"Schema Abstraction" in a Multiple-Trace Memory Model

Douglas L. Hintzman
University of Oregon

A simulation model of episodic memory, MINERVA 2, is applied to the learning of concepts, as represented by the schema-abstraction task. The model assumes that each experience produces a separate memory trace and that knowledge of abstract concepts is derived from the pool of episodic traces at the time of retrieval. A retrieval cue contacts all traces simultaneously, activating each according to its similarity to the cue, and the information retrieved from memory reflects the summed content of all activated traces responding in parallel. The MINERVA 2 model is able to retrieve an abstracted prototype of the category when cued with the category name and to retrieve and disambiguate a category name when cued with a category exemplar. The model successfully predicts basic findings from the schema-abstraction literature (e.g., differential forgetting of proto- types and old instances, typicality, and category size effects), including some that have been cited as evidence against exemplar theories of concepts. The model is compared to other classification models, and its implications regarding the abstraction problem are discussed.

How is abstract knowledge related to specific experience? In present-day terms, this question concerns the relationship be- tween episodic and generic memories. This article explores the possibility that there is only one memory system, which stores episodic traces, and that abstract knowledge as such does not have to be stored but can be derived from the pool of traces of specific experiences at the time of retrieval. I demonstrate how this might work by applying a simulation model of a multiple- trace memory theory to the schema-abstraction experimental paradigm, which is widely believed to capture in the laboratory the processes by which generic or abstract ideas are formed.

Multiple-trace theories assume that each event to which one attends gives rise to its own memory trace. Thus, repetition of an item such as a word in a list does not strengthen a prior representation (i.e., one predating the experiment or one laid down by the item's first experimental occurrence); rather, it produces a new trace that coexists in memory with traces of other occurrences of the same item. Experiments supporting the multiple-trace assumption have been primarily concerned with the ability of subjects to remember an item's presentation frequency, list membership, presentation modality, exposure duration, serial position, and so forth (e.g., Hintzman, 1976; Hintzman & Block, 1971; Hintzman, Block, & Summers, 1973; Hintzman, Grandy, & Gold, 1981; Hintzman, Nozawa, & Irm- scher, 1982). These are all episodic memory tasks, and multiple-

trace theories are primarily accounts of how repetition affects episodic memory.

We become familiar with concepts, however, through re- peated encounters with category members, and so repetition also plays a role in the acquisition of abstract categories. Is this effect of repetition mediated by a different mechanism than the one involved in episodic memory tasks? One view is that it is, and that repeated exposure to exemplars of a category produces not only traces of the individual events in an episodic memory system but also a unitary, abstract representation of the cate- gory in a functionally separate generic memory system (e.g., Tulving, 1983). The alternative view under consideration here is that only traces of the individual episodes are stored and that aggregates of traces acting in concert at the time of retrieval represent the category as a whole. The simulation model shows, in principle, how the latter might be accomplished. It demon- strates, moreover, that a multiple-trace model comprising a simple set of episodic memory assumptions can account for many of the phenomena that have been reported in the litera- ture on the schema-abstraction task.

There appear to be two basic ways in which people learn to classify objects and events. One is through the presentation and naming of category exemplars (as a child typically learns to tell dogs from cats). The other is through explicit communication about definitions (e.g., that of prime number), exceptions (e.g., a bat is not a bird), and the like. Learning through examples appears to be a prerequisite to learning through communica- tion, in that the latter depends on the learner's knowing the meanings of at least some of the words. One reason for concen- trating on the schema-abstraction task, rather than naturally learned concepts, is that the task appears to capture in the labo- ratory the more basic of these two types of learning, uncon- founded by the acquisition of information through language. A second reason is that the task allows careful control of variables crucial to the understanding of learning—such as the frequency, recency, and similarity of exemplars—which are usually un- known in the case of naturally learned concepts. For these rea-

sons, in its present form the theory advanced here attempts to deal only with learning by example. It is assumed, however, that linguistically communicated information is encoded in the same episodic memory store.

The present multiple-trace theory of classification represents the "exemplar" approach to concepts (e.g., Smith & Medin, 1981, pp. 143–161) taken to a logical extreme—where even an individual, if encountered (and attended to) on several occasions, is represented by more than one memory trace. Thus certain capabilities of exemplar models in general can be inferred from the results reported here.

The simulation model, dubbed MINERVA 2, consists of a set of core processes that interact with various task environments, in several different computer programs. To date, the model has been tested with some success on recognition memory, memory for frequency, and paired-associates learning tasks, in addition to schema abstraction. A preliminary summary of this work is found in Hintzman (1984), and frequency judgments and recognition memory are detailed in Hintzman (1986). Only the schema-abstraction work will be discussed here. In presenting the model, I shall attempt to describe the core processes generally but in enough detail so that a reader could duplicate the essentials of MINERVA 2 in any high-level programming language. However, the computer model is only a metaphor for the theory, and so it is appropriate to begin with an overview of the theoretical ideas that the model is intended to represent.

Theoretical Overview

Every experience is assumed to be represented internally as an active configuration of primitive properties (many of these could be the output of Pylyshyn's, 1985, "transducers"). The number of primitive properties is assumed to be large and to range from simple emotional tones and modality-specific sensory features (e.g., basic colors and odors) through properties accessible by more than one modality (e.g., intermittency, spatial location), to primitive abstract relations (e.g., before, same as, greater than, has as parts). These properties are not acquired by experience, although the ability to label them may be.

In assuming a large set of primitive properties of which some are abstract, the present theory takes a rationalist, as opposed to empiricist, view (Fodor, 1981). What is most important for present purposes, however, is that the number of primitive properties, although large, is much smaller than the number of experiences that a person has. Thus, experiences share properties, and the similarity of any two experiences can be related roughly to the number of properties they share.

A memory trace is a record of an experience or episode, and so it preserves—perhaps imperfectly—the configuration of primitive properties making up the experience. Every conscious experience gives rise to its own memory trace, no matter how similar it may be to an earlier one. Thus, phenomena that are repeated but nevertheless command attention will be represented in memory over and over again. Because it will be necessary in what follows to distinguish between the active representation of the current experience, on the one hand, and the vast pool of largely dormant memory traces, on the other, I shall adopt James's (1890) terminology and refer to the former

as in *primary memory* (PM) and the latter as in *secondary memory* (SM).

