are interpreted as having essential properties. An adequate theory of concepts must respect the generic nature of stimulus and response classes.

Words Are Not Things

Because of a taste for a priori definitions, psychologists tend to classify behavior only by topography, ignoring the complexity of antecedent–behavior relations (Moore, 2000). But a response of a particular form does not reflect the potential variability of controlling stimuli. This circumstance is particularly troublesome when behavior is verbal, because verbal behavior can be transcribed to serve as controlling variables for behavior at a later time. The reader is removed even further than the listener from the variables that control the original verbal operant. It is easy to forget that texts are products of behavior, and that notions such as ''reference'' can be understood only with respect to the relation between the original verbal response and its controlling variables. Moreover, texts that are produced by a cat walking over a keyboard or by an experimenter stringing together words from bins labeled ''noun,'' ''adjective,'' and so on, are not verbal behavior at all and have no meaning.

As noted earlier, traditional interpretations of verbal behavior lead to considerable confusion, because they encourage discussion of strings of words as though they were discriminated operants. Thus, in *Concepts* phrases such as *striped apple, U.S. monarch,* and *4th century saxophone quartet* evoked considerable debate (Laurence & Margolis, 1999, p. 36). It is hard to see how anything sensible can be said about such phrases, for they give the illusion of meaning but are meaningless. Distinctions among the classes of verbal operant, such as *tact*, *textual*, *intraverbal*, and so on, are just what is required to organize an interpretation of the relation between verbal responses and concepts.

CONCEPTS AS STIMULUS CLASSES AND RESPONSE CLASSES

Concept is not a technical term in behavior analysis, although it is common to speak of a stimulus class as a concept; it is less common, but consistent, to speak of a response class as a concept as well. We are likely to use the term when a stimulus class is complex and cannot be easily described by identifying a few stimulus dimensions, particularly if the stimulus class controls our own behavior differentially. We are less likely to speak of the many olfactory ''concepts'' of a bloodhound, or of the auditory ''concepts'' of a bat, although, for all we know, such stimulus classes might be quite complex.

Human operant research that investigates concepts as stimulus classes is flourishing, particularly in the topic of stimulus equivalence, but some of the fundamental processes are best revealed in animal research, in which tight experimental control is possible. For example, Herrnstein and Loveland (1964) and Malott and Siddall (1972) trained pigeons to respond differentially to pictures containing people in a wide variety of dress, settings, and poses, a stimulus class that defies simple description. In that differential responding was acquired over the course of these experiments, it is tempting to assume that the training procedures established pictures of people as a stimulus class, or as a concept, but because pigeons had prior exposure to people, it can be argued that the procedures merely created a new function for a preexisting stimulus class, perhaps even an innate stimulus class. This interpretation, however, can hardly be maintained for subsequent experiments. Herrnstein, Loveland, and Cable (1976) replicated the experiment using pictures of, among other things, a single unfamiliar individual. Pictures of other individuals were included among the nonexamples. Watanabe (2001) trained pigeons to respond differentially to paintings by Van Gogh and Chagall, replicating and extending an earlier study

that showed that pigeons could discriminate impressionist and cubist art. In all of these experiments, pigeons generalized to novel exemplars of the relevant stimulus classes, and at least in the latter two experiments, it cannot be argued that the stimulus classes were innate or had been established prior to the experiment.

These studies avoided the essentialistic pitfalls discussed earlier. In the Herrnstein and Loveland (1964) experiment, for example, the effective stimulus class was determined empirically by ranking the trials by response rate. Two clear stimulus classes emerged, but their boundaries overlapped, revealing the full variability expected from selection contingencies. Thus, although the experimenters had an a priori notion of the concept *person*, the behavior of the pigeons was the ultimate arbiter of the effective class. Moreover, because these studies rest upon a broad foundation of research on stimulus generalization and discrimination, the validity of the theoretical terms adduced is well established.

Missing from the analysis of these experiments is any suggestion that the pigeons acquired structured mental representations, that they encoded a set of necessary and sufficient conditions for the application of concepts, or that they organized their experiences to conform to a mental theory. In short, even if research on stimulus classes were not tainted by its association with behaviorism, it is unclear that it would excite the editors of *Concepts*. But the essentialist notion of concepts is a mirage; it lures the seeker, but there is nothing there.

CONCLUSION

Of the five theories of concepts featured in the book and summarized here, four suffer from the error of imputing essential qualities to behavioral phenomena, hence to products of contingencies of selection. It is taken for granted that a theory of concepts should permit ''reference determination,'' that is, the definitive sorting of candidates into examples and nonexamples. From a selectionist perspective, this is simply an error. Variability is fundamental to products of contingencies of selection, and this variability must be respected by both the methodology and the theoretical constructs of our field. Moreover, none of the proponents of the various theories seem aware of the importance of distinguishing between words as independent things and words as behavior. The merits of various theories are evaluated by inferences about strings of words that have no relevance to verbal behavior. These shortcomings impose a gulf that cannot be bridged between such theories and a behavioral theory of concepts. The fifth theory, prototype theory, is the exception. It would be fairly easy to translate the prototype theory into a theory of response classes, response strength, response differentiation, and stimulus discrimination. The papers included in the present volume to represent the prototype theory do not themselves offer anything of value to the behavior analyst, but they allude to an experimental literature that might do so. The empirical work of Eleanor Rosch, in particular, seems promising. It is revealing that in the present volume the prototype theory is considered flawed for reasons that I take to be strengths. The book is a triumph of psychological essentialism.

REFERENCES

Armstrong, S., Gleitman, L., & Gleitman, H. (1999). What some concepts might not be. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 225–259). Cambridge, MA: MIT Press. (original work published 1983)

Carey, S. (1999). Knowledge acquisition: Enrichment or conceptual change? In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 459–487). Cambridge, MA: MIT Press. (original work published 1991)

Chomsky, N. (1959). Review of Skinner's *Verbal Behavior*. *Language, 35,* 26–58.

Dawkins, R. (1993). Gaps in the mind. In P. Cavalieri & P. Singer (Eds.), *The great ape project* (pp. 80–87). New York: St. Martin's Press.

Fodor, J., Garrett, M., Walker, E., & Parkes, C. (1999). Against definitions. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 491–512). Cambridge, MA: MIT Press. (original work published 1980)

Fodor, J. A., & Lepore, E. (1996). The red herring and the pet fish: Why concepts still can't be prototypes. *Cognition, 58,* 253–270.

Gelman, S., & Wellman, H. (1999). Insides and essences: Early understandings of the nonobvious. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 613–637). Cambridge, MA: MIT Press. (original work published 1991)

Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concept in the pigeon. *Science, 146,* 549–551.

Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976).

Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 2,* 285–302.

Jackendoff, R. (1999). What is a concept, that a person may grasp it? In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 305–333). Cambridge, MA: MIT Press. (original work published 1989)

Katz, J. (1999). On the general character of semantic theory. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 125–149). Cambridge, MA: MIT Press. (original work published 1972)

Laurence, S., & Margolis, E. (1999). Concepts and cognitive science. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 3–81). Cambridge, MA: MIT Press.

Malott, R. W., & Siddall, J. W. (1972). Acquisition of the people concept in pigeons. *Psychological Reports, 3,* 3–13.

Margolis, E., & Laurence, S. (1999). *Concepts: Core readings.* Cambridge, MA: MIT Press.

Medin, D., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). New York: Cambridge University Press.

Moore, J. (2000). Words are not things. *The Analysis of Verbal Behavior, 17,* 143–160.

Osherson, D., & Smith, E. (1999). On the adequacy of prototype theory as a theory of concepts. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 261–278). Cambridge, MA: MIT Press. (original work published 1981)

Palmer, D. C., & Donahoe, J. W. (1992). Essentialism and selectionism in cognitive science and behavior analysis. *American Psychologist, 47,* 1344–1358.

Plato. (1999). Euthyphro. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 87–99). Cambridge, MA: MIT Press.

Putnam, H. (1999). Is semantics possible? In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 177–187). Cambridge, MA: MIT Press. (original work published 1970)

Quine, W. V. O. (1999). Two dogmas of empiricism. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 153–169). Cambridge, MA: MIT Press. (original work published 1953)

Ridley, M. (1993). *Evolution.* Boston: Blackwell.

Rosch, E. (1999). Principles of categorization. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 189–206). Cambridge, MA: MIT Press. (original work published 1978)

Rosch, E., & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605.

Skinner, B. F. (1935). The generic nature of the concepts of stimulus and response. *Journal of General Psychology, 12,* 40–65.

Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.

Skinner, B. F. (1971). *Beyond freedom and dignity.* New York: Knopf.

Watanabe, S. (2001). Van Gogh, Chagall, and pigeons: Picture discrimination in pigeons and humans. *Animal Cognition, 4,* 147–151.

Wittgenstein, L. (1999). Philosophical investigations, Sections 65–78. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 171–174). Cambridge, MA: MIT Press. (original work published 1953)

Received May 10, 2002
Final acceptance May 31, 2002