

## **Essentialism and Selectionism in Cognitive Science and Behavior Analysis**

David C. Palmer *Smith College*  
John W. Donahoe *University of Massachusetts/Amherst*

*Contingencies of selection, be they phylogenetic or ontogenetic, merely set boundaries on units; they do not provide blueprints. Thus, variability is fundamental to all units of selection. Skinner, by characterizing the units of analysis in behavior as generic in nature, established his science squarely within the selectionist paradigm, thereby avoiding the tendency, common throughout psychology, to slip into essentialist analyses. The distinction between essentialism and selectionism is refined in this article, and prominent examples of essentialism in linguistics, theories of memory, theories of representation, associationism, and even in behavior analysis are identified. Recent trends in cognitive science—specifically, research on adaptive networks—are amenable to a selectionist interpretation, suggesting the possibility of future fruitful interactions with behavior analysis.*

In this article we import a distinction from evolutionary biology—that between selectionism and essentialism—to discuss contrasting trends in cognitive science. Largely because of the prestige of Darwin's theory, essentialism is out of fashion as an explicit doctrine in science. However, one can pay lip service to selectionism and still subscribe to essentialist assumptions, employ essentialist locutions, define essentialist units of analysis, and worse, pursue research guided by these assumptions, units, and locutions. In contrast to most of his contemporaries, B. F. Skinner consistently repudiated essentialism (although he never used the term) both in his science and in his verbal behavior. Of particular significance, we argue, was Skinner's early methodological claim that the appropriate units of analysis in a science of behavior are to be defined empirically, rather than a priori (Skinner, 1935, 1938). By putting this claim into practice, Skinner set the stage for a thoroughgoing selectionist science and so avoided much of the fruitless inquiry engendered by implicit essentialist assumptions. The field of behavior analysis has generally, although not always, heeded Skinner's precepts and remains psychology's most consistently selectionist enterprise. In contrast, many contemporary cognitive scientists, while accepting selectionism at the phylogenetic level, explicitly reject Skinner's position without subscribing to an alternative selectionist methodology; consequently, much normative cognitive science, we argue, is prone to essentialist assumptions and locutions that have engendered research, which, from a selectionist's point of view, is uninterpretable. In our opinion, the distinction between behaviorism and cognitivism is less fundamental than the distinction between selectionism and essentialism. Behavior analysts are well advised to consider those cognitive analyses that are selectionist and question any behaviorist analyses that smack of essentialism. We close on the optimistic note that current trends in cognitive science, particularly the growing interest in adaptive networks, or parallel-distributed processing, are amenable to selectionist interpretations. This suggests the possibility of fruitful interactions between such models of cognition and radical behavioral accounts of complex human behavior.

## Essentialism and Selectionism Contrasted

Darwin's theory of evolution by natural selection is still hotly disputed in some quarters, but there are few in the scientific community that doubt that the theory, at least in its broad outlines, accounts for the extraordinary complexity and diversity of living things. The elegance of the theory lies in the unparalleled simplification that it achieved; countless acts of special creation were replaced by the repeated action, over the eons, of a relatively few, elementary processes. Although selective breeding had been practiced for millennia, the power of selection to explain adaptive complexity in nature was not cogently argued until Darwin unveiled *The Origin of Species* in 1859. Despite its elegance, Darwin's theory languished without widespread acclaim until its synthesis with population genetics in the 1930s. If selection was slow to be recognized for its role in phylogeny, it was slower still to be recognized as a nonteleological explanation for a wide variety of other complex phenomena, including operant conditioning (e.g., Thorndike, 1898), the immune response (Jerne, 1955), problem solving and the acquisition of knowledge (e.g., Campbell, 1960, 1974; Popper, 1972), cultural practices (e.g., Campbell, 1975; Skinner, 1948, 1971, 1981), perception (Campbell, 1956b), neural networks (e.g., Edelman, 1987), and technological innovations (e.g., Basalla, 1988; Root-Bernstein, 1989). Selectionist interpretations have been provided for phenomena as disparate as the locomotion of protozoa (Baldwin, 1895; Campbell, 1956a) and the orderly orbits of planetary bodies (Donahoe, Burgos, & Palmer, 1993). As in the case of speciation, even when underlying mechanisms are poorly understood, selectionist accounts are appealing because of their simplicity and power.

Selectionism is gradually replacing what the evolutionary biologist Ernst Mayr called *essentialist thinking*, the tendency to view categorical phenomena in nature as reflections of universal, enduring qualities intrinsic to each class or unit (Mayr, 1976, 1982, 1988).<sup>1</sup> In Darwin's day, the dominant view of living things was essentialist in this sense; species and other classes of organisms were seen as collections of individuals that all shared some essential property that defined the group, and taxonomy was largely a matter of identifying these essential properties. Individuals within a group might vary widely, but they were all seen as variants of a single

---

<sup>1</sup>Mayr (1976, 1982, 1988) borrowed the term *essentialism* from Popper (1957), who coined it as an unambiguous substitute for the overworked term *realism*. The realists were Scholastic philosophers who, quibbling over the exegesis of Aristotle, held that categories are defined by essential properties that transcend the specific members of the categories. The category of white things, for example, is defined as those elements that possess the property of "whiteness." Whiteness is an essence, a "thing" that we come to know in its own right through our experience with white objects. This position can be traced back in some form to Plato and Parmenides. In contrast, nominalists (chiefly following William of Occam) held that universal terms such as *white* were mere labels that we use for a collection of objects of a common color. They do not represent things with an existence or status independent of the set of white objects. Hobbes (1651/1968) captured the nominalist position in his discussion of the distinction between proper nouns and common nouns. Of the latter he noted, "Every one of which, though but one name, is nevertheless the name of diverse and particular things: in respect of all which together, it is called a *universal*, there being nothing in the world universal but names: for the things named are every one of them singular and universal, one universal name is imposed on many things for their similitude in some quality, or other accident" (p. 102). Locke (1690/1975), in his *Essay on Human Understanding*, foreshadowed much of the discussion in the present article by distinguishing between what he called "real essences" and "nominal essence," that is, between properties that are intrinsic and those that are conventional. He held that only humanly defined categories such as *circle* and *Whig* could be said to have real essences. Natural categories such as *man* were conventional: he noted that occasionally infants were born so deformed that it was unclear whether they qualified as exemplars of the category *human*. Although circles have defining properties that, in a sense, precede any instance of a circle, natural categories such *sheep*, *tree*, or *man* do not have defining qualities, except by convention.

template. Individual variability could be explained as the outcome of less fundamental factors—crossbreeding, environmental stress, accident, or other vicissitudes.

Of course this position leaves unexplained the origin of the templates. It is characteristic of essentialism that phenomena are said to reflect some ideal, some essence, or some template that itself remains unexplained. The devout presumably attribute the origin of templates to a deity. Certainly, much of the bitterness with which Darwin was attacked arose from the implication that natural phenomena could be explained without reference to a designer. Nevertheless, one does not have to be a deist to believe that species have essential properties; the skeptic merely adds the origin of templates to the list of natural phenomena that we accept as unexplained, perhaps even as unexplainable.

Darwin's (1859/1950) discussion of the term *species* is clearly at odds with the essentialist position. He noted,

No one definition has satisfied all naturalists; yet every naturalist knows vaguely what he means when he speaks of a species. Generally, the term includes the unknown element of a distant act of creation. . . . It is certain that many forms, considered by highly-competent judges to be varieties, resemble species so completely in character, that they have been thus ranked by other highly competent judges. But to discuss whether they ought to be called species or varieties, before any definition of these terms has been generally accepted, is vainly to beat the air. . . . It will be seen that I look at the term species as one arbitrarily given, for the sake of convenience, to a set of individuals closely resembling each other, and that it does not essentially differ from the term variety, which is given to less distinct and more fluctuating forms. The term variety, again, in comparison with mere individual differences, is also applied arbitrarily, for "convenience" sake. (pp. 24-29)

Of course it was not Darwin's linguistic or philosophical predilections that aroused his contemporaries, but his exposition of selection as a non-teleological explanation of the diversity of life. Evolution requires only that there be heritable variation among individuals and contingencies of selection that operate over time. Selection is a process that necessarily takes time; hence if contingencies of selection are stable over the duration of one's observations, species may appear to have essential properties, but Darwin showed that this appearance of stability is quite consistent with a selectionist account. When viewed over time, it is clear that the individual is a unique constellation of properties that can only be understood, not by considering one's group membership, but by considering, in detail, the environment-organism interactions of one's ancestors.