Communication between primary and secondary memory is assumed to be restricted to two simple operations: (a) A retrieval cue or "probe" can be sent from PM to all traces in SM, and (b) PM can receive a single reply or "echo" that emanates back from SM.

The probe is an active representation of an experience, in PM, that is communicated in parallel to all traces in SM. This communication process can be imagined as activation traveling upward, over connections linking primitive-property nodes, at a lower level, to episodic-trace nodes, at a higher level. Alternatively, it can be thought of as a kind of resonance process in which the information in PM is broadcast simultaneously to all the traces in SM (Ratcliff, 1978). Whichever imagery one adopts, each trace is assumed to be activated according to its similarity to the probe. Thus, traces sharing many properties with the probe are activated strongly, whereas traces that overlap little with the probe are activated hardly at all.

Activation of a trace implies activation of all its primitive properties. That is, as the node-and-link image suggests, activation of the trace via the properties it shares with the probe spreads to the trace's other properties, which it does not share with the probe. In this way, information that the probe itself does not contain may be activated in memory. This is the basis of associative learning.

The echo that is returned to PM following a probe has two characteristics: intensity and content. The intensity of the echo depends on the total amount of SM activation that was triggered by the probe. The greater the similarity of traces to the probe and the greater the number of such traces, the greater is the intensity of the echo. Echo intensity thus can serve as a kind of familiarity signal. Recognition memory, judgments of frequency, and judgments of familiarity are assumed to be based primarily on the intensity of the echo (Hintzman, 1986).

The content of the echo is the pattern of activation among primitive properties caused by the reactions of all SM traces to the probe. The echo may itself be an experience in PM—that is, an active configuration of the properties shared by many of the traces that were strongly activated by the probe. It is the way in which the echo content is determined that distinguishes the present theory from most other current approaches to memory, and so the point needs to be emphasized. The probe activates all SM traces simultaneously, and all traces contribute simultaneously to the echo. Thus the content of the echo reflects the summed contributions of all the traces in SM, each responding according to its similarity to the probe. If several traces are strongly activated, then the content of the echo will reflect primarily their common properties, and—much as one sour note or peculiar accent is drowned out in a chorus of voices—a characteristic that distinguishes one activated trace from the others will be masked in the echo emanating back from SM.

In a system such as this, the structure of the probe is crucial, because it determines which particular combination of SM traces will contribute importantly to the echo. Depending on the specificity of the probe, the subset of strongly activated traces may be large or small, and depending on the unanimity of the traces in that subset, the echo content may be ambiguous or clear. In this way, information of various degrees of abstrac-

tion can be quickly retrieved from a single store of episodic memory traces. Apparently, a full-fledged memory theory using this scheme to select a few episodic traces from a very large pool and abstract information from them in parallel was first proposed by Richard Semon (1909/1923). The present theory is compared with Semon's, among others, later in this article.

The Model

In MINERVA 2, memory traces are represented as lists of features, much as they are in Bower's (1967) multicomponent model of memory. An experience—or a trace representing an experience—is a string of integers in the range -1, 0, $+1$. A value of $+1$ in position j may be thought of as excitation of property j, a value of -1 as inhibition of the property, and a value of 0 as indeterminate. In most cases, feature lists are generated by a random process in which -1 and $+1$ are equally likely. This scheme was chosen primarily for its simplicity and for some of its mathematical properties, which will be mentioned below.[1]

Learning consists of copying the features of an experience into a trace. In MINERVA 2 this is done probabilistically, each feature being encoded with probability L. Forgetting is treated as the inverse of learning; each nonzero feature has probability F of reverting from -1 or $+1$ to 0. Because learning and forgetting are complementary processes, learning with $L = .25$ would be equivalent in its effects to learning with $L = 1.00$, followed by forgetting with $F = .75$. It would be inappropriate to assume that the model somehow captures all of the variables that contribute significantly to the performance of human subjects in an experiment, and so no attempt has been made to fit data quantitatively. Rather, the simulations were intended to investigate the functional relationships that the model predicts under the manipulation of a variety of experimental variables. For this purpose the exact values of parameters are usually not crucial, and so in most of the simulations to be reported L was set at 1.

Activation of SM traces by the probe is depicted in Figure 1. Let $j = 1 \ldots n$ index feature position (from left to right in Figure 1), where n is the total number of primitive properties, and let $i = 1 \ldots m$ index traces (from top to bottom in Figure 1), where m is the number of traces in SM. Then $P(j)$ represents the value of the jth feature of the probe, and $T(i, j)$ represents the value of the jth feature of trace i. The similarity of trace i to the probe is given by

$$S(i) = (1/N_R) \sum_{j=1}^{n} P(j)T(i, j), \quad (1)$$

where N_R is the number of features that are relevant to the comparison, the number for which either $P(j)$ or $T(i, j)$ is nonzero. $S(i)$ behaves much like a Pearson r, being 0 if the probe and trace are orthogonal and having a maximum value of $+1$. Negative values are also possible. Also, $S(i)$ can be interpreted as X, where $X(j) = +1$ when $P(j)$ and $T(i, j)$ are identical, $X(j) = -1$ when $P(j) = -T(i, j)$, and $X(j) = 0$ when either $T(i, j) = 0$ or $P(j) = 0$. The numerator is a version of Tversky's (1977) "contrast model" of similarity. Dividing it by N_R gives $S(i)$ upper and lower bounds of $+1$ and -1.

