In Verbal behavior and general behavior theory.
T. R. Dixon and D. L. Horton (Eds.). Englewood
Cliffs, N.J., Prentice-Hall (1968).

A Formal Limitation of Associationism

24 T. G. BEVER, J. A. FODOR, M. GARRETT

Classical associationism has attempted to elucidate principles whereby complex ideas are constructed out of simple ones. Similarly modern associationism has attempted to devise a body of learning principles which can explain how complex skills are constructed out of simple operants and reflexes. Given such a set of principles, one might validate them in either of two ways. In the first case, one might attempt to show that complex behaviors can be added to the repertoire of an organism by the formation of associations between appropriate rudimentary behaviors. That has been the course that associationists have typically chosen for testing their principles. They have tried to show that manipulation of the variables upon which associative strength depends permits the laboratory simulation of complex behaviors (such as the construction of "response chains" in maze learning) or of complex psychological phenomena (such as the production of "selective forgetting" by variation of stimulus order). For the cognitive psychologist in particular, the continuing interest of associationism rests on the assumption that conceptual behaviors can be approximated in the laboratory by techniques based on the putative laws of association.

There is, however, a second approach to the validation of associative principles. For, given any theoretical principles for psychological description, one may study the kinds of behavioral repertoires their operation can represent *in principle*. That is, assuming that the principles exhaustively characterize the learning mechanisms available to a hypothetical organism, one can determine the limits their operation imposes upon the organism's behavioral repertoire. A partial ordering can thus be imposed upon the set of behaviors so that learning principles capable of describing the assimilation of the more complex of them can describe the assimilation of the simpler ones, but not conversely. Thus, for associative principles in particular there is an upper bound on the richness of the repertoires they are capable of explaining, and we can ask of any particular behavioral ability whether it lies above or below that bound.

The important point for present purposes is that certain human abilities lie beyond the upper bound on *any* set of learning principles that could reasonably be called "associative." Certain kinds of conceptual competences fall outside the explanatory power of associationism, given the kinds of constraints on learning principles that have tradi-

tionally defined associationism. Moreover, it can be shown that there are infinitely many such counterexamples to the adequacy of associationistic accounts of learning.

We assume that the following meta-postulate is a necessary condition on any set of principles being called "associative": that is, by definition, no theory of learning counts as associative unless it satisfies this postulate.

> **The Terminal Meta-Postulate: Associative principles are rules defined over the "terminal" vocabulary of a theory, i.e., over the vocabulary in which behavior is described. Any description of an n-tuple of elements between which an association can hold must be a possible description of the actual behavior.**

Notice, first, that the satisfaction of this meta-postulate is independent of the particular choice of a vocabulary for describing behavior. It does not matter whether psychological relations are taken to be relations of ideas, as in classical associations, or relations among stimuli and responses. The postulate requires only that the vocabulary chosen for psychological descriptions of output states must also be the vocabulary over which the associative rules are defined. That is, the psychological theory will not contain any element which is abstractly related to the elements of the behavior.

Second, the terminal meta-postulate does not preclude associations between 'overt behaviors' and 'intervening states' so long as the internal processes can be described in the same vocabulary (or isomorphic derivatives) the theory uses to describe overt behavior. In particular, the postulate is satisfied by "mediation" theories, since such theories suppose the intervening states in associative chains to be drawn from the stimulus and response elements in which the behavior itself is described (see Fodor, 1965).

A corollary of the terminal postulate is that, since behavior is organized in time, every associative relation is a relation between left and right elements of a sequence. In a well-known article, Lashley (1951) showed that a special case of this corollary is unsatisfied for many sorts of behavior; namely, the case in which each right member of a behavior chain is associated with the immediately preceding left member. Lashley showed that, for a great variety of behaviors (of which typing mistakes may be considered paradigmatic), a left member of a chain is dependent upon a nonadjacent right member (as when we type *Lalshey* for Lashley).

Lashley's argument was, in effect, generalized in Chomsky's 1957 monograph, which showed that there are indefinitely many learnable behaviors not describable by principles which allow association (or any other form of dependency) only between left and right members of a behavior chain; in particular, by principles which satisfy the terminal postulate.

Consider what a subject does when he learns to recognize mirror-image symmetry in figures without explicitly marked contours. The infinite set of strings belonging to a mirror-image language is paradigmatic of such symmetrical figures, i.e., all sequences of a and b of the form $X\bar{X}$,

accept	*reject*
aa	ab
abba	aaab
aabbaa	baabba
abbbba	abbabb
etc.	etc.

