Storage and Later Recognition of Exemplars of Concepts[1]

JUDITH S. REITMAN[2]

The University of Michigan

AND

GORDON H. BOWER

Stanford University

This paper concerns the problem of abstraction: whether when we en-
counter several exemplars of a concept, we retain only the abstracted con-
cept, only the exemplars, or both. Although many studies concur that both
are stored, a recent article argued strongly that only the abstracted concept
is stored. The present study, aimed at replication of this recent finding,
follows the earlier procedural details but adds appropriate controls and uses
simpler material. A set of 24 exemplars of four concepts, in the form of
four-tuples of letters and numbers, was presented to Ss who, after presenta-
tion, rated a larger set of exemplars for recognition. One group of Ss experi-
enced the conceptual exemplars; control group Ss experienced items that
were similar in composition but not exemplars of a concept. Two major
results appeared: Unlike the study on which this was based, all Ss were
able to distinguish those items that were originally experienced from those
that were not. And, the more completely an exemplar fit the concept (the
longer the item), the more confident the S was that it had been presented.
In contrast, in the control condition, the longer the item, the more con-
fident the S was that it had *not* been presented earlier. Two models are
described to account for these results. One is based on the S's initial storage
of the exemplars in a concept-plus-correction format; the other is based
on a procedure whereby the S can make recognition judgments without
having previously abstracted and retained the concepts.

This paper concerns the age-old problem of abstraction. There are
two fundamental questions that learning theorists have asked regarding
abstraction: (1) what are the processes by which general principles are

[2] Requests for reprints should be sent to Judith S. Reitman, The University of
Michigan, Mental Health Research Institute, 205 North Washtenaw Place, Ann Arbor,
MI 48104.

194

abstracted from experience with different exemplars; and (2) what specific information is retained regarding the exemplars themselves? This paper is concerned with the second question, the retention of exemplar information.

The extreme positions on this question, of course, are that either everything or nothing is retained about each exemplar. Both approaches assume that the abstract principle is learned. The "retain everything" position is incorporated by several induction programs developed in artificial intelligence research on pattern recognition (Hunt, Marin, & Stone, 1966; Nilsson, 1965; Sebestyen, 1965). These programs accumulate complete information on many different instances in the process of calculating a decision rule to classify the instances. The "retain nothing" position has been expressed in the writings of Bartlett (1932), Oldfield (1954), and others. This position states that only general principles or "schema" are abstracted from S's experience with exemplars and retained, and that the exemplar information is lost when assimilated into the schema.

The truth doubtless lies between these extremes: Along with the general principle, Ss probably retain some but not all of the exemplar information. This outcome is clear in studies on memory for dot patterns (Posner & Keele, 1968, 1970; Strange, Keeney, Kessel, & Jenkins, 1970). For example, Posner and Keele (1968) taught their Ss to classify distortions of several "prototype" concepts defined by a pattern of dots in a matrix. They later tested retention by noting how Ss categorized old distortions of the prototypes, new distortions, and the prototypes themselves. Since Ss categorized the prototype patterns as accurately as the old distortions, it was concluded that they had inferred the abstract schema from the set of exemplars. However, Ss also categorized old distortions, seen before, more accurately than new distortions, never seen before, leading to the conclusion that something about the original, specific instances was stored. In other words, along with the prototype schema some exemplar information had also been stored.

Recently, however, the "store nothing" position has been argued quite forcefully by Bransford and Franks (1971) and Franks and Bransford (1971) on the basis of their data. Since our experiment follows the line started by Bransford and Franks, their procedure should be described. Their Ss were exposed to many different exemplars generated from a simple principle; later, Ss were tested for recognition memory for the exemplars presented earlier, for exemplars not presented earlier, and for non-exemplars which violated the general principle. The principle for generating exemplars was of the following form:

$$S \rightarrow (A) \quad (B) \quad (C) \quad (D) \qquad\qquad (1)$$

That is, each string was generated by selecting and concatenating one or more of the optional elements A, B, C, or D. In Bransford and Franks (1971), the elements were "atomic" propositions, and the concatenation operation was that of clause embedding in a complex sentence. Suppose, for example, that the four atoms were A = The car is old, B = The car pulled the trailer, C = The car climbed the hill, and D = The hill was steep. Then several exemplar strings generated by Rule 1 would be AC = The old car climbed the hill, BC = The car pulling the trailer climbed the hill, and ABCD = The old car pulling the trailer climbed the steep hill. There are $2^4 - 1 = 15$ "legal" strings generated by this rule from the four atoms, and about half of these were presented to the Ss in the Bransford and Franks experiment. Moreover, their Ss were exposed concurrently to four such sets of exemplars (i.e., the atomic elements, A to D, differed among sets; each set called a concept), all intermixed in scrambled order. The Ss did not know during presentation that retention of the specific exemplars would be tested.

Two primary results were yielded by the recognition memory tests in the Bransford and Franks experiment. First, their Ss were completely unable to discriminate between presented versus nonpresented exemplars of the concepts, although they accurately rejected illegal test strings which violated the concepts, e.g., "The new car followed the trailer down the hill." Second, Ss' ratings of confidence in recognition were higher the greater the number of elements in the legal test string. In fact, the string ABCD evoked the highest recognition confidence rating (of having been presented) despite the fact that no four-element string had ever been presented during the initial exposure. The conclusions, then, are that Ss remember only general concepts (e.g., Rule 1 applied to certain sets of elements), not the particular exemplars, and that a "recognition rating" reflects the proximity of the test string to the full-form of the concept. This second finding was neither expected nor satisfactorily explained by Bransford and Franks (1971).

A few factors in the Bransford and Franks design militate against accepting the conclusion that Ss usually retain only schema and not the particular exemplars. One problem is that memory must be overloaded when the S is exposed concurrently to seven or eight exemplars of each of four different concepts, and is then asked to remember which specific combinations of elements occurred, e.g., that AC and ABC occurred but AB did not. The overloading might have been even more severe because each element was itself a full sentence and contained peripheral syntactical information. Too, in formal terms, judgments of which specific combinations of elements within a concept had occurred should exhibit massive amounts of interference due to the formal overlap and similarity

of the exemplars (as well as the similarities among the different concepts). These remarks argue that the Bransford and Franks materials and procedures may have arranged conditions in such manner as to cause massive forgetting of the particular exemplars.

Also, the Bransford and Franks experiment was missing a "nonconcept" control group of Ss. Subjects in such a group would be exposed to a series of overlapping strings constructed not by applying Rule 1 to certain ordered sets of elements, but by randomly selecting and combining elements from a small unordered set of elements. If these Ss also later failed to distinguish between presented and nonpresented strings, one might then interpret the earlier findings in terms of the interference and overloading of memory in the situation rather than in the forgetting of the particular exemplars when they are assimilated into a schema.

The Noncases which Bransford and Franks' Ss rejected—and which were offered as evidence for their having learned something—were semantic deviates of the atomic propositions. But the issue is not whether the basic elements of the concept were stored; both theories, the "retain everything" and the "retain nothing" theory, suppose that they would be. Even a control group presented with random combinations should be able to reject test sentences composed of atomic propositions which deviate in identity from those heard earlier. The important issue requiring demonstration is retention of specific combinations of atoms, not retention of the atoms themselves. In this regard, then, Bransford and Franks failed to demonstrate that their Ss retained anything beyond the atomic propositions. This was perhaps to be expected, since in their experiment each atomic proposition was presented many times (2–6) during the study phase and many times during the recognition testing series, whereas a specific combination of atoms (e.g., ABD) would occur at most once.

For such reasons, we decided to repeat the essentials of the Bransford and Franks (1971) procedure but to use simpler learning materials and to include an appropriate control condition.