However, it means something to say, for example, that a fox is a fox and not a mouse. Presumably all foxes share a common ancestor, and as they breed with one another and not with dissimilar species, they have a strong family resemblance. In addition, all foxes are subject to similar contingencies of selection. Do not the contingencies of natural selection specify the essential properties of species? To argue so is to overlook a crucial feature of evolutionary phenomena. Foxes will vary from one another within bounds determined by the selection contingencies. Any variant that satisfies the contingencies can make a contribution to future generations of foxes. Moreover, although we speak of contingencies of natural selection in general terms, the contingencies are as individual as the organisms themselves. For example, we

might observe that a changing climate favored foxes with heavier coats or that brilliant plumage in a species of bird was more effective in attracting mates. However, survival is a matter not of global contingencies but of the moment-to-moment contingencies of an individual's life. A light-coated fox might thrive in a region where food was abundant, and a modestly plumed bird might find a mate if competition were scarce. Although, on the average, it may be true that the race is to the swift and the battle to the strong, the average organism is an abstraction; only individuals exist, and time and chance happeneth to them all.

The role of genes—unknown to Darwin—does not affect our conclusion. Genes are templates, to be sure, but constellations of genes are variable within a species, even from parent to offspring, and they are continually subject to mutation. Indeed, without this variability, evolution would be impossible. Selection does not produce organisms stamped out of a common mold. To the contrary, a bizarre mutation that eliminated variability in a population would surely prove fatal when prevailing contingencies changed.

Thus, contingencies of selection do not yield rigid, static, or idealized species, nor do they select rigid, static, or idealized properties of species. The selected property, be it a morphological feature or a behavior, can vary in any arbitrary characteristic that is incidental to the contingency, but, more fundamentally, it can even vary along the dimensions that are defined by the contingency. A selection contingency merely sets minimum standards for a property; it does not provide a blueprint. Variation within the boundaries of the selection contingencies will be constrained only by those mechanisms that generate variability in the property. The critical difference between essentialism and selectionism, then, is that selectionism regards variability within classes of phenomena as fundamental, whereas essentialism regards it as a misleading irrelevance.

### **Selection and Behavior**

The analogy between natural selection and learning has struck several observers, apparently independently (e.g., Baldwin, 1895, 1909/1980; Campbell, 1956a; Pringle, 1951; Skinner, 1953; Staddon, 1983). Thorndike (1898) provided the first systematic analysis of reinforcement: A cat in a puzzle box emits a wide variety of behaviors; some variants affect the apparatus in such a way that the door falls open, allowing the cat to escape; on later trials these successful variants are more likely to occur. As in evolution, variable elements have different consequences; those with certain kinds of consequences are strengthened relative to the unsuccessful elements. Order emerges without appeal to a designer or to intentionality. Skinner (1938) noted that highly organized, complex behavior could be shaped from relatively undifferentiated baseline behavior by successive contingencies of reinforcement in which the selection criterion was gradually altered to more closely approximate some target behavior, just as gradually changing climate, habitat, competing species, and so on, presumably shaped highly organized, complex structures or innate behaviors in organisms (Skinner, 1966b). Thus, both evolution and reinforcement operate by repeatedly selecting elements from a variable substrate to produce orderly classes of phenomena. In behavior as in morphology, variability is required if selection is to create new forms.

The behavioral repertoire of an organism is affected by both phylogenetic contingencies and contingencies of reinforcement. Consequently, variability is a fundamental characteristic of behavior, even under the most-restricted circumstances. Abstraction, idealization, categorization, and averaging cannot eliminate this variability; it can only mask it. We err in our science if we

treat the variability of our subject matter as an annoying irrelevance that can be eradicated by such practices.

### **Implications for Units of Analysis in Cognitive Science**

In the behavioral sciences no one has been a more thoroughgoing selectionist than Skinner, for he has interpreted all behavior—from lever presses to perception, verbal behavior, and thinking—in terms of principles of selection. But Skinner's place in the pantheon of selectionists does not rest primarily on his verbal interpretations, important though they are in showing the scope of these principles. Rather, his main contribution was the methodology that arose from his prescient grasp of the nature of his subject matter. Skinner recognized that variability was fundamental to behavior and fashioned his methodology accordingly. In particular, he realized that the appropriate units of analysis in a science of behavior are generic in nature. In this regard, so far as we know, he was unique.

#### ***Skinner's Empirical Units of Analysis***

As Skinner (1935, 1938) observed, if we want our units of analysis to respect lines of fracture in nature, we must define them empirically. We cannot define them from our armchairs; rather, we must survey the variability of our subject matter and adopt working definitions according to the order we find. Skinner was not making a philosophical claim; he observed that the order that emerged in his investigations rested upon adopting empirically determined, generic units of analysis.

According to Skinner, the search for orderly units of behavior and environment begins arbitrarily. The experimenter has a hunch, follows a precedent, or picks defining properties of his units at random. If the dynamic properties of behavior are not orderly when the units are so defined, the experimenter systematically restricts or modifies the definitions until orderly relationships between variables emerge. For example, in the study of the flexion reflex in a spinal preparation, Skinner (1938) wrote,

If we are measuring fatigue, for example, we shall not obtain too smooth a curve if our stimulus varies in such a way as to produce at one time one direction of flexion and at another time another; but as we restrict the stimulus to obtain a less variable response, the smoothness of the curve increases. (p. 36)

The process of modifying these definitions can continue to the point at which both environmental and behavioral units are completely restricted. Here we choose to count only those stimulus and response events that meet very narrow definitions. For example, we might define the response in terms of specific effectors, precise location, force, latency, and so on. However, Skinner noted that nothing is gained by continuing to restrict these definitions past a certain point:

The generic nature of the concepts of stimulus and response is demonstrated by the fact that complete induction obtains (and the dynamic changes therefore reach an optimal uniformity) *before* all the properties of stimulus and response have been fully specified in the description and respected in each elicitation. (p. 37)

Extending the analysis to operant behavior, Skinner noted that, in the process of restricting our definitions, an inflection point is reached at which our data are most orderly. Continued restriction actually leads to a deterioration in the orderliness of the data. Consider bar pressing in

the rat:" Before we can see precisely what a given act consists of, we must examine the changes it undergoes in strength. Here again we merely specify what is to be counted as a response and refuse to accept instances not coming up to the specification. A specification is successful if the entity which it describes gives smooth curves for the dynamic laws. . . . The number of distinguishable responses on the part of the rat that will give the required movement of the lever is indefinite and very large. They constitute a class which is sufficiently well defined by the phrase 'pressing the lever.' . . . The members of the class are quantitatively mutually replaceable in spite of their differences. If only such responses as had been made in a very special way were counted (that is, if the response had been restricted through further specification), the smoothness of the resulting curves would have been *decreased*. The curves would have been destroyed through the elimination of many responses that contributed to them

. . . *A respondent, then, regarded as a correlation of a stimulus and a response and an operant regarded as a functional part of behavior are defined at levels of specification marked by the orderliness of dynamic changes.* (pp. 37-40)

Note that Skinner was not recommending that we simply manipulate our independent variables until order emerges; he was recommending that we modify our *definitions* until order emerges. That is, a single set of observations might yield either orderly or disorderly relationships, depending on what we choose to count as stimulus and response events.

Skinner explicitly rejected the practice of ad hoc categorizing of stimulus and response units. When a boy hides from a dog, it is a mistake, he averred, to assume uncritically that the dog is a stimulus or that hiding is a response. Such practices may be useful in interpreting behavior outside the laboratory but should be avoided in the basic science. Skinner also rejected Watson's (1930) definition of a response as "anything the animal does, such as turning toward or away from a light, jumping at a sound, and more highly organized activities such as building a skyscraper, drawing plans, having babies, writing books and the like" (cited in Skinner, 1938, p. 42). Clearly, the latter activities do not share the orderly properties of bar pressing or turning toward a light.

Skinner justified his position on pragmatic grounds, declining to speculate whether methodological advances will permit fruitful analyses of fully restricted units. However, he might have justified it on principled grounds as well: Generic units of analysis follow from a commitment to selectionism. As noted earlier, contingencies of selection cannot yield idealized units; variability is fundamental. Selection contingencies merely set bounds on what is possible; they do not prescribe designs.

The effectors that enable a rat to press a bar were presumably selected for their contribution to foraging, climbing, running, and so on. A particular act, say, extension of the forepaw, contributes to each of these activities, and the structures participating in the act can be explained in part by reference to contingencies of natural selection. Not only do these contingencies tolerate variability in form from one instance to the next, but such variability is actually necessary if reinforcement contingencies are to shape effective locomotion in an uneven environment. Variability of response topography is not only unavoidable, it is adaptive. Skinner's generic units, then, are analogous to species: The units are orderly but are neither arbitrary nor invariant.

Skinner's analysis of behavioral units establishes his science squarely within a selectionist

paradigm. Moreover, without his methodological precepts there could be no justified inductions in the analysis of behavior and hence no science of behavior; nor could there be plausible interpretations of complex phenomena based on that science. Despite the importance of Skinner's analysis, his views on the subject are virtually unknown outside the field of behavior analysis. Actual practice in normative cognitive science is more consistent with essentialism. Units of analysis are usually defined a priori, formally, or are taken as self-evident.