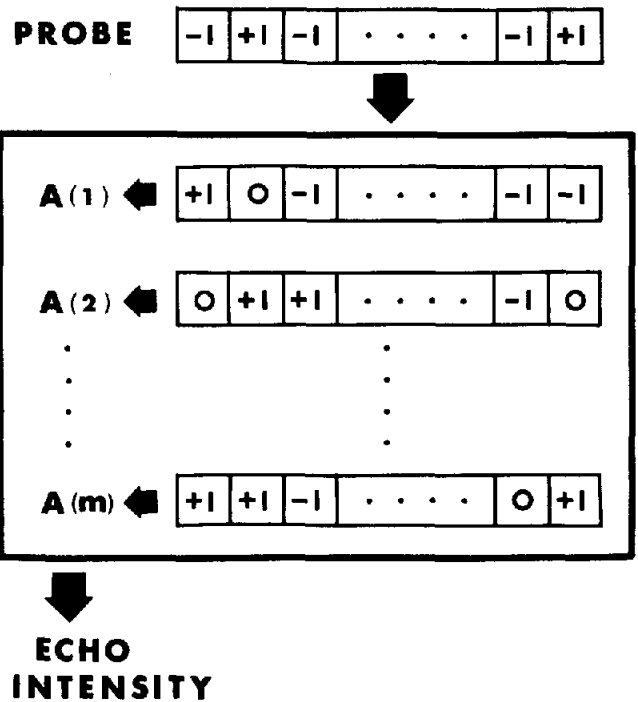The degree to which trace i is activated by the probe is a nonlinear function of similarity:

Figure 1. Trace activation. (Each trace is activated according to its similarity to the probe. Feature values [$j = 1 \ldots n$] are listed from left to right, and traces [$i = 1 \ldots m$], from top to bottom. $A(i)$, the activation level of trace i, depends on the proportion of features it shares with the probe. Echo intensity is the sum of the $A(i)$ values.)

$$A(i) = S(i)^3. \quad (2)$$

Raising the similarity value to the power 3 produces generalization gradients that are concave upward and, in effect, increases the signal-to-noise ratio in the echo. Consider, for example, a target trace and a nontarget trace having similarities to the probe of .8 and .2, respectively. If activation were a linear function of similarity, just four such nontarget traces could overshadow the effect of the target trace. As given by Equation 2, however, the $A(i)$ values of the two traces will be .512 and .008, a ratio of 64 to 1. The power 3 also has the advantage of preserving the sign of the similarity measure, so that when trace i is orthogonal to the probe, the expected value of $A(i)$ is 0. If the power to which $S(i)$ is raised is considered a parameter of the model, then other activation functions are acceptable as long as the sign of $S(i)$ is retained in $A(i)$. In the present work, however, the only activation function used was that given in Equation 2.

As Figure 1 indicates, all traces in SM contribute to the intensity of the echo. Echo intensity is found by summing the activation levels of all m traces:

$$I = \sum_{i=1}^{m} A(i). \quad (3)$$

---

[1] This representation should be seen as a computational device only. No claim is made that the complex structure of experience can be captured in simple feature lists.
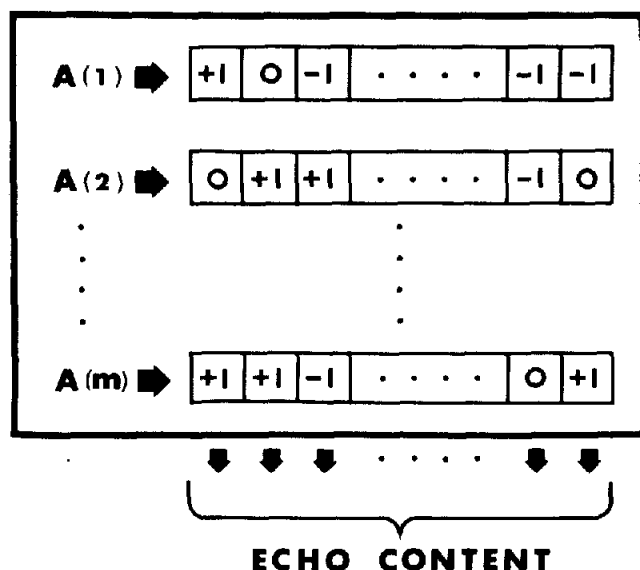
Figure 2. Retrieval of echo content. (Activation initiated by the probe [see Figure 1] is passed down to all features of each trace, as the product of $A(i)$ and the feature value. For each feature, j, the products are summed over traces to yield $C(j)$. Echo content is the set of $C(j)$ values.)

If there are no target-related traces in memory, positive and negative $A(i)$ values should roughly balance, and the expected value of I will be 0. In general, the more SM traces match the probe and the greater their similarity to the probe, the greater is the value of I. The I dimension has been used to simulate performance in the recognition memory and frequency judgment tasks (Hintzman, 1984, 1986); however, in the present simulations no use is made of I.

Derivation of the content of the echo is shown in Figure 2. The activation level of each trace is fed down to its constituent features by multiplying each feature value by $A(i)$. The activation of each feature in the echo is then found by summing these products over traces:

$$C(j) = \sum_{i=1}^{m} A(i)T(i,j). \qquad (4)$$

The histogram or pattern of $C(j)$ values shows which features are shared by the strongly activated traces. Large positive or negative $C(j)$ values are common to a large majority of the activated traces, whereas small $C(j)$ values are not.

The histogram of $C(j)$ will always resemble that of $P(j)$, because the probe strongly activates only traces to which it is similar. However, the activated traces may contain information not included in the probe, and so the histograms of $C(j)$ and $P(j)$ can systematically differ. This is how MINERVA 2 accomplishes associative recall. Because the schema-abstraction task requires that category names be produced when category exemplars are presented, the present simulations make use of this capacity.

Consider a way in which associative information about name–face pairs could be represented in SM and retrieved. Let $j = 1 \ldots k$ represent features of the names, and $j = k + 1 \ldots n$ represent features of the faces. Then each trace with the full

complement of features $j = 1 \ldots n$ filled in encodes an experience with a name–face pair. Now, to retrieve a face given a name, a probe is constructed that has $P(1) \ldots P(k)$ filled in, and $P(k + 1) \ldots P(n)$ empty—that is, all equal to zero. In the echo content, the name features $C(1) \ldots C(k)$ will be fairly faithful replicas of those in the probe, whereas features $C(k + 1) \ldots C(n)$ will resemble the features of the associated face. Retrieval of a name given a face can be done in the opposite fashion.

Schema-Abstraction Task

Before turning to the simulation results, we need to consider some of the basic phenomena that are to be explained. Research with the schema-abstraction paradigm originated with the experiments of Posner and Keele (1968), in which subjects learned to classify random dot patterns. In a typical schema-abstraction experiment (e.g., Homa, Cross, Cornell, Goldman, & Schwartz, 1973), subjects may be trained to classify patterns into three categories of different sizes (e.g., 3, 6, and 9 exemplars per category). All exemplars of a particular category are distortions of a single prototype pattern; however, the subjects do not see any of the three prototypes during learning, and they are not told that prototypes exist. After the subjects have learned to classify the 18 exemplars correctly, they are tested for their ability to classify several types of patterns: (a) the "old" training exemplars, (b) the three prototypes, (c) new low-level (moderate) distortions of the three prototypes, (d) new high-level (extreme) distortions of the three prototypes, and (e) "random" patterns unrelated to the three prototypes. The same testing procedure may be carried out, using an independent group of subjects, after a retention interval ranging up to several weeks.