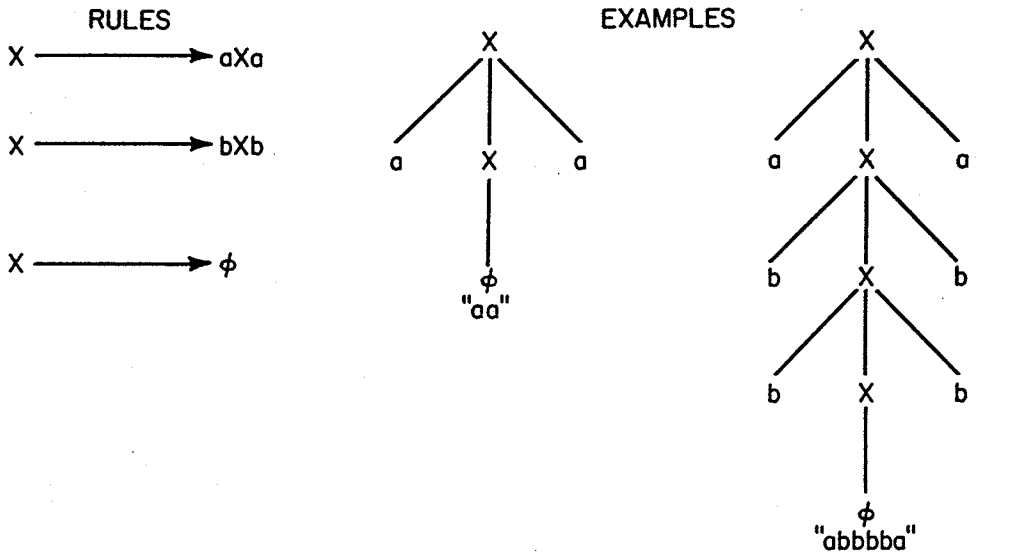
Someone who learns this language has acquired the ability to accept as well formed in the language all and only strings consisting of a sequence

of *a*'s and *b*'s followed immediately by the reverse of that sequence.

The question is whether or not an organism whose behavior is determined solely by associative principles can select just the set of sequences that satisfy this criterion. In fact, it is provable that the answer to this question is no. This is *not* simply because the set of strings consonant with the rule is infinite (it is easy to design an automaton with a finite memory but an infinite behavioral repertoire, e.g., the language containing any number of *a*'s followed by any number of *b*'s). It is rather because of the particular kind of relation holding between the left and right halves of strings in the mirror-image language; namely, that dependencies are allowed to nest within dependencies (for further discussion, see Chomsky, 1957, 1963).

Interestingly, the weakest system of rules which allows for the construction of a mirror-image language is precisely one which violates the terminal postulate. That is, it is one which allows rules defined over elements that are precluded from appearing in the terminal vocabulary (e.g., rules defined over items other than *a*'s and *b*'s in the example just cited). Such rules yield a simple characterization of the mirror-image language. (The rules are unordered; "ϕ" stands for the null element in the diagram below.)

Notice that the X in these rules explicitly violates the terminal postulate of associationism. If it were to appear in a terminal string, the system would generate strings not in the mirror-image language (e.g., aXa). (Intuitively, the X in the rules above is a formal representation of the hypothetical "center" around which each element on the left is rotated to the right.) Thus, an organism that has learned the mirror-image language has learned a concept that cannot rest on the formation of associations between behavioral elements. In general, behavioral abilities which involve recursion over abstract elements violate the terminal meta-postulate, since there are usually elements in the description which do not appear in the behavior. Such abilities include the distinguishing of sentences from non-sentences, verbs from nouns, and many other abilities related to natural language.



This argument appears to us to provide a conclusive proof of the inadequacy of associationism for these kinds of natural behaviors. It might be replied that there are indefinitely many behavioral repertoires which *can*

be described by associationistic principles (e.g., all the finite state languages in the sense of Chomsky *op. cit.*). Hence, continued research on associationism could be justified by the illumination it might cast on such behaviors. However, this is to evade the point of our argument. We have considered associationism to require certain constraints upon the formulation of learning principles. Theories that are more powerful than associationism are at least theories that have weaker constraints. Hence, any behavior that can be characterized by associative principles can *ipso facto* be characterized by the more powerful models. Such models should not, therefore, be considered as alternatives to associative models; rather, associative rules are simply special cases of the rules employed by more powerful theories. If the rules are allowed, you are allowed the associative rules, but not conversely. As Sol Saporta has put it, anything you can do with one hand tied behind your back, you can do with both hands free.

References

Chomsky, N., *Syntactic Structures*. The Hague, Netherlands: Mouton and Co., 1957.

———, "Formal Properties of Grammars," in *Handbook of Mathematical Psychology*, Vol. II, R. Luce, R. Bush, and E. Galanter, eds. New York: Wiley & Sons, Inc., 1963.

Fodor, J. A., "Could Meaning Be an r_m?" *Journal of Verbal Learning and Verbal Behavior*, 4 (1965), 73–81.

Lashley, K. S., "The Problem of Serial Order in Behavior" in *Cerebral Mechanisms in Behavior*, L. A. Jeffress, ed. New York: John Wiley & Sons, Inc., 1951.

Bever, T.G., Fodor, J.A., & Weksel, W. (1968). A formal limitation of associationism. T.R. Dixon & D.L. Horton (Eds.), Verbal behavior and general behavior theory. Prentice-Hall, Inc.