### *A Priori Units of Analysis*

Some concepts have formal definitions, presumably proposed by one individual and adopted, without complaint, by others. Centuries ago, someone defined a dodecahedron as a regular polyhedron with 12 faces, and the definition has served us well ever since. Deciding whether something is a member of the class is merely a matter of inspecting the candidate in the light of the definition. If asked whether a particular garnet crystal is a dodecahedron, we can examine it and assert, with complacency, that no, although it has 12 faces, it is not *regular* at the third decimal place. Some may balk at such pedantry—surely the concept of dodecahedron was defined in order to capture the regularity of garnets and other crystals—but there is no requirement that a concept with a formal definition map onto the dusty and spotted stuff of the real world.

To the contrary, concepts with formal definitions *do not* map onto the natural kinds of the real world. We treat some phenomena categorically before we try to define them—garnet crystals, for example. In such a case we may try to define the class, but our attempt requires not *imposing* a definition but *discovering* one. We must study crystals, and we may, at length, settle on a definition that seems to embrace an appropriate group. Nevertheless, on further study we are sure to find a crystal that seems as though it *ought* to qualify as a garnet but that just fails to meet our definition. The crystal has some impurity or irregularity that we had not anticipated. (Defining the class in terms of its chemical structure, for example, would include only pure garnets—if there are any—or pure parts of impure crystals, thus missing the generalization that we set out to capture.) We may modify our definition, or we may just decide that, although imperfect, it is good enough. Our empirical definition will have some slack in it, uncharacteristic of a priori formal definitions. Thus, the problem for the scientist is to determine the natural lines of fracture of the phenomena under study.

The task of determining these lines of fracture is the task of finding order in our subject matter. Order is partly, of course, in the eye of the beholder, thanks to a particular genetic makeup, a particular personal history, or a particular practical contingency—all of which involve contingencies of selection. Although garnet crystals may seem an obvious and well-established category to a mineralogist, they surely do not to a sea slug or to many an unschooled human, for that matter. A chicken will see little to choose between a fox and a coyote but will make much of the difference between a worm and a caterpillar. Honeybees will swarm to buckwheat in the morning but forage elsewhere in the afternoon, bewildering the farmer, who sees no difference. On the other hand, regularities in the world presumably do not wait upon an observer; the apple would still have fallen, even if Newton had overslept. But how these regularities affect the observer is a matter of contingencies of selection.

Categories with formal definitions or categories defined a priori can be said to have essential properties, properties that, in a sense, precede any example of the category. We are free to define

such categories as we please, and they need not reflect distinctions in nature. If we choose to define human beings as featherless bipeds, then we cannot object to including *Tyrannosaurus rex*. As it is considered good scientific practice to define one's terms, researchers may be tempted to define their terms before the data are in, and, in effect, provide a formal definition for a concept that is more appropriately defined empirically. Doing so may be harmless, but it invites essentialist thinking into a domain shaped, at least in part, by contingencies of selection. The problem is not unique to cognitive science. The evolutionary biologist Benjamin Burma (1949) voiced the following lament about the species concept:

In some respects it is extremely unfortunate that names ever get attached to ideas or objects. The false attachment of names to ideas or objects similar but not identical with the original can work harm far exceeding the benefits conferred by having a convenient label. The name "species" has come to such a state. (p. 369)

In biology and psychology, orderly classes of phenomena will have all the variability characteristic of products of selection. Such categories, then, will be fluid both because the exemplars of a category will have varying properties and because the boundaries of the category will depend upon the demands and characteristics of the observer. But it is a rare cognitive scientist who defines his terms empirically. In its flight from the restrictions of behaviorism, cognitive science has abandoned this important methodological constraint. Consequently many cognitive concepts retain a strong essentialist flavor.

### ***Some Difficulties With Essentialist Units of Analysis***

It is our thesis that essentialist thinking is no more appropriate in behavioral or cognitive science than it is in evolutionary science. Selection is the only natural process yet proposed for explaining adaptive complexity in nature (Dawkins, 1986), and we argue that this is as true for behavior as it is for morphology. (Note that this is not to claim that *all* structures or behaviors have been specifically selected. As we have argued, selection is a blunt tool that permits, and therefore ensures, variability.) An analysis of a behavioral phenomenon in essentialist terms may well be better than nothing, but it can be more profitably recast in selectionist terms. Moreover, essentialist analyses are prone to the following problems.

*Temptation to circular reasoning.* In the worst case, essentialism attempts to explain a behavioral phenomenon by inventing a property of the organism responsible for the phenomenon. Because behavioral phenomena are presumably, at least in part, a function of properties of the organism, this seems an innocent step. However, the hypothetical property often is later invoked as an explanation for the phenomenon; it transubstantiates from a tautological construct to an essence with causal status (cf. Skinner, 1963). Of course, this is circular, and it is rare scientist who deliberately engages in circular reasoning. However, it is often easier to ridicule circular reasoning than to avoid it. William James (1907) noted that although no one seriously thinks that wealth is an explanation for having money, it is not uncommon to attribute sickness to poor health, or muscular feats to great strength, or problem-solving skills to great intelligence. In cognitive science it is not uncommon to attribute language acquisition to linguistic competence, or ineffective performance on a recall task to limited capacity. Patently circular usage is typically avoided by the original proponents of a concept. Circularity often emerges, however, when a concept becomes familiar; it tends to become reified, especially by subsequent researchers, students, and writers of secondary texts. (Indeed, Medin & Ortony,

1989, have suggested that such reification may be fundamental to cognition.)

*Curtailed inquiry.* A second shortcoming of essentialist explanations of phenomena is that they tend to cut off inquiry prematurely (Skinner, 1950). Essential properties are treated as givens; they need no further explanation. Note that attributing a behavioral phenomenon to an innate structure of the organism is not essentialistic for innate structures are presumably selected by evolutionary processes. However, the structure will not be ideal, universal, or fixed, but will reveal the variability characteristic of all physical features that are the products of selection. As noted, natural selection permits variability and, indeed, may select for variability. Far from being ideal, structures are often opportunistically cobbled together from the available "raw materials" of ancestral species (Gould, 1980). Investigating the structure, its variability, and its evolutionary origins enriches our understanding of the behavioral phenomenon.

However, the widespread practice of gratuitously invoking the genetic endowment is essentialism in selectionist's garb. To attribute a behavioral phenomenon to an invented property of the organism and then to "explain" the property by casually alluding to the genetic endowment is vacuous unless the origin of the property can plausibly be identified and discussed in selectionist terms. The short-term memory register, for example, is a metaphor, not a structure. As a metaphor it may be useful, but it has no explanatory force, for it is accorded just those properties necessary to explain the disparate phenomena grouped under the heading *short-term memory* (cf. Crowder, 1982). To suggest that it is innate is a statement of faith, not a serious proposal. At best, the suggestion that it is innate is the beginning of inquiry, not its end, for we now have assumed responsibility for accounting for the natural selection of short-term memory. It is most often treated, not as a variable outcome of selection, but as a fixed property of the organism. No attempt is made to treat it as a product of contingencies of selection. .

*Unparsimonious explanations.* A third drawback of essentialist concepts, and the last that we shall mention here, is that they all require separate explanations. Darwin's theory achieved a dramatic simplification, as many apparent acts of creation were shown to be the product of a few common processes, and seemingly unrelated phenomena were shown to be intimately related to one another. Unconstrained by selectionist thinking, the structures postulated within cognitive science may proliferate in an unconstrained manner. For example, when the data obtained in memory experiments are not readily accommodated by the distinction between short-term and long-term memory, an intermediate memory may be proposed. Similarly, when not all data obtained with a procedure used to study long-term memory can be accommodated by a single set of processes or structures, different subtypes of long-term memory—semantic versus episodic or declarative versus nondeclarative—may be proposed. These proposed subdivisions of memory may begin as convenient shorthand descriptions of observed differences but become reified into "types" of memory whose characteristics are just those required to encompass the data that were the impetus for their postulation.

Often, such distinctions lead to pointless debate over whether a given phenomenon is really evidence for a given concept or whether it is better taken as evidence of some other concept. For the same reasons it makes no sense to ask whether *Archaeopteryx* is really a bird or a reptile, there is little to be gained by asking whether a given datum in a memory experiment is really from long-term rather than short-term memory or from episodic rather than semantic memory. Observations of intermediate forms and intermediate memories may both be of great value because of what they reveal about the course of selection, but not because of what they purport to

reveal about membership in one or another essentialist "type." Problems with lack of parsimony, as with circularity and curtailment of inquiry, are not necessary consequences of assigning names to classes of observations, but they are often the essentialist accompaniments of such a practice.

### **Cognitive Science and Essentialism**

Researchers in cognitive science are seldom explicitly essentialistic, of course. However, most cognitive analyses have, until recently, shown a taste for formal or a priori units of analysis and have posited complex, high-level structures as explanations of cognitive phenomena, without reference to either physical mechanisms explained by contingencies of natural selection or acquired repertoires explained by contingencies of reinforcement. Consequently, these analyses are implicitly essentialistic, and subsequently, they come to be treated so by other researchers, students, and writers. We will outline several examples.