Such experiments routinely produce the following results:

1. Classification of prototypes is more stable over time than is that of old exemplars. A common pattern is for classification of old exemplars to be clearly better on the immediate test but to decline to about the same or a lower level than that of the prototypes after a long delay (e.g., Homa et al., 1973; Posner & Keele, 1970; Strange, Kenney, Kessel, & Jenkins, 1970). This differential forgetting result has been taken as evidence that a representation of the central tendency of the category (a schema or prototype) is abstracted during learning and that this representation decays more slowly than do traces of the training exemplars themselves.

2. Old exemplars are classified better than new exemplars on both the immediate test and the delayed test. The persistence of this superiority has been interpreted as showing that, while the prototype representation plays an increasingly stronger role in classification with the passage of time, traces of the old exemplars remain and may continue to contribute to performance.[2]

---

[2] The assumption seems to be that if only the prototype representation remained then the old and new exemplars, being equally similar to the prototype, would be classified equally well. However, it is similarity to the abstracted schema or empirical prototype derived through experience with the old exemplars that is crucial, not similarity to the prototype itself. Just as a sample mean is usually closer on average to the scores in its own sample than to the scores of another sample from the

3. Transfer of classification from the old exemplars to new patterns is best for prototypes, intermediate for low-level distortions of the prototypes, and worst for high-level distortions. This gradation provides further support for the view that each category is represented by a kind of prototype, and the similarity of the result to typicality effects found with natural concepts is a prime reason for the recent resurgence of interest in the schema-abstraction task (Rosch, Simpson, & Miller, 1976).

4. Transfer to new patterns improves with increasing category size, that is, the number of training exemplars in the category (e.g., Homa et al., 1973; Homa, Sterling, & Trepel, 1981; Homa & Vosburgh, 1976). This result has been taken as evidence of the importance of variation; the greater the number and, hence, variety of exemplars, the more effective the hypothetical prototype abstraction process becomes.

5. The tendency to erroneously assign patterns, particularly random patterns, to a category increases with category size (e.g., Homa et al., 1973, 1981). This result has been seen as an anomaly, for it is difficult to reconcile with the conclusion that large categories are more accurately representated than small ones. (The tendency can be reduced or eliminated by instructions—a result that will be discussed later.)

The functional relations just described appear to be particularly robust, and so their explanation should be a central goal of a model of the schema-abstraction task. In what follows, these and several other findings from the schema-abstraction literature will be considered.

Simulations

Method

With exceptions that will be noted, most of the simulation runs to be described shared the following details:

1. Each memory trace and each probe were represented as a vector of 23 feature values ($n = 23$). Ten values ($j = 1 \ldots 10$) were reserved for the category name, and 13 values ($j = 11 \ldots 23$) for the stimulus pattern.

2. Three categories were used. For each simulated "subject," three category names were generated at random, and these were paired with three randomly generated prototype patterns. For each feature, $Pr(-1) = Pr(+1) = .50$.

3. Exemplars of a category were generated by altering features of the prototype. For a low-level distortion, exactly 2 randomly selected features of the prototype, $j = 11 \ldots 23$, were multiplied by -1. For a high-level distortion, 4 features were changed in this way. (A change of 4 out of 13 features may not appear extreme; however, changing 6.5 of the features would produce a pattern orthogonal to its prototype and, thus, not an exemplar of the category in any meaningful sense.)

4. Exemplars of each of the three categories were generated according to rules that differed from one simulation run to another. In all cases, however, each exemplar was combined with its category name to form a 23-feature event encoded in a single memory trace.

5. Testing was carried out using either name probes lacking stimulus features (i.e., $P(11) \ldots P(23)$ set at 0) or stimulus probes lacking cate-

gory name features (i.e., $P(1) \ldots P(10)$ set at 0). Details of the testing procedure differed in different simulation runs.

Before we turn to simulations of classification data, let us examine the associative recall process in more detail. There are two main points to be made: One concerns the retrieval of abstract category representations, and the other concerns the problem of ambiguity in the retrieval of category names.

Abstraction

With traces of category exemplars stored in SM, what is the effect of a probe with a category name? To find out, 3, 6, and 9 high-level distortions of three category prototypes were generated. Each was combined with its category name, and one copy was stored in SM, as just described, for a total of 18 memory traces. Testing was then carried out using each of the three category names as probes (features $j = 1 \ldots 10$ were filled in and $j = 11 \ldots 23$ set at 0). Twenty subjects were simulated by carrying out the entire procedure, beginning with the generation of new prototypes, 20 times.

An example of the outcome is shown in Figure 3. Here, the name of the 9-item category was used as the probe. The top panel shows the histogram of the original prototype of the category, and the bottom panel shows the histogram of the echo content across the 13 stimulus features, that is, the values of $C(11) \ldots C(23)$. Clearly the echo is not an exact copy of the prototype, but it does strongly resemble it. The correlation between prototype and echo is $r = .82$.

Moreover, the echo resembles the prototype more than it does any of the category exemplars that were stored. We know this because each exemplar was generated from the prototype by a rule ensuring that exactly four feature values were reversed in sign, and in the echo only one feature is opposite in sign from the prototype. This shows that MINERVA 2 is capable of retrieving an "abstract idea," even though an abstraction as such was never stored.