METHOD

The Bransford and Franks procedure was replicated with the following main exceptions. First, the type of material differed; elements consisted of letters and digits, although strings were generated by Rule 1. Second, each string was presented only once for acquisition and tested only once for recognition. Finally, Ss in two control groups experienced material that was identical in elemental composition, but different in the existence or obviousness of concepts.

The experimental session was divided into an acquisition session followed by an unexpected recognition test. Acquisition was done in the guise of testing S's short-term memory for the exemplar strings. On an acquisition trial, the S read aloud a single string which was presented in the window of a memory drum for 2 sec. The drum then turned to a three-digit number from which S subtracted threes aloud as rapidly as he could. After 10 sec of subtracting, he attempted to recall the string. After a 10-sec recall period, the next item and three-digit number were presented. This procedure continued until 24 strings were presented, eight of each of three different concepts.

For the recognition session, the S received a sheet listing either 45 or 65 items, depending on his particular condition. He marked each item on a scale from −5 to +5, the −5 indicating a judgment that the item clearly had not been presented, and the +5 that the item clearly had been presented.

There were three conditions. In the first two conditions, the acquisition strings were exemplars of three concepts, Rule 1 applied to three different sets of elements. For example, for the concept in which Rule 1 is applied to the elements 1234, the 15 legal exemplars are: 1234, 123, 124, 134, 234, 12, 13, 14, 23, 24, 34, 1, 2, 3, and 4. The first two conditions differed in the obviousness of the concepts and presumably in the difficulty of acquiring them. Condition Ob, the Obvious concept condition, involved concepts such as 1234, ABCD, and wxyz. Condition Df, the Difficult concept condition, involved concepts such as 41AB, 3yzD, and w2Cx. The 12 elements in the Obvious condition were mixed to form the three Difficult concepts.

During acquisition, the Ss in Conditions Ob and Df experienced eight of the strings generated by Rule 1 applied to each of the three concepts. Half of the possible one-, two-, three-, and four-tuples were presented. For recognition, all 45 legal items (15 × 3 concepts) were presented, plus 20 illegal items. Fifteen of the 20 illegal items were constructed by using the elements from within a single set but by violating their customary left–right order. For the concept 1234, item 143 exemplifies an order violation. Five of the 20 illegal items were composed by selecting elements from different sets and violating their customary order. Given 1234, ABCD, and wxyz as concepts, items such as 12BA and 41x are examples of double violations. The legal items presented in acquisition are called OLD items; the matching half of the legal items not presented, NEW:LEGAL items; and the two kinds of violation items, NEW: ILLEGAL(ORDER) and NEW:ILLEGAL(ORDER and SET), respectively.

In the third condition, the Random condition (R), neither concepts

nor rules existed. The 12 elements from Conditions *Ob* and *Df* were randomly grouped, with the restriction that the acquisition and recognition sets include the same number of one-, two-, three-, and four-tuple strings as in the first two conditions.

The entire experiment was replicated three times, each time with a new set of 12 elements. Thirty-six Ss, 12 per Condition, from the departmental pool of paid Ss participated individually in the experiment. All were Stanford University students; half were female and half were male, balanced over the three conditions. Each S was paid $1.75 for the single-hour session.

RESULTS

Each S in Conditions *Ob* and *Df* made 65 judgments of recognition, scores ranging from −5 to +5. Those in Condition *R* produced only 45 judgments, due to the absence of possible NEW:ILLEGAL items. Where appropriate these scores were separated into the four item categories: OLD, NEW:LEGAL, NEW:ILLEGAL(ORDER), and NEW:IL-LEGAL(ORDER and SET). An average rating was calculated for items of equal length within each category for each S. Two analyses of variance were then applied to these average scores. The first assessed (OLD versus NEW:LEGAL) × (Conditions *Ob, Df,* and *R*) × (Length of item, 1–4); the second assessed (Kind of NEW item) × (Conditions *Ob* and *Df*) × (Length of item, 2–4).