#### ***Verbal Behavior***

Much of contemporary linguistics, particularly that associated with Noam Chomsky, is an effort to employ essentialist analyses to shed light on behavioral phenomena. Here the essentialism is not an inadvertent outcome of the careless use of terms but is explicit policy adopted in order to isolate crucial variables. The goal is not merely to provide an abstract structural description of language but to use that abstract description to tell us something about human beings, how we speak and understand our language, and how we learn to do so in a few short years. Chomsky (1980a) suggested that an inquiry into language be modeled on inquiry into biological systems, such as the visual system, and that it consider its function, abstract structure, physical structure, individual development, and evolutionary development (p. 227). Although a complete analysis of language will include many idiosyncratic phenomena, the most interesting questions, in his view, concern universal properties of language, particularly those that are genetic in origin, for they are "part of an innate endowment that defines the 'human essence' " (p. 92). Chomsky pointed out that "the study of biologically necessary properties of language is a part of natural science: its purpose is to determine one aspect of human genetics, namely, the nature of the language faculty" (p. 29).

These biological questions would appeal to any selectionist, but the terms *abstract structures*, *language faculty*, and *human essence* all introduce a flavor of essentialism. When we look at the realization of Chomsky's program, however, there is no longer any doubt. Citing methodological and ethical problems in pursuing the other issues mentioned earlier, Chomsky (1965) invested most of his effort into characterizing the abstract structure of language. He quickly removed to an idealized world in which everything has essential properties. "Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly" (p. 3). "I have in mind certain biological properties that are genetically-determined and characteristic of the human species, which I assume for the purposes of this discussion to be genetically uniform, a further idealization" (Chomsky, 1980a, p. 28). "I have argued that the grammar represented in the mind is a 'real object,' indeed that a person's language should be defined in terms of this grammar" (p. 120). A highly abstract formal system is proposed. The unit of analysis is assumed a priori: "The basic elements we consider are sentences; the grammar generates mental representations of their form and meaning" (p. 143). Thus, the Sentence is defined formally, not operationally or empirically. In fact, few, if any of the terms of the analysis can be given operational definitions: Chomsky

asserted that "'wellformedness' is a theoretical concept for which we cannot expect to find a precise set of necessary and sufficient operational criteria" (p. 198). "Furthermore, there is no reason to expect that reliable operational criteria for the deeper and more important theoretical notions of linguistics (such as 'grammaticalness' and 'paraphrase') will ever be forthcoming" (1965, p. 19). Although linguistic intuitions of native speakers play a role in the enterprise, they have limited authority: "In fact, as soon as sentences become moderately complex, judgments begin to waver and often fail, which is neither surprising nor particularly important, but leaves the question of existence of a decision procedure wide open" (1980a, p. 121).

Chomsky's approach is avowedly essentialistic; indeed, he regards an abstract characterization of grammar as the most promising line of inquiry, given the constraints on research with humans. He clearly expects that the analysis will provide insights into the imperfect and variable domain of living organisms where selectionist principles hold sway, at least in part. Is there any objection, in principle, to this approach? We do not argue so, although we regard it as an open question. The Linnaean classification of species based on structural features was essentialistic, and although it is not identical with a system based on evolutionary considerations, it proved useful in organizing a bewildering variety of data. However, the two classification systems are qualitatively different; the Linnaean system did not contain the seeds of the evolutionary one. Selectionist and essentialist units of analysis are incommensurate; the terms in a formal analysis cannot be simply translated into those of an empirical analysis. When our units of analysis are defined formally, as in a grammatical analysis, there is no guarantee that they will have the orderly properties of units defined empirically. Moreover, if Chomsky was correct that many grammatical concepts cannot be given operational definitions, the translation becomes even more formidable, if not impossible. We believe that this point is insufficiently appreciated by Chomsky and his followers.

We question, for example, that the sentence is an appropriate unit of analysis. In a formal system, sentences are defined by a grammar, and grammars, at least those proposed for human languages, can generate an infinite number of sentences. Given a finite vocabulary, it is obvious that only an infinitesimal fraction of these infinite sentences would have fewer than, say, a billion words. This imperfect fit between the sentences generated by the formal system and actual units of behavior is dismissed by Chomsky as simply reflecting the difference between competence and performance. Similarly, performance variables are invoked to explain the fact that people often do not speak in sentences and the fact that grammatical intuitions are notoriously fickle. It seems to us that the competence-performance distinction is being used to mask the fundamental variability of behavior in order to shore up the formal edifice. The sentence is the preferred unit because it permits a formal analysis, not because it is an orderly unit of behavior.

Second, we question the device of characterizing language abstractly and then imputing the formal edifice to the organism as a part of its innate endowment. It is taken for granted by Chomsky and his colleagues that people use a grammar when speaking and listening. Linguists have found it to be a difficult matter to provide a fully adequate grammar for natural languages; no doubt there seems to be no alternative to assuming that people actually use such a system. However, there is no force to this conclusion. A child can learn to catch a baseball presumably because of the orderliness of the laws of motion, but we do not assume that the child has induced these laws. The possibility cannot be rejected, but until the account is fleshed out in biological terms, imputing formal laws to the child as an explanation of his behavior is circular.

Chomsky attempted to embed his formal system within the natural sciences by arguing that he was providing an abstract characterization of properties of the genome. His formal analysis of language reveals many regularities that can be described by abstract rules. No one teaches these rules to children, and no one consciously uses them when speaking or writing. Some rules facilitate communication, and one might argue that they are learned as a consequence of communicative contingencies. Many of the rules, however, are wholly arbitrary; utterances would be reasonably intelligible if the rule were violated. Chomsky (1975) pointed out that, for some rules, it is unlikely that many people have had experience that even bears on the rule; indeed, "it is often a difficult problem even to discover examples that bear on the hypothesis in question" (p. 175). In other cases, induction from personal experience would suggest violations of the rule, violations that are not, in fact, observed. Chomsky (1980a) argued,

The knowledge acquired and to a large extent shared involves judgments of extraordinary delicacy and detail. The argument from poverty of the stimulus leaves us no reasonable alternative but to suppose that these properties are somehow determined in universal grammar, as part of the genotype. There is simply no evidence available to the language learner to fix them, in many crucial cases that have been studied. (p. 66)

This argument has a superficial cogency. In our present understanding of biology, if a characteristic cannot have been acquired, we assume that it is genetic in origin. However, as has been noted, learning principles and evolutionary principles are parallel selectionist principles. Both account for complex phenomena by appealing to variation, selection, and retention. The same evidence that is adduced to argue that a rule cannot have been learned can be used to argue that it cannot have evolved through natural selection. Arbitrary rules that have no communicative or practical consequences cannot confer a selective advantage to an individual who uses such rules. If it is hard even to find examples that bear on the issues, it is unlikely that the grammatical distinctions will contribute to evolutionary fitness. Acknowledging this difficulty, Chomsky (1980b) argued that although the child must learn his language in a few years, evolution has had many thousands of years to select universal grammar. However, the length of time that selection is at work is irrelevant if the grammatical distinctions are arbitrary. The "argument from poverty of the stimulus" applies to any selectionist account, be it ontogenetic or phylogenetic. Attributing a complex phenomenon to the genetic endowment without requiring that the evolutionary account be plausible, or even possible, is comparable to arguing for the special creation of species.

The conclusions that Chomsky (1969) drew from his formal analyses appear inconsistent with principles of selection. Rather than question the analyses, Chomsky preferred to question selectionism:

It is, in fact, perfectly possible that the innate structure of mind is determined by principles of organization, by physical conditions, even by physical laws that are now quite unknown, and that such notions as "random mutation" and "natural selection" are as much a cover for ignorance as the somewhat analogous notions of "trial and error," "conditioning," "reinforcement," and "association." (p. 262)

In this appeal to additional but unknown principles, even to "physical laws. . . now quite unknown," Chomsky (1969, p. 262) explicitly cast himself loose from the selectionist principles

that unify the biological sciences and that provide an explanation of diversity and complexity at both the ontogenetic and phylogenetic levels. The retreat to an ideal, essentialist world is no longer a temporary recourse but an end in itself.

Not all linguists are equally dismissive of the need to provide a selectionist interpretation of linguistic analyses. Pinker and Bloom (1990) noted, "If a current theory of language is truly incompatible with the neo-Darwinian theory of evolution, one could hardly blame someone for concluding that it is not the theory of evolution that must be abandoned but the theory of language" (p. 708). They then provided a masterful interpretation of the adaptiveness of human language. Their account succeeds admirably in relating the facts of language to selectionist theory but it falls short of showing that an essentialist interpretation of these facts is justified. As Herrnstein (1990) asked,

What evolutionary pressure selects for the case filter or structure dependence or the binding theory or X' theory? It seems at first blush that a perfectly serviceable communication system that did not mark "abstract" case on [noun phrases] could be just as good a medium of communication as one that does. (pp. 735-736)

The force of Pinker and Bloom's argument rests on the lack of a precise alternative account for the grammatical regularities they discussed. However, we argue next that formal models in linguistics do not themselves provide explanations of these regularities.