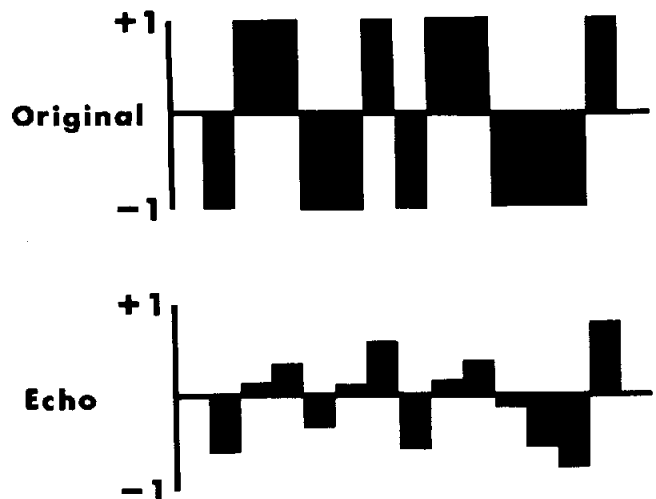Not surprisingly, the degree to which the echo resembles the

Figure 3. Retrieving an abstracted pattern. Top panel: the original prototype of a 9-exemplar category. Bottom panel: the pattern returned in the echo when the category name was used as the probe.

same population, the empirical prototype should be closer to the old exemplars of the category than to new ones. Qualitatively, then, a difference between old and new exemplars in classification is consistent with the prototype-abstraction hypothesis.

original prototype increases with category size. This can be seen by comparing the results of the 3-, 6-, and 9-exemplar conditions averaged across the 20 simulated subjects. Mean echo-prototype correlations for the three conditions (standard deviations in parentheses) were .62 (.07), .73 (.06), and .79 (.06), respectively. The example in Figure 3 is fairly typical of the 9-exemplar condition, having $r = .82$.

Ambiguous Recall Problem

What happens if the prototype of a category is used as a probe in an attempt to retrieve the category name? In 20 additional simulation runs with traces generated as above, each of the three prototypes was used as a probe, and $C(1) \ldots C(10)$ of the echo were correlated with the corresponding features of the original category name. For the 3-, 6-, and 9-exemplar conditions, the mean correlations (standard deviations in parentheses) were .58 (.31), .83 (.20), and .88 (.15), respectively. Because the category names were perfectly encoded in memory ($L = 1$), the noise reflected in these correlations is due to interference from the other items stored in SM. For an example of the degree of match between the original name and the echo in the 9-exemplar condition, see the top two panels of Figure 4.

These far-from-perfect correlations illustrate the problem of ambiguous recall. It would be intolerable for a device that must reliably communicate hundreds or even thousands of category names to produce them in such a sloppy way. To be sure, if there are only three alternatives as there are here, then the obtained echo–name correlations are high enough to allow fairly reliable identification of a retrieved category name. The test item can simply be assigned to the category whose name yields the highest positive correlation with the echo. This is the procedure that was adopted in the simulations that follow.

The procedure is unsatisfying as a general solution to the ambiguous recall problem, however, because it assumes an external memory holding copies of all the acceptable alternatives, to which the echo can be compared. This assumption was made explicitly by Eich (1982), whose holographic network model, like MINERVA 2, suffers from ambiguous recall. Eich assumed that the output of the memory system was sent to a semantic memory having the ability to identify ambiguous patterns. Clearly, however, Eich's solution would violate the spirit of MINERVA 2, which attempts to explain episodic and generic memories without postulating two separate systems.

Fortunately, MINERVA 2 appears to provide an elegant solution to the ambiguous recall problem. As was already mentioned, the top two panels of Figure 4 show an original 9-exemplar category name and the corresponding echo produced when the prototype of that category was used as the probe. The correlation of the retrieved pattern with the original name is .74. The third panel (Echo 2) shows the result of normalizing Echo 1 into the -1 to $+1$ range and using it as a secondary probe. More precisely, in normalization each $C(j)$ value was multiplied by g, where $g = 1/\text{MAX}[C(j)], j = 1 \ldots 23$. Converting Echo 1 into a probe in this way and feeding it back into memory produced Echo 2, which was a better replica of the original name than was Echo 1. Converting Echo 2 into a probe, in turn, produced Echo 3, shown in the fourth panel of the figure. A subse-
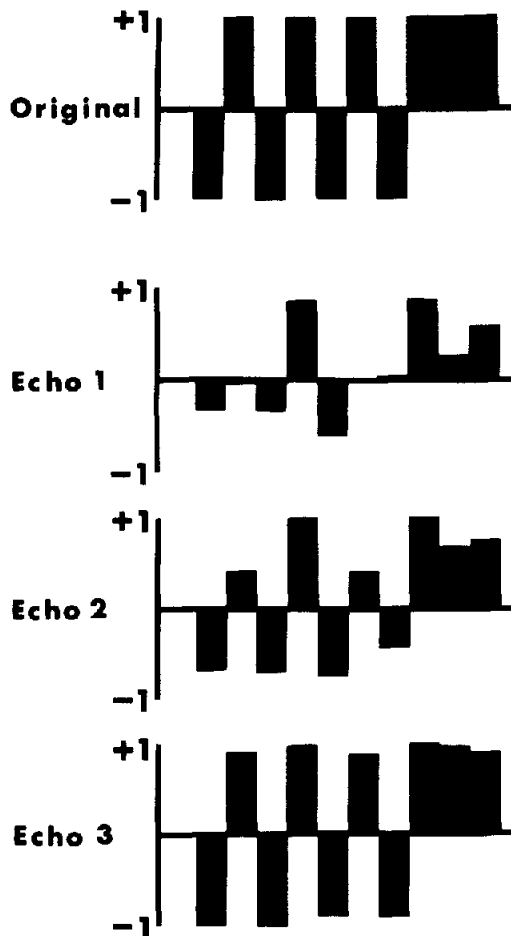

Figure 4. Retrieving and cleaning up a category name. (From top to bottom: [a] the original name of a 9-exemplar category; [b] the pattern returned in the echo when the category prototype was used as the probe; [c] the pattern returned in the second echo when the first echo was converted into a probe; [d] the echo resulting from a second echo–probe conversion.)

quent Echo 4, not shown, was indistinguishable from the original category name.

Figure 4 illustrates that MINERVA 2 is capable of bootstrapping itself out of the ambiguous recall problem. In the rather limited explorations of this process that have been conducted so far, the illustration appears to be entirely typical. Three or four echo–probe conversions are usually sufficient to produce a virtually perfect copy of one of the category names that were originally stored. The final result is usually the correct name, but sometimes it is a perfect copy of one of the alternatives. In the latter case, the subsequent echoes drift away from the correct answer rather than toward it.