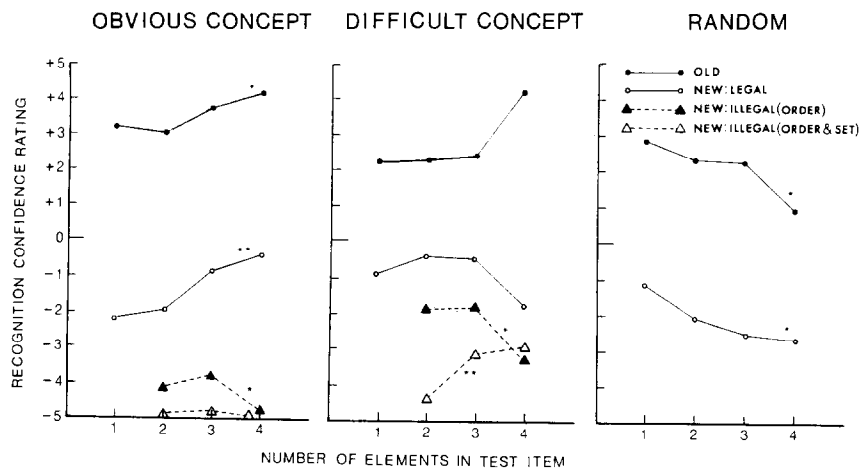The primary results are shown in Fig. 1, one panel for each condition.

FIG. 1. Recognition confidence ratings for OLD and legal and illegal NEW items of various lengths for three conditions.

A first, strong result is that Ss discriminate well between OLD items seen previously and NEW:LEGAL items that fit the rule but were not seen previously. This contrast was very significant overall ($F(1,33) = 224.05$, $p < .01$) and was present in each condition. There was no significant main effect for the Length of the test string, but there was an interesting interaction between Length and Condition ($F(2,33) = 5.76$, $p < .01$). We will say more about this interaction later. Compared across the three panels of Figure 1, the overall level of ratings seen to be somewhat lower for Condition R ($F(2,33) = 4.14$, $p < .05$) which is itself of little interest. The several kinds of NEW items were rated quite differently ($F(2,33) = 56.99$, $p < .01$); NEW:LEGAL strings were rated highest followed by NEW:ILLEGAL(ORDER) strings, and then by NEW:ILLEGAL(ORDER and SET) strings. There is also an interaction between kind of NEW item and Condition ($F(2,44) = 10.77$, $p < .01$), reflecting lower ratings for NEW items in Condition Ob.

A more detailed analysis was performed on the function relating confidence ratings to the length of the test item, and how this relationship varied across conditions. For this purpose, a linear trend component from a one-way analysis of variance was computed for each S's judgments of items of various lengths, one trend component for each of the four categories. The trend component was then assigned to each S as his score for that kind of item. The scores were then tested for difference from zero.[3] This test assesses the relationship between the item lengths and ratings, without involving the large variances between Ss. This test indicates which of the 10 lines in Figure 1 have slopes significantly different from zero.

Accordingly, those lines in Fig. 1 that have a reliable nonzero slope are marked with a single asterisk if significant beyond the .05 level, or a double asterisk if beyond the .01 level. Furthermore, the slopes of the lines for OLD and NEW:LEGAL items in Condition Ob were compared, respectively, with those corresponding lines in Condition R and found to be significantly different ($t(\text{OLD}) = 3.16$; $t(\text{NEW}) = 7.52$, each $p < .01$, d.f. = 11).

In sum, Ss appeared not only able to abstract the rules which generated the items (that is, NEW:LEGAL > NEW:ILLEGAL), but also able to retain the specific items seen (OLD > NEW). With an obvious rule to generate the items, the longer the test item, the more confident the S was that he had seen it before (Condition Ob, lines sloping upward). When there was no rule, the longer the test item, the more certain the S

---

[3] As suggested by J. E. Keith Smith, personal communication. A discussion of the linear trend component analysis can be found in Winer (1962).

was that he had *not* seen it before (Condition *R*, lines sloping down-ward). When there was a rule to generate the items but it was difficult (Condition *Df*), there was no regular overall relationship between item length and recogniition ratings.

Since Ss in Condition *Ob* produced a positive relationship and Ss in Condition *R* a negative one, it is probable that Condition *Df* represents some mixture of these effects. This mixture may occur within Ss or be-tween Ss. Since the variance of the trend scores is larger for Condition *Df* than for Conditions *Ob* and *R*, the mixture is more likely between Ss. About half of the Ss may have noticed the concepts and rules, behaving as Ss in Condition *Ob*, while the remainder may not have noticed the rules, behaving as Ss in Condition *R*. For this reason, discussion will involve only the "pure" results from Conditions *Ob* and *R*.

DISCUSSION

The experiment was designed around three questions: Does S retain for long the particular experiences that he abstracts into concepts? How do his judgments of recognition differ when he has acquired a concept versus when he has merely experienced a list of similar items? Why do Ss recognize complete instances of concepts with greater confidence than incomplete ones? These questions will be discussed below in that order.

Is there specific memory for the experiences that were abstracted into concepts? From the results of this study and those reviewed other than Bransford and Franks, we may conclude that there is some specific mem-ory for the experiences that were abstracted into concepts. The material and procedures in the Bransford and Franks studies would appear to be a hindrance to retention of the particular exemplars experienced. Perhaps the more accurate emphasis is to view the Bransford and Franks experi-ments as demonstrating that conceptual rules may be abstracted and retained despite conditions which hinder retention of much information about particular experiences.

Do recognition judgments differ when the experiences are unrelated but confusing rather than related by a rule? When Ss had acquired con-cepts, the relationship between the length of an item and recognition confidence ratings was clearly different from when the Ss were simply confused by many similar experiences. The relationship was positive when the items were generated by a rule but negative when items were similar but not rule generated.

Why are recogniiton confidence ratings related to the lengths of test items? In this as in the Bransford and Franks study, the longer items were ones that were more complete instances of the full concept. Thus, this question can be rephrased: Why are recognition confidence ratings

related to the goodness-of-fit of the test item to the full form concept? Below, we develop several plausible answers to this question.

One theory could be that in learning the items presented, S stores the full unit presented, but that tag "generalizes" to subunits of the pattern. For example, when presented with the string 134, S would attach tags in memory to the elements 1, 3, 4, to the diagrams 13 and 34 (but not 31 or 43), and to the triplet 134. Later, one could reconstruct with fair accuracy from information about subunit frequency which items must have been presented earlier. This reconstruction process might be detailed as follows.

Imagine that when S is presented with a test item for recognition, he searches his memory for tags associated to the individual elements, to the pairs of adjacent elements (digrams) in the test item, to the triplets of adjacent elements (trigrams), etc. The confidence rating for an item of length n would be computed in n passes. First is a check for recognition of single elements. The memory representations of the single elements of the test string are examined for frequency of associated tags. If the frequency of tags on a given element exceeds a threshold number, we say it is "recognized"; otherwise, it is rejected. In this first pass, if any element in the string is not recognized, then the entire test string is rejected and given a low rating. (None of these types of strings was included in our experiment.)

If each individual element is recognized and if the string contains two or more elements, then a second pass begins, in which the recognition rating will be augmented according to the digrams in the string. The memory representation of each digram of the string is consulted for the number of tags associated to it, indicating frequency of past occurrence. If the number of tags on the digram exceeds a threshold amount, then that digram is recognized and a tally is added to a counter for the string. If the frequency is below threshold, that digram fails its recognition test and a tally is subtracted from the string's counter. The counter accumulates tallies across the $(n - 1)$ digrams of the test string. Similarly, if the item has three or more elements, a third pass is made to collect tally points for the $(n - 2)$ trigrams. The passes continue in like manner until tallies are collected for all n-gram sub-units of the string. The final confidence rating is a linear transformation of the accumulated tally for the test string.

The outcome of this process can be described mathematically. Let p denote the probability that a particular n-gram has a training frequency (number of tags) sufficient to exceed the threshold for placing $a + 1$ in the tally counter. Then the total tally for a homogeneous string of length n is expected to be

$$\sum_{i=1}^{n-1} (n - i)[p(+1) + (1 - p)(-1)] \quad \text{or} \quad (2p - 1) \sum_{i=1}^{n-1} (n - i)$$

or $(p - .5)\ n\ (n - 1)$. This expected tally is increasing when $p > .5$ and decreasing when $p < .5$). For Ss in the rule-generated condition, the legal strings are composed only of legal n-grams, and we shall assume that for all of them $p > .5$. Therefore, in the rule condition, the ratings for legal strings will have positive slope. By definition, illegal strings contain one or more nonoccurring n-grams (for which $p = 0$), so they will therefore have a lower overall tally.

For the "random non-concept" condition, it would be argued that $p < .5$, yielding a negative slope for the relation between item length and recognition rating. In support of this argument, it should be noted that since many more different n-grams would occur during training in the random condition than in the rule condition, particular n-grams will be infrequent and consequently will not be learned very well. In terms of the model, there would be more n-grams over which a fixed number of frequency tags would be spread; the number of n-grams that consequently have a sufficient number of tags to exceed threshold would be much lower.

The separation between OLD and NEW strings in Fig. 1 is represented in terms of the value of p. Frequently occurring n-grams have $p > .5$, infrequently occurring n-grams have $0 < p < .5$, and nonoccurring n-grams have $p = 0$. This model accounts for most of the data in Fig. 1.

In summary, then, this view is that the conceptual rule *per se* is not learned; rather, the set of training strings generates a certain frequency distribution into the memory registers for sub-units, and recognition ratings for test strings are computed from these registers. The theory could apply in principle to other materials, whether the elements are atomic propositions (Bransford & Franks, 1971), geometric forms in a larger picture (Franks & Bransford, 1971), or the locations of dots in a matrix (Posner & Keele, 1968). For example, with matrix dot patterns, the elementary structural features (corresponding to the n-grams) would be a set of "masks" or "templates" for small local regions obtained from fragments of the training patterns (cf. Uhr & Vossler, 1963). With linguistic material, the basic "n-grams" counted could be embeddings of n subject-predicate constructions.

A statistical decision strategy such as this may explain why it sometimes appears that S can identify and remember a nonpresented prototype better than presented distortions of the prototype (Posner & Keele, 1970; Strange, Keeney, Kessel, & Jenkins, 1970). The prototype simply

contains a greater total number of frequently presented subunits than does any single distortion. For example, if every string of the concept ABCD had been presented and stored except for the quadragram itself, still ABCD would receive a tally of 4 (=5 minus 1 for the tag missing on the quadragram) whereas the presented trigram ABC would only receive a tally of 3. This theory argues then that the storage of particulate information is veridical—it is not degraded by assimilation to a proto-type. Rather, the later recognition-judgmental process is "biased" or "distorted" in the ways indicated.

The n-gram tally theory above has several difficulties. First, it seems unlikely that p, the probability that a particular n-gram exceeds thres-hold, is invariant across n, the length of the particular n-gram. For example, since at presentation each string is broken down into all subunit n-grams and each of those n-grams is tagged, over the entire presentation sequence, more digrams will be tagged than quadragrams. Consequently, the probability that there will be enough tags to exceed threshold should be greater for digrams than for quadragrams. We cannot, therefore, as-sume p to be the same for both short and long n-grams. An augmented model accommodates these differing probabilities. The expected total tally for any one test string of length n would be

$$\sum_{i=1}^{n-1} (n-i)[p_i(+1) + (1 - p_i)(-1)] \quad \text{or} \quad \sum_{i=1}^{n-1} (n-i)(2p_i - 1).$$

From the above argument, it is expected that for $i < j$, $p_i > p_j$. The resulting tally would have a positive slope only when

$$\sum_{i=1}^{n-1} p_i > \frac{n-1}{2},$$

negative slope when

$$\sum_{i=1}^{n-1} p_i < \frac{n-1}{2},$$

and be zero otherwise, for n being the length of the longest item.

The augmented model now presents another difficulty. It must rely upon a fortuitous combination of circumstances and threshold settings to get the p_i's (or just the p's in the simpler model) to sum appropriately. A third, and more serious difficulty with both the above models is that they predict that the S would have to do a great deal of computing to construct the concepts if he were asked. The reaction time for produc-

tion or recognition of an item which fit the full concept would be unreasonably longer than for any presented subunit.

The chief alternative theory would suppose that the prototype effect, the influence of the proximity of the string to the concept, is a learning phenomenon, not a judgmental one. Importantly, after a concept has been learned, new presented strings can be identified as the concept plus one or another minor distortion or correction. For instance, the string ACD could be encoded as "concept ABCD, B missing"; the string AB as "concept ABCD, C missing, D missing." In this way, the prototype "concept ABCD" would accumulate a high implicit frequency during training and so yield higher recognition ratings and show greater resistance to forgetting than would the specific instances ("corrections") which have lower frequencies. This theory would then predict that the S's later production of the concept should not take a longer time than any particular exemplar since it does not have to be calculated from the stored information. The basic problem with this approach is its failure to specify how the concept or prototype is acquired unless some specific particulars are stored about early instances. The intermediate position is that particular exemplar information is stored initially until the conceptual rule is inferred; thereafter, encoding of instances changes so that particulate information is no longer entered or maintained.

This discussion has not so much resolved as restated and recast the basic "retain everything" versus "retain nothing" hypotheses about conceptual induction. We have tried to place in perspective a novel finding by Bransford and Franks (1971) which appeared to disprove, or at least seriously challenge, one of these theories. A theoretical and experimental critique of their findings shows that reports of the demise of the ancient problem have been premature: It is still very much alive and kicking.

REFERENCES

BARTLETT, F. C. Remembering. Cambridge University Press, 1932.

BRANSFORD, J. C., & FRANKS, J. J. Abstraction of linguistic ideas. Cognitive Psychology, 1971, 2, 331–350.

FRANKS, J. J., & BRANSFORD, J. D. Abstraction of visual patterns. Journal of Experimental Psychology, 1971, 90, 65–74.

HUNT, E. B., MARIN, J., & STONE, P. Experiments in induction. New York: Academic Press, 1966.

OLDFIELD, R. C., & ZANGWILL, O. L. Head's concept of the schema and its application in contemporary British psychology. British Journal of Psychology, 1942, 32, 267–286; and 1943, 33, 58–64, 113–129, 143–149.

OLDFIELD, R. C. Memory mechanisms and the theory of schemata. British Journal of Psychology, 1954, 43, 14–23.

NILSSON, N. J. Learning machines. New York: McGraw-Hill, 1965.

POSNER, M. I., & KEELE, S. W. On the genesis of abstract ideas. *Journal of Experimental Psychology,* 1968, **77,** 353–363.

POSNER, M. I., & KEELE, S. W. Retention of abstract ideas. *Journal of Experimental Psychology,* 1970, **83,** 304–308.

SEBESTYEN, G. *Decision making processes in pattern recognition.* New York: McGraw-Hill, 1965.

STRANGE, W., KEENEY, T., KESSEL, F. S., & JENKINS, J. J. Abstraction over time of prototypes from distortions of random dot patterns. *Journal of Experimental Psychology,* 1970, **83,** 508–510.

UHR, L., & VOSSLER, C. A pattern-recognition program that generates, evaluates, and adjusts its own operators. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought.* New York: McGraw-Hill, 1963.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.