Chomsky holds that humans are endowed with an innate device that will generate a grammar appropriate to their speech community, based on a limited set of exemplars. Even with impoverished input, a child will master grammatical distinctions and so avoid errors that might be suggested by induction from the exemplars. This appears to be a reasonable proposal until we ask how such a device would work. There are many examples of behavior that are evidently largely genetic in origin (e.g., courtship rituals, maternal behavior, warning cries, feeding). In each case it is possible to identify the properties of the class of stimuli that elicits the behavior. Often it is possible to "trick" the organism by exposing it to supernormal stimuli or inappropriate stimuli within the class, as when a duckling is imprinted on a human or a hawk. In the case of Chomsky's putative generator of grammars, however, there is no way to characterize the input to the device; the input is entirely arbitrary. A child in France will be exposed to French, the child in Cuba will be exposed to Spanish, and the deaf child will be exposed to signing. An innate mechanism can make no use of such arbitrary inputs. In order for such a device to be possible, the inputs to the device must already be segmented and analyzed into phrases, words, and parts of speech. Here we are faced with the familiar problem that such terms are essentialistic; they have meaning only in the world of formal analysis, not in the world of fundamentally variable classes of environmental and behavioral events. The grammatical classes of the formal analysis cannot be defined in terms of such events; neither, according to Chomsky, can they be operationally defined. Rather than solving the child's problems, Chomsky's proposal seems to raise new ones that have no solution.

This objection is often countered by the suggestion that the child acquires knowledge of grammatical classes by a kind of "bootstrapping" operation (e.g. Grimshaw. 1981; Pinker, 1982. 1984). Nouns are often uttered in the presence of "things" and verbs in the presence of "actions." Through interacting with the world and with proficient speakers, the child learns such distinctions and refines them with further experience. Once armed with a modest repertoire of word-concept pairs, the child sets to work analyzing patterns in the sentences he or she hears.

Through this sort of distributional analysis the child can generalize to abstract nouns and verbs and can identify determiners, suffixes, and so on.

Such proposals do not specify how this bootstrapping operation actually works, for the formal conceptual apparatus of linguistics is not equipped to analyze environmental and behavioral classes of events. What, specifically, are the stimulus properties of a "thing"? When is a stimulus a thing, as opposed to an attribute? What events entail reference, as opposed to coincidence? It is not sufficient to allude to the child's knowledge of such things, for that merely rephrases the mystery; we need an analysis of what happens at the level of stimulus input to the organism without taking refuge in the child's "knowledge" or "concepts." Finally, exactly how does the nonverbal child conduct the elaborate hypothesis testing necessary to conduct a distributional analysis of language? As Pinker (1984) himself observed, "If the theory is correct, then I have placed a considerable burden on theorists of conceptual and communicative development, saddling them with an inventory of cognitive abilities far richer than many current cognitive theories allow for" (p. 362). To us, the proposed solution seems as formidable a problem as the original mystery.

We do not doubt that a child learns his or her language through the kinds of interactions implied by the semantic bootstrapping hypothesis, but we question that he or she will learn the essentialist concepts of Chomskyan linguistics. What a child learns from such interactions are not grammatical terms but, curiously enough, the behavioral units analyzed by Skinner (1957) in *Verbal Behavior*. Skinner's account avoids these problems because his generic units are perfectly suited to the analysis of inherently variable environment-behavior interactions. If Chomsky's proposals are to find a place in contemporary biological thought, they must eventually be translated into selectionist terms; when they are, it will be found that Skinner got there first.

We have argued that Chomskian linguistics, one of the most important themes in contemporary cognitive science, is heavily essentialistic and, for just that reason, promises little help in integrating cognitive science with the rest of biology (cf. Andresen, 1990; Palmer, 1986). Chomsky (1959) has explicitly rejected an account of language in terms of reinforcement contingencies, and as we have indicated, the mechanisms he proposed as an alternative seem to be immune to contingencies of natural selection. Moreover, Chomsky has not shown how to translate his formal apparatus into the biological world with its fuzzy boundaries and permissive contingencies of selection. His proposed innate mechanisms, in solving one problem for the organism, pose new problems that to us seem insuperable.

### ***Remembering***

Other examples of essentialist thinking in cognitive science can be found in conventional views of memory. Memories are seen as *things*, things that are stored, things that are retrieved, and things that can explain subsequent behavior (cf. Watkins, 1990). We learn on Tuesday that a man's name is Johnson; on Wednesday we hail him by name. Within cognitive science, it is conventional to say that the man's name was stored as a memory and was retrieved the following day. The storage metaphor—a metaphor that is accepted uncritically by nearly everyone in cognitive science—supports this view of memory; *to store* is a transitive verb, and that which gets stored must be a thing. On different occasions we hail Johnson by name, we picture his face, we write him a note. The memory is not seen as the utterance, the image, or the scrawl; rather, the memory is seen as a unitary thing underlying the various behavioral manifestations, just as,

to an essentialist, the species is a prototype of which individuals are imperfect manifestations.

Few scientists take the storage metaphor literally; the brain is a memory storehouse only in a figurative sense. Certainly the nervous system changes with experience, and these changes may endure, but it is not these changes that we speak of when we speak of memories. We do not retrieve synaptic changes when we "retrieve" a memory. We suspect that the more the physiological bases of learning are understood, the less compelling the storage metaphor will be. We are not tempted to say that the patellar reflex is stored in the spinal cord. A reflex is a relationship between a class of stimuli and a class of responses; apart from this relationship it does not exist. When we identify a neural pathway, a "reflex arc," we have not isolated a reflex; we have just shown how it works. Similarly, memories have no existence apart from a relationship between those conditions that evoke them and the various behavioral outcomes from which memories are inferred. The underlying physiology is a pathway (or a network of pathways), not an essence.

Even though few take the storage metaphor literally, it appears to guide much theorizing in the field. Semantic network models and structural models of every sort subscribe to it, and the impact of these models has been tremendous. Difficulties with the Atkinson-Shiffrin model may be apparent to memory theorists, but the concepts of the model pervade the thinking of cognitive scientists and laymen alike. For a selectionist, however, questions about memory are most generally conceived as problems in stimulus control: What are the conditions under which present behavior is affected by selection by reinforcement of behavior in prior environments? In contrast, within most of cognitive science, remembering is approached as a set of problems in storage, search, and retrieval. From the selectionist perspective, an instance of recall is an instance of *current* behavior (some of which is possibly covert) and should be investigated just as any other instance of behavior. The antecedents of a memory are often much more complex and difficult to observe than the antecedents of an eye blink, but a difference in complexity is not a difference in kind. In a recall task we search not for memory traces but for controlling stimuli (often response produced). The resulting behavior is a unique event; although it may be, and usually is, a member of a response class, a response does not represent or stand for other responses. All behavior is equally a product of *current* controlling variables. From a selectionist's point of view, just as a given morphological feature can have different evolutionary histories in different organisms, a response of a given topography can have many antecedents, even in a single organism. In a recall task, a response of a particular topography is said to be "correct," regardless of the antecedents of the response. The storage metaphor is misleading in that it suggests that the response is not the product of a unique confluence of current events but of an invariant, hidden, surrogate of behavior.

A selectionist view draws our attention to the role of current variables and asks how they can account for memory phenomena. It is evident that some instances of recall are the direct outcome of past selections. When we fumble for the windshield-wiper switch of an unfamiliar car, we twist knobs and pull levers until the wipers go on. On the next occasion, we may reach for the correct switch immediately. Similarly, we may respond "1588" when asked "When was the Spanish Armada?" because that response to the question was reinforced in grammar school. However, many memory phenomena require a more elaborate interpretation. When asked, "Where were you last Thursday?" we do not reply with a stock answer that was reinforced in the past. When asked such a question we engage in recall strategies that supplement current controlling variables with those that are generated by our own responses: "Let's see. . . today's

Monday. . . On Wednesdays I have Historical Society meetings. . . . I stayed out late Wednesday night. . . . I must have stayed in bed late on Thursday," and so on. The question does not directly evoke the response; rather, it initiates an acquired procedure of generating responses that, together with the initial question, may not be adequate to control the "answer." Thus, whether or not a response is directly controlled by a stimulus, we attempt to analyze memory phenomena as current behavior under control of current variables, using units of analysis consistent with Skinner's formulation. (See Donahoe & Palmer, in press; Palmer, 1991, for a more extensive discussion of these points.)

We are not suggesting that all findings that are obtained in research guided by the storage metaphor are without merit. Researchers, regardless of their theoretical commitments, are constrained to investigate manipulable independent variables, not memory traces, and much of the resulting body of data is important, regardless of one's perspective. However, we believe that most theories of memory, *to the extent that they are essentialistic*, are misleading and contribute little to the value of research in the field.