One might object to the echo–probe conversion process on the grounds that, in order to know when to initiate or terminate the process, the system must know whether or not the retrieved echo is what it is looking for. However, echo–probe conversion, or its equivalent, might be an automatic component of recall. One possibility is that the procedure is always executed a certain number of times. Another is that each echo is compared

with the probe that initiated it, and another echo–probe conversion is performed only if a discrepancy is found. Because successive echoes appear to be quickly captured by one of the acceptable responses, arrival at a steady state could be used as a signal to stop and produce the response.

What allows MINERVA 2 to clean up its own output as is illustrated by Figure 4? Notice that in the initial probe only features of the stimulus pattern can activate traces, because the name features ($j = 1 \ldots 10$) are all set to zero. Echo 1 returns values for name features, however, and so when it is converted into a probe, traces containing similar name–feature patterns are activated more and traces containing dissimilar patterns are activated less than they were initially. Over successive echo–probe conversions, the strongest activation becomes more and more confined to a particular subset of traces containing a particular category name.[3]

Because the procedure that allows MINERVA 2 to bootstrap its way out of the ambiguous recall problem consumes a considerable amount of computer time, the capability was not used in the following simulations. Instead, whenever the probe was a stimulus pattern, the name features of the subsequent echo, $C(1) \ldots C(10)$, were correlated with each of the category names and the pattern was assigned to the category yielding the highest positive correlation. In case of a tie, one of the tied categories was selected at random, and in case none of the correlations was positive, a residual "junk" classification was used.

Basic Findings

To evaluate the model's performance on the schema-abstraction task, 3, 6, and 9 high-level distortions of three prototypes were generated, respectively, and each was paired with the appropriate category name and encoded in SM as described earlier, for a total of 18 SM traces. Classification testing was carried out using one probe per category representing each of the following conditions: (a) an old exemplar (one that was originally stored in SM), (b) the category prototype, (c) a new low-level distortion of the prototype, (d) a new high-level distortion of the prototype, and (e) a random pattern. A forgetting cycle followed, during which each feature stored in SM—that is, each $T(i, j)$—either remained the same with $Pr = .25$ or reverted to zero with $Pr = .75$. Then testing was done again in the same way. Three hundred subjects were simulated by repeating the entire procedure, from the generation of new prototypes and names through the final test, 300 times.

Correct classification percentages are shown in the form of forgetting curves in Figure 5, with the 3-, 6-, and 9-exemplar data presented in separate panels. Notice first that for each category size, differential forgetting of prototypes and old exemplars was obtained. That is, classification of the prototypes suffered less from forgetting than did that of the exemplars that had been stored in SM. As was indicated, this is the typical result with human subjects (e.g., Posner & Keele, 1970). Essentially the same outcome was produced, using a best-match or nearest-neighbor classification strategy, by MINERVA 1 (Hintzman & Ludlam, 1980). The MINERVA 2 model bases classification on all traces, each weighted according to its similarity to the probe, rather than on a strict nearest-neighbor rule. However, the explanation of the differential forgetting result is basi-
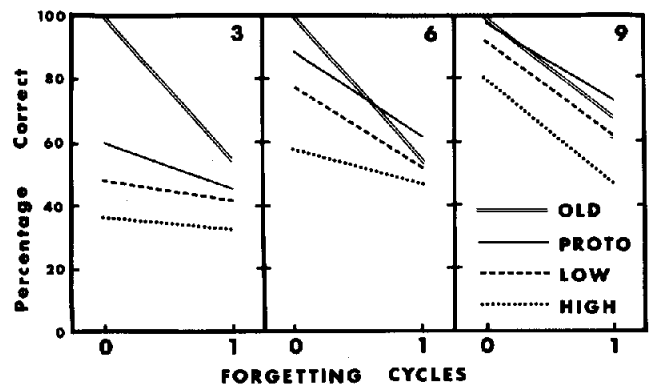

Figure 5. Correct classification percentages as a function of category size (3, 6, or 9); probe type; and number of forgetting cycles. (One forgetting cycle equals 75% forgetting. Secondary memory [SM] contained one copy of each exemplar. Each point is based on $N = 300$.)

cally the same for both models: When little information has been lost from memory, excellent performance is obtained if the probe is identical to one of the originally stored exemplars, but after much information has been lost, a probe that is moderately similar to several of the original exemplars has a statistical advantage in classification over a probe that is identical to only one of them. Put differently, the old exemplar has the edge in the original quality of the relevant information, whereas the prototype has the edge in redundancy. The more noise is added to the communication channel through forgetting, the more important redundancy becomes; hence the obtained interaction.

It is also clear from Figure 5 that classification of the prototype improves dramatically with increasing category size, and transfer to new exemplars, both high- and low-level distortions, follows this same pattern. Thus, in general, as is found with human subjects, transfer is an increasing function of category size (e.g., Homa et al., 1973). Note further that performance on old exemplars always exceeds that on new exemplars of the same degree of distortion—another parallel with human data.

As for the ordering of the transfer conditions, for all three categories the prototypes were categorized best and new high-level distortions worst, with performance on low-level distortions falling in between. This demonstrates the typicality effect that characterizes natural concepts—that is, the fact that category members vary in how representative of the category they are (e.g., Rosch et al., 1976). Further, it is evident that family resemblance must be the basis of classification (e.g., Rosch & Mervis, 1975); defining features do not exist for these categories, and even if they did the model would not encode or use those particular features in any special way.

The only major discrepancy between the simulated data of Figure 5 and data from human subjects is that the forgetting rates—particularly for transfer items in the larger categories—are greater than human data show. A second simulation to be described below demonstrates that these decay rates can be sub-

---

[3] This capture process appears similar to the effect of feedback on the neural net model described by J. Anderson, Silverstein, Ritz, and Jones (1977).
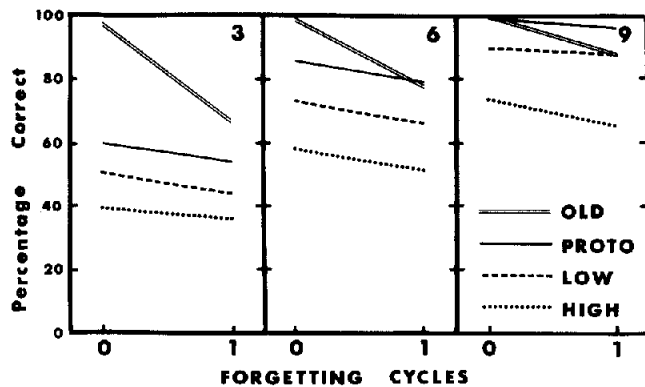
Figure 6. Correct classification percentages when secondary memory [SM] contains six copies of each exemplar. (All other parameters are the same as Figure 5. Each point is based on $N = 400$.)

stantially reduced by increasing the number of traces per exemplar. (In the present simulation this parameter was set at 1.)