### ***Representation***

A third manifestation of essentialism within cognitive science, one that seems to reach into all corners of the field, is that of representation. A representation is more than a copy; it is a symbol, capable of being stored, manipulated, and transformed. The thing being represented—usually some stimulus, event, or action—metamorphoses from a unique environment-organism relationship into an abstraction, an ideal, and in so doing crosses the boundary into the essentialist domain of formal analysis. The effect of this transubstantiation from an event in the physical world to an abstraction in an essentialist world of representation may sometimes be harmless, but it is all too easy to forget that one has made an idealization and, then, to ask questions that make sense only in the world of symbols, not in the world of stimulus and response classes. We are particularly apt to do so when discussing meaning and reference. Nothing is more natural than to ask the meaning of a word or sentence, as if meaning were a property of an utterance, independent of an individual speaker or listener.

When spoken, a verbal expression is a response, to be understood as a product of the current situation, the organism, and a particular selection history. We may be reading the expression from a book, solving an anagram, repeating what someone else has said, alerting our listener to some state of affairs, recalling an anecdote, reciting a famous quotation. We provide a meaning of the utterance by considering such variables, along with the past experience of the speaker; the topography of the response need not be considered. For the listener, bringing a different history to bear, the meaning is different. For him, the utterance is not a response but a stimulus, and meaning is to be sought partly in his experience with elements of the class exemplified by that stimulus.

Perhaps it is just because utterances of a particular topography can have so many diverse origins that we strive to attach meaning to the elements themselves, something that will transcend the idiosyncratic instance. There is a typical, or modal, circumstance in which a typical speaker might emit such an utterance, and meaning is distilled from these considerations. Superficially, this view appears *to* simplify matters considerably. Meaning, as a property of a word (either as stimulus or response) becomes common to speaker and listener, in fact, to the entire verbal community. This notion of meaning is useful in everyday affairs; we ask "What

does that mean?" when confused, and we would be quite lost without dictionaries. Verbal behavior would appear to be pointless unless there were something common to speaker and listener, and the term *meaning* seems to capture that common element.

However, when we analyze the behavior of a single speaker or listener in terms of contingencies of selection the term *meaning* does not appear, just as the term *species* does not appear in an analysis of a morphological feature in an individual organism. The term *meaning* dissolves into myriad experiences of the individual just as a morphological feature is the product of countless unique instances of variation and selection (cf. Skinner, 1957. pp. 7-10). The simplification achieved by treating meaning as a property of a stimulus may be useful in casual discourse, but it can cost us dearly if we take it seriously, that is, if we use an essentialist concept such as meaning to guide our theories about the products of selection (i.e., about organisms). Yet, the notion that words and phrases have meaning is central to semantic network models, models with lexicons, and computerized translation algorithms, and it is suffused throughout most analyses of language and intelligence.

Research guided by essentialist views of meaning and reference typically presents subjects with contrived strings of words, usually written, and tests recall, recognition, reaction time to answer questions, and so on, in an attempt to determine the components of concepts, the constituent structure of propositions, the system of storage of semantic units, and to distinguish between the predicate theory and the case theory of propositions-and so on. Implicit in this approach is the view that meaning is a property of a string of words, independent of the history of our subjects. From a selectionist view, such experiments tell us only about the typical verbal practice of the community from which the subjects were recruited. There is no general truth, no fact of the matter to be discovered. A string of words drawn from a hat, or carefully constructed by an experimenter to serve as a stimulus, has no meaning whatever. Once presented to a subject, the string has a meaning in the sense that it has an effect on the subject, just as a tone might have a meaning to a laboratory rat. But the meaning of the string, like the meaning of the tone, is a product of the history of the subject, not an intrinsic property of the stimulus. Most of our subjects are sophisticated enough to pretend that the stimulus string was generated by someone responding to a real state of affairs; the generality of such experiments depends on such dubious inferences.

Verbal communities maintain common contingencies so that most members respond in similar ways to given verbal stimuli. The uniformity and universality of these contingencies provide the order that we try to capture when we speak of meaning or reference. However, as noted above, such contingencies can be satisfied in many idiosyncratic ways. As we have said, contingencies of selection, be they contingencies of reinforcement or contingencies of natural selection, do not yield classes with essential properties. We may mask this variability by averaging our data, but such variability is not error, it is an unavoidable and important outcome of the processes of selection. It is the variant, not the mean, that is most profitably examined, for selection operates on individuals or elements, not on means.

### **Essentialism in Other Areas of Psychology**

The tendency to slip into essentialist modes of thinking is not confined to one perspective in psychology. The avowed essentialist is rare within any theoretical tradition, but even the most committed selectionist finds it difficult to avoid reifying distinctions once they have been named.

## ***Behavior Analysis***

We have claimed that radical behaviorism is the most thoroughgoing selectionist approach within psychology. One of the central distinctions of that approach is the operant-respondent distinction, which captures the generalization that some response systems, toward one end of a continuum, are narrowly committed to particular functions and are recruited only by specific stimuli or their antecedents, whereas other response systems, toward the other end of the continuum, are virtually unconstrained in their potential interactions with the environment. Accordingly, two different experimental procedures, the operant and classical conditioning procedures, must be used to investigate the full range of relations between the environment and behavior that may be selected by reinforcers. The experimental procedures themselves have defining properties, specified a priori by the canons of the paradigm. The procedures, then, like dodecahedrons, have essential properties.

It does not follow, however, that the responses studied with the operant and classical procedures have essential properties (i.e., that operants and respondents are two different *types* of responses requiring two different *types* of conditioning principles for their description). And yet, it is not uncommon to find such distinctions debated within the technical literature and generally accepted within the secondary literature. Indeed, textbooks on learning often classify responses as operant or respondent, independent of procedural considerations, as if they differed in some essential quality of "operancy." For example, it is sometimes asked if a particular response is "really" a respondent, or if a particular event is not "really" an eliciting stimulus rather than a discriminative stimulus. None of this is to say that the operant-respondent distinction is not a critically important one for many purposes, but that the distinction can lend itself to essentialist misinterpretations (Donahoe, Millard, Crowley, & Stickney, 1982).

Another distinction within the radical-behavioral tradition that lends itself to essentialist interpretation is the distinction between contingency-shaped and rule-governed behavior (Skinner, 1966a). Contingency-shaped behavior is most immediately understood in terms of its prior consequences; rule-governed behavior is most immediately understood in terms of verbal, contingency-specifying antecedents (Cerutti, 1989; Schlinger & Blakely, 1987). Thus, a person who ingratiates him or herself with another by listening attentively to his or her conversation may be doing so because, in the past, such behavior has been reinforced with social reinforcers provided by the other person. Alternatively, attentive listening may be the result of just having taken a Dale Carnegie course on "How to win friends and influence people." Behavior of the same topography—attentive listening—can be regarded as contingency shaped in the first instance and rule governed in the second. .

The distinction between contingency-shaped and rule-governed behavior exemplifies one of Skinner's most important contributions—the rejection of topography of response as a sufficient defining variable in the study of behavior. As illustrated in the preceding example, a response of the same topography may be brought about by many different means—through social and nonsocial contingencies, through imitation of others, or—in individuals with appropriate histories—through the effect of instructions on behavior. When we first sit down at a computer, a phonograph, or an automobile, our behavior is guided almost entirely by instructions; with practice, natural contingencies exert their effects. Although behavior on the two occasions may be topographically similar, important differences remain, differences that can be attributed to the shaping contingencies.

The distinction between rule-governed behavior and contingency-shaped behavior arose from these considerations, and the labels are frequently helpful. However, contrary to the 'Spirit of the original distinction, the terms are sometimes used as if the distinction lay in the behavior itself, as if there were two different *types* of behavior. For example, rule-governed behavior, we are told, may be less sensitive to its consequences than contingency-shaped behavior. The data underlying this conjecture (e.g., Matthews, Shimoff, Catania, & Sagvolden, 1977) are of indisputable importance, but the essentialist cast of the interpretation of the different effects of reinforcers on responses of these two origins is unhelpful. It obscures the fact that rule-governed behavior is itself the product of prior selection by its consequences. If the essentialist tone of the distinction between contingency-shaped and rule-governed behavior persists, we should not be too surprised to find a controversy emerge over whether a particular behavior, shown to be sensitive to its consequences, is "really" rule governed.

### ***Associationism***

The other major noncognitive alternative to understanding behavior is associationism. Although there are many selectionist aspects to associationism, one distinction—that between learning and performance—is often redolent of essentialism. When the observed relation between the environment and behavior is not as orderly as the experimenter might wish, the irregularity is often attributed to perturbations of an underlying order—learning—by processes of a different type—those affected by performance variables. The performance variables are thought to obscure the underlying, idealized process much as they are said to do in the formal analysis of syntax. Although the learning-performance distinction reflects an important insight, notably that almost every instance of behavior is a function of many variables, to attribute invariant, thing-like properties to the effects of some of those variables and to deny them to others is akin to attributing some of the characteristics of an organism to its species and others to different types of constructs (e.g., varieties and individual differences; cf. Skinner, 1950; see Donahoe, 1984, for other comments on essentialism within the associationist approach).

### **Connectionism and Selectionism**

Although no domain in psychological science seems to be entirely free of essentialist analyses, we detect a welcome selectionist trend in some recent work. Evolutionary epistemology, ethology, and radical behaviorism have always been avowedly selectionist, of course, but the recent resurgence of interest in connectionism, or parallel distributed processing, is a conspicuous example of a shift of emphasis away from essentialism in cognitive science.

Connectionist models attempt to simulate complex behavior by means of adaptive networks made up of simple processing units, loosely analogous to neurons. The strengths of interconnections among units, which determine the input-output relations that the network can simulate, are often modified by some feedback mechanism (Rumelhart, Hinton, & Williams, 1986). The feedback mechanism that supervises learning has the effect of selecting successful patterns of activation of the pathways in the network. Part of the appeal of adaptive networks lies in their power and their parsimony: Complex operations can be performed by the repeated action of relatively primitive processes. An equally important part of their appeal arises from the fact that the analogy with the nervous system is plausible enough to encourage the hope of a profitable interdisciplinary attack on the problems of complex human behavior (Smolensky, 1988). Indeed, many proponents of adaptive-network models are biologists hoping to simulate

the activity of portions of the nervous system (e.g., Edelman, 1987). Some regret that adaptive-network models are not more closely guided by neurophysiological and neuroanatomical data (e.g., Crick & Asanuma, 1986; Minsky & Papert, 1988), and neural plausibility will surely emerge as one criterion for evaluating such models. Insofar as these models simulate biological processes, that is, insofar as they are informed by experimental analyses of physiology and behavior, they are likely to be consistent with selectionism rather than essentialism (cf. Trehub, 1991).

Three features of adaptive networks are fundamental to their implementation of a selectionist account of complex behavior (Donahoe, 1991). First, a criterion output, given a particular input, can generally be achieved by an indefinite number of different combinations of activated pathways. Second, the terminal topography of the output is the product of a particular sequence of selections by the feedback mechanism. Third, although the initial state of the network may affect the speed with which a criterion output is achieved, given particular feedback contingencies, a wide variety of initial states will all evolve into networks that mediate the criterion input-output function. Taken together, these features reduce the likelihood that adaptive-network models will be given essentialist interpretations. The criterion output is not dependent on anyone pattern of activated pathways (representations). The pattern of activated pathways that produces the criterion output pattern differs with different selection histories, and the selection history—which reflects the structure of the environment—determines the pattern of activated pathways more importantly than does the initial structure of the network (cf. Norman, 1986).

Because the input-output relations mediated by adaptive networks are not the product of invariant, underlying, idealized structures within the network. One is less tempted to scrutinize a network for units (nodes) or patterns of activation that can be labeled *memories*, *engrams*, *images*, *meanings*, or rules—entities that are commonplace in the metaphorical models of traditional cognitive theories. In general, there is no one unit or pattern of activity within the network that corresponds to any of these constructs. To quote Rumelhart and McClelland ("1986), "Many of the constructs of macrolevel descriptions such as schemata, prototypes, rules, productions, etc. can be viewed as emerging out of interactions of the microstructure of distributed models" (p. 125). And in the same vein:

Schemata are not *things*. There is no representational object which is a schema. Rather, schemata emerge at the moment they are needed from the interaction of large numbers of much simpler elements all working in concert with one another. . . In the conventional story, schemata are stored in memory. . . In our case, *nothing stored corresponds very closely to a schema*. What is stored is a set of connection weights which, when activated, . . . generate states that correspond to instantiated schemata. (Rumelhart, Smolensky, McClelland, & Hinton, 1986, pp. 20-21)

The fact that our microstructural models can account for many of the facts about the representation of general and specific information. . . makes us ask why we should view constructs like logogens, prototypes, and schemata as anything other than convenient approximate descriptions. (Rumelhart & McClelland, 1986, p. 127)

Clearly, adaptive network models do not encourage essentialist interpretations.

### *Essentialist Temptations Within Connectionism*

Most adaptive-network researchers pursue their work with little, if any, regard for the essentialism-selectionism issue. Nevertheless, most such models instantiate the three selectionist steps to complexity: variation, selection, and retention. The state of the network is in constant flux, patterns of activation that mediate criterion outputs are strengthened, and successful pathways are preserved as stable connection weights between units in the network.

Although adaptive networks provide a theoretical framework within cognitive science that is congenial to the Selectionist approach to complexity, there are several points at which temptations to essentialism exist. As is the case with other selectionist approaches, such as radical behaviorism, adaptive-network theory does not confer immunity to the charms of essentialism. Only two such enticements are mentioned here—the search for simple correspondences between the nodes and pathways of the network and verbally characterized cognitive processes, and the effort to endow the feedback mechanisms that select the connections with essentialist properties.

*Verbal characterization of adaptive networks.* As mentioned earlier, in an adaptive network no simple relation exists between the units or patterns of activity and any normative cognitive construct. More precisely, there is, in general, little correspondence between our verbal descriptions of the activity of specific elements of the network and our verbal description of the cognitive processes that we observe in ourselves and infer in our subjects. The patterns of activity within the network *specify* our verbal descriptions (Gibson, 1979) in the sense that they constitute sufficient conditions for the descriptions, but their dimensions are incommensurate with those of the verbal description. Minsky and Papert (1988) made the point this way:

It is because our brains primarily exploit [adaptive networks] that we possess such *small* degrees of consciousness, in the sense that we have so little insight into the nature of our own conceptual machinery. . . . What appear to us to be direct insights into ourselves must be rarely genuine and usually conjectural. (p. 280)

In spite of the absence, in general, of a one-to-one correspondence between network elements and cognitive constructs, we may anticipate continued efforts to identify such simple correspondences. Furthermore, because some apparent correspondences may indeed be occasionally found, such efforts are likely to persist for some time. It must be appreciated, however, that even when correspondences are identified, they are most likely to be constituents of transient patterns of neural activity rather than enduring essentialist entities. For example, one may find in the brain of a sheep a cell whose firing is highly correlated with the presence of the complex constellation of stimuli provided by faces of other sheep (Kendrick & Baldwin, 1987). Nevertheless, that cell is best interpreted not as an abstract representation of *face*, but as a cell that is only one element of a network of neural activity reliably activated by the complex stimuli provided by faces (and, perhaps, other objects as well). The selectionist origin of face cells is revealed in the finding that face cells in sheep do not respond to inverted faces of sheep, whereas monkey face cells do correspond to their inverted counterparts (Kendrick & Baldwin, 1987). Monkeys have a selection history that includes the inverted faces of conspecifics; sheep do not.

*Nature of feedback mechanisms.* Feedback mechanisms in adaptive-network theory serve to modify the strength of connections so that a given input from the environment will cause a given

output to occur from the network. Various versions of adaptive network theory use different feedback mechanisms, and some of these feedback mechanisms have essentialist characteristics. For example, one commonly used mechanism, the delta rule (Rumelhart, Hinton, & Williams, 1986), requires that the correct output pattern be known from the beginning of training so that differences between the obtained output on a trial and the ideal output may be fed back to the network. Such a feedback mechanism is essentialistic in that it assumes that which must be discovered by the learner—the correct response. Feedback rules of this form are both regrettable and unnecessary. They are regrettable in that they introduce an element of teleology: The network must know in advance what the ideal output should be. They are unnecessary in that a feedback mechanism that does not make such an assumption and that is consistent with the behavioral and neurophysiological analysis of reinforcement is sufficient (cf. Donahoe & Palmer, 1989; Donahoe et al., in press; Sutton & Barto, 1981; Widrow & Hoff, 1960). In summary, the emergence of parallel distributed processing admits a selectionist approach within cognitive science, but, as with radical behaviorism, it does not ensure that a selectionist interpretation will be consistently adopted.

### Summary

Contingencies of selection are capable of yielding orderly classes of complex phenomena, and these classes may be stable so long as the contingencies remain stable. However, Selection contingencies merely set limits on class membership; they do not provide blueprints. Elements within classes will vary within these limits, constrained only by those mechanisms that generate variability in the first place. If the duration of our observations is small relative to the duration of the contingencies, such classes may seem to have essential properties. In any case, once classes are named, the fundamental variability of all products of selection is obscured.

Skinner was perhaps unique within psychology in correctly grasping the variable nature of his subject matter and specifying an appropriate methodology. In doing so he wove his science seamlessly into the fabric of modern biology, permitting interpretations of behavior that draw, with equal facility, from ontogenetic and phylogenetic contingencies. In contrast, many contemporary cognitive scientists celebrate their freedom from behaviorist methodological constraints mindless, apparently, of the reasons for heeding these constraints. To the extent that their analyses are implicitly essentialistic, their science is discontinuous with biology. Although recent practice in virtually all areas of psychology reveals examples of essentialist thinking, current interest among cognitive scientists in adaptive networks and the current interest among linguists in natural selection suggest a trend that may ultimately yield a selectionist theme that unifies cognitive science and permits a fruitful dialogue with radical behaviorism.

### REFERENCES

- Andresen, J. T. (1990). Skinner and Chomsky: Thirty years later. *Historiographia Linguistica*, 175, 145-165.
- Baldwin, J. M. (1895). *Mental development in the child and race*. New York: Macmillan.
- Baldwin, J. M. (1980). *Darwin and the humanities*. New York: AMS Press. (Original date of publication, 1909)
- Basalla, G. (1988). *The evolution of technology* Cambridge, England: Cambridge University Press.

- Burma, B. H. (1949). The species concept: A semantic review. *Evolution*, **3**, 369-370.
- Campbell, D. T. (1956a) Adaptive behavior from random response. *Behavioral Science*, **1**, 105-110.
- Campbell, D. T. (1956b). Perception as substitute trial and error. *Psychological Review*, **65**, 330-342.
- Campbell, D. T. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review*, **67**, 380-400.
- Campbell, D. T. (1974). Evolutionary epistemology. In P.A. Schlipp (Ed.), *The philosophy of Karl Popper* (Vol. 14, pp. 413-463). LaSalle, IL: Open Court.
- Campbell, D. T. (1975). On the conflicts between biological and social evolution and between psychology and moral tradition. *American Psychologist*, **30**, 1103-1126.
- Cerutti, D. T., (1989). Discrimination theory of rule-governed behavior. *Journal of the Experimental Analysis of Behavior*, **51**, 259-276.
- Chomsky, N. (1959). Review of *Verbal Behavior* by B. F. Skinner. *Language*, **35**, 26-58.
- Chomsky, N. (1965). *Aspects of a theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1969). Some empirical assumptions in modern philosophy of language. In S. Morgenbesser, P. Suppes, & M. White (Eds.). *Philosophy, science, and method: Essays in honor of Ernest Nagel* (pp. 260-285). New York: St. Martin's Press.
- Chomsky, N. (1975). *Reflections on language*. New York: Pantheon.
- Chomsky, N. (1980a). *Rules and representations*. New York: Columbia University Press.
- Chomsky, N. (1980b). Rules and representations [Author's response]. *The Behavioral and Brain Sciences*, **3**, 1-61.
- Crick, F. H. C., & Asanuma, C. (1986). Certain aspects of the anatomy and physiology of the cerebral cortex. In D. E. Rumelhart, J; L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol 2. Psychological and biological models* (pp. 333-371). Cambridge, MA: MIT Press.
- Crowder, R. G. (1982). The demise of short-term memory. *Acta Psychologica*, **50**, 291-323.
- Darwin, C. (1950). *The origin of species*. London: Watts. (Original work published 1859).
- Dawkins, R. (1986). *The blind watchmaker*. New York: Norton.
- Donahoe, J. W. (1984). Skinner-The Darwin of ontogeny? *The Behavioral and Brain Sciences*, **7**, 287-288.
- Donahoe, J. W. (1991). Selectionist approach to verbal behavior. Potential contributions of neuropsychology and adaptive networks. In L. J. Hayes & P. N. Chase (Eds.). *Dialogues on verbal behavior* (pp. 119-145). Reno, NV: Context Press.
- Donahoe, J. W., Burgos, J. E., & Palmer, D. C. (1993). Selectionist approach to reinforcement. *Journal of the Experimental Analysis of Behavior*, **60**, 17-40.
- Donahoe, J. W., Millard, W. J., Crowley, M. A., & Stickney, K. A. (1982). A unified principle of reinforcement. In M. L. Commons, R. J. Herrnstein, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 2. Matching and maximizing accounts* (pp. 493-521). Cambridge, MA: Ballinger.
- Donahoe, J. W., & Palmer, D. C. (1989). The interpretation of complex human behavior: Some reactions to *Parallel distributing processes*. *Journal of the Experimental Analysis of Behavior*, **51**, 399-416.
- Donahoe, J. W. & Palmer, D. C. (1994). *Learning and complex behavior*. Boston: Allyn & Bacon.
- Edelman, G. (1987). *Neural Darwinism*. New York: Basic Books.

- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Gould, S. J. (1980) *The Panda's thumb*. New York: Norton.
- Grimshaw, J. (1981). Form, function, and the language acquisition device. In C. L. Baker & J. J. McCarthy (Eds.). *The logical problem of language acquisition* (pp. 165-182). Cambridge, MA: MIT Press.
- Hobbes, T. (1968). *Leviathan*, Harmondsworth, England: Penguin. (Original work published 1651).
- Hornstein, N. (1990). Selecting grammars. *Behavioral and Brain Sciences*, **3**, 735-736.
- James, W. (1907). *Pragmatism*. New York: Longmans Green.
- Jerne, N. K. (1955). The natural selection theory of antibody formation. *Proceedings of the National Academy of Sciences*, **4**, 849-856.
- Kendrick, K. M. & Baldwin, B. A. (1987). Cells in temporal cortex of conscious sheep can respond preferentially to the sight of faces. *Science*, **236**, 448-450.
- Locke, J. (1975). *An essay concerning human understanding*. Oxford, England: Clarendon Press. (Original work published 1690)
- Matthews, B. A., Shimoff, E., Catania, A. C., & Sagvolden, T. (1977). Uninstructed human responding: Sensitivity to ratio and interval contingencies. *Journal of the Experimental Analysis of Behavior*, **27**, 453-467.
- Mayr, E. (1976). *Evolution of the diversity of life*. Cambridge, MA: Belknap Press.
- Mayr, E. (1982). *The growth of biological science*. Cambridge, MA: Harvard University Press.
- Mayr, E. (1988). *Toward a new philosophy of biology: Observations of an evolutionist*. Cambridge, MA: Belknap Press.
- Medin, D., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-195). Cambridge, England: Cambridge University Press.
- Minsky, M. L., & Papert, S. A. (1988). *Perceptrons* (Expanded ed.). Cambridge, MA: MIT Press.
- Norman, D. (1986). Reflections on cognition and parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (pp. 531-546). Cambridge, MA: MIT Press.
- Palmer, D. C. (1986). Chomsky's nativism: A critical review. In P. N. Chase & L. J. Parrott (Eds.), *Psychological aspects of language* (pp. 49-60). Springfield, IL: Charles C Thomas.
- Palmer, D. C. (1991). A behavioral interpretation of memory. In L. J. Hayes & P. N. Chase (Eds.) *Dialogues on verbal behavior* (pp. 261-279). Reno, NY: Context Press.
- Pinker, S. (1982). A theory of the acquisition of lexical interpretive grammars. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 655-726). Cambridge, MA: MIT Press.
- Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Pinker, S., & Bloom, P. (1990). Natural selection and natural language. *Behavioral and Brain Sciences*. **13**. 707-784.
- Popper, K. (1957). *The poverty of historicism*. Boston: Beacon Press.
- Popper, K. (1972). *Objective knowledge: An evolutionary approach*. Oxford, England: Clarendon Press.
- Pringle, J. W. S. (1951). On the parallel between learning and evolution. *Behaviour*, **3**, 175-

- Root-Bernstein, R. S. (1989). *Discovering*. Cambridge, MA: Harvard University Press.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 318-362). Cambridge, MA: MIT Press.
- Rumelhart, D. E., & McClelland, J. L. (1986). PDP models and general issues in cognitive science. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 110-146). Cambridge, MA: MIT Press.
- Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes in PDP models. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (pp. 757). Cambridge, MA: MIT Press.
- Schlinger, H., & Blakeley, E. (1987). Function-altering effects of contingency-specifying stimuli. *The Behavior Analyst*, **10**, 41-46.
- Skinner, B. F. (1935). The generic nature of the concepts of stimulus and response. *The Journal of General Psychology*, **12**, 40-65.
- Skinner, B. F. (1938). *The behavior of organisms*. Englewood Cliffs, NJ: Prentice-Hall.
- Skinner, B. F. (1948). *Walden two*. New York: Macmillan.
- Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review*, **57**, 193-216.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.
- Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1963). Behaviorism at fifty. *Science*, **140**, 951-958.
- Skinner, B. F. (1966a). An operant analysis of problem solving. In B. Kleinmuntz (Ed.), *Problem solving: Research, method, teaching* (pp. 225-257). New York: Wiley.
- Skinner, B. F. (1966b). The phylogeny and ontogeny of behavior. *Science* **153**, 1205-1213.
- Skinner, B. F. (1971) *Beyond freedom and dignity*. New York: Knopf.
- Skinner, B. F. (1981). Selection by consequences. *Science*, **213**, 501-213.
- Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, **213**, 1-23.
- Staddon, J. E. R. (1983). *Adaptive behavior and learning*. Cambridge, England: Cambridge University Press.
- Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, **88**, 135-170.
- Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review Monograph Supplements*, **2**(4, Whole No.8).
- Treuhub, A. (1991). *The cognitive brain*. Cambridge, MA: MIT Press.
- Watkins, M. J. (1990). Mediationism and the obfuscation of memory. *American Psychologist*, **45**, 328-335.
- Watson, J. B. (1930). *Behaviorism*. New York: Norton.
- Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. *Institute of RadioEngineers, Western Electric Show and Convention. Convention Record* (Pt. 4), 96-104.