Finally, the classification of random stimuli into the three categories increased with category size, as it usually does with human subjects (Homa et al., 1973, 1981). Of the random patterns tested ($N = 1,800$, collapsing over forgetting cycles), 12% were assigned to the junk category, and 22%, 31%, and 34% to the 3-, 6-, and 9-exemplar categories, respectively. The reason for this category–size gradient is obvious: The more exemplars of a category are stored, the more likely it is that one or more of them will share several features with a randomly generated probe. In MINERVA 2, random stimuli are more likely to be erroneously placed in large categories than in small ones for essentially the same reason that members of large categories are more often correctly classified. In both cases, the more traces there are belonging to a category, the more likely it is that one or more will generalize to the probe and influence its classification.

There is evidence that the tendency to place random stimuli in larger categories can be reduced (Hartley & Homa, 1981) or eliminated (Omohundro, 1981) by forewarning subjects that the categories will be used equally often on the test. This has been interpreted as an indication that the usual finding represents a response bias that this special instruction overrides. However, the instruction could induce a bias (to use the categories about equally often) rather than eliminating one. If the instruction does eliminate a bias, then prototype-abstraction theory should predict that subjects will assign fewer random stimuli to large categories than to small ones, because the large category representations are supposed to be better defined. This result does not seem to have been obtained.

Figure 6 shows the results of a replication of the previous simulation ($N = 400$), with two changes: Each exemplar was represented in SM by six traces rather than just one, and classification of random stimuli was not done. The differential forgetting result was again obtained, as were the category-size and typicality effects. However, observed forgetting rates were lower than in the previous simulation, especially for the largest category, and as a consequence Figure 6 resembles human data more closely than does Figure 5. Comparison of the two figures shows that the observed forgetting rate depends not only on the

underlying parameter F but also on the number of traces in memory. Therefore, the model predicts an overlearning effect: Observed forgetting rates in classification tasks should decrease with increases in the number of training trials.

Prototype Similarity

As categories become more similar to one another, classification learning becomes more difficult (Robbins et al., 1978), and transfer to new exemplars suffers (Goldman & Homa, 1977). In Experiment 1 of the Robbins et al. (1978) study, intercategory similarity was varied by manipulating similarities among the category prototypes. The original simulation described above used category prototypes that were always independently generated or orthogonal, corresponding to the "unrelated" condition of the Robbins experiment. For comparison with Figure 5, basically the same simulation was repeated using similar prototypes. This was done by first generating a master prototype and then producing category prototypes by changing exactly 3 of the 13 master-prototype features, chosen at random. High-level distortions of the three prototypes were then encoded in SM as before, and then each of the 24 conditions represented in Figure 5 was tested. A total of 500 subjects were simulated.

Although the pattern of classification accuracy across conditions was qualitatively the same as that shown in Figure 5, the absolute level of performance was lower. This was true for 12 of the 12 conditions before the forgetting cycle and for 10 of the 12 conditions after (in the 2 exceptions, performance in the similar-prototype run was virtually tied with that in the orthogonal-prototype run). Overall, classification accuracy was 66% when prototypes were orthogonal and 57% when they were similar. Between-category similarity thus makes categorization more difficult for MINERVA 2, as it does for human subjects.

Category Breadth

Posner and Keele (1968) trained one group of subjects to classify medium-level distortions of prototypes and another to classify low-level distortions. The group that had learned with the more extreme distortions showed better transfer to new items. In a similar study by Homa and Vosburgh (1976), two groups of subjects learned to classify patterns into three categories, having 3, 6, and 9 exemplars, respectively. One group was trained using mixed distortion levels (high, medium, and low) and another using only low-level distortions. Transfer testing showed an interaction between group and category size: Subjects learning low-level distortions displayed better transfer on the 3-exemplar category, and those learning with mixed-level distortions showed better transfer on the larger categories. Both studies have been interpreted as showing that categories are better learned when training experiences are highly variable than when variability is low.

The MINERVA 2 model was applied to the category breadth problem by repeating the simulation shown in Figure 5 in all details except that low-level distortions of the three prototypes were stored in SM. Classification was better than that shown in Figure 5 (where high-level distortions were used). This is the opposite of the outcome that summaries in the literature frequently report (e.g., Homa et al., 1981, p. 420).

However, this may not be the failure that it at first appears to be. In both the Posner and Keele (1968) and the Homa and Vosburgh (1976) studies, original training was carried out until a performance criterion was reached, and subjects learning with more extreme distortions required many more trials to reach the criterion than did those in the low-level distortion groups. However sensible a performance criterion may seem from some theoretical perspectives, from that of a multiple-trace theory it is clearly inappropriate. One cannot match two groups on degree of learning by matching their classification performance when the respective difficulties of the classification tests are not the same. This requirement is clearly violated when within-category generalization is higher for one group than for the other, as was true during the learning trials in these studies. The issue is crucial because, as was shown above, the number of training trials affects the model's transfer performance, particularly if many trace features are missing (e.g., if either $F > 0$ or $L < 1$). This suggests that the reported positive effect of category breadth on transfer may be an artifact, owing to the confounding of breadth with number of training trials.

In accordance with this argument, data published recently by Homa and Cultice (1984, Figure 2) show clearly that when number of training trials is held constant, subjects trained with low-level distortions show better transfer—not worse—than those trained with high-level distortions. To complicate matters still further, Peterson, Meagher, Chait, and Gillie (1973) conducted four experiments in which subjects were trained to criterion before testing, and consistently found the opposite of the Posner and Keele (1968) result. Homa and Vosburgh (1976) attributed this to the small (3-exemplar) category size that Peterson et al. (1973) used but failed to mention that this was the same category size used by Posner and Keele (1968).

As a test of the model's ability to account for the range of findings regarding category breadth, four simulations were done. In three of them, one trace of each exemplar was stored in SM, and the training exemplars were high-, medium- (3 features changed), or low-level distortions of the category prototypes (Ns = 400, 400, and 600, respectively). A lower learning rate was used than previously (L = .50). As before, there were three category sizes, and classification was tested for one exemplar, the prototype, and one high- and one low-level distortion from each category. The results are shown in the H-1, M-1, and L-1 panels of Figure 7. These results compare favorably with those of Homa and Cultice (1984), displaying a clear trend toward better classification performance with decreasing training-exemplar breadth.

The fourth simulation was carried out to determine the effect of extended training. Ten copies of each medium-level training pattern were stored in SM (N = 300). The data are shown in panel M-10 of Figure 7. In comparison with the M-1 data, classification was better in every condition, demonstrating the redundancy advantage of 9 additional traces of each exemplar. More importantly, a comparison of the M-10 and L-1 curves shows that M-10 performance was better than L-1 in 9 of the 12 test conditions, including all 6 comparisons involving transfer to new low- and high-level distortions. The conclusion is clear: The memory theory underlying MINERVA 2 predicts better transfer after learning with low-level than with higher level distortions when number of trials is held constant, but if extra
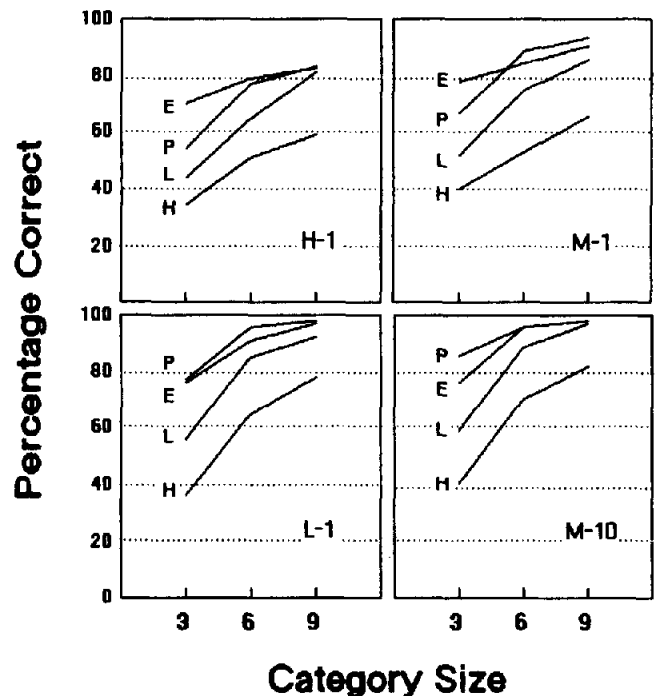

Figure 7. Classification performance after learning with one copy of each high-level distortion (H-1), medium-level distortion (M-1), and low-level distortion (L-1), or 10 copies of each medium-level distortion (M-10). (L = .50.)

training is given in the latter condition, this relationship may be reversed. The model does not account for every detail of the somewhat contradictory findings on category breadth but does appear generally consistent with the major results.

Distortions of Exemplars

Homa et al. (1981) trained subjects to classify patterns into categories of size 5, 10, and 20, and then tested for transfer to new stimuli that were distortions, not of the category prototypes but of the training exemplars themselves. Transfer to these stimuli declined with the increasing distortion level of the test item and improved with increasing category size. The effect of distortion level was less for the larger categories than for the smallest, however. This interaction was interpreted as showing that as the category becomes larger, individual stored exemplars play a decreasing role in classification and an abstracted representation of the protoype has an increasingly powerful effect. Homa et al. (1981) tested and rejected three exemplar models, but none included the assumption of MINERVA 2: that classification is determined by all SM traces in concert, with the contribution of each weighted according to its similarity to the probe.

To investigate the effects of category size on MINERVA 2's classification of distortions of exemplars, a simulation was done using three categories of 3, 7, and 15 high-level distortions. Three copies of each exemplar were stored in SM using L = 1. Testing was carried out with one old exemplar from each cate-
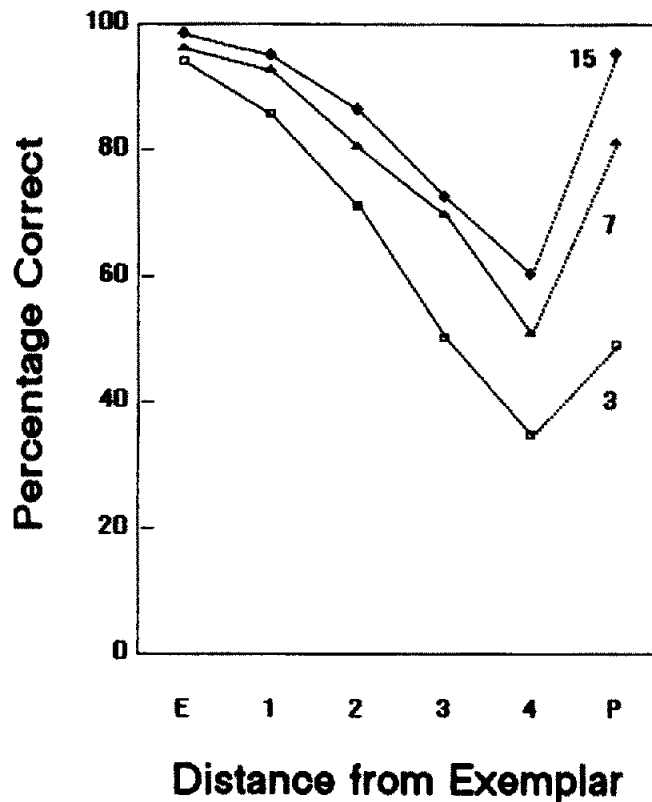
Figure 8. Effects of category size and probe type on classification of distortions of the training exemplars. (Three copies of each exemplar were stored.)

gory, and distortions of that exemplar produced by changing exactly 1, 2, 3, or 4 of its 13 features. The prototype of each category, which was a distance of 4 from each stored exemplar, was also classified. The model's performance is shown in Figure 8 ($N = 600$), which appears similar to graphs published by Homa et al. (1981). In particular, it shows the same interaction between distance from the exemplar and category size. Thus, even though it is an exemplar model, MINERVA 2 reproduces the essentials of the Homa et al. (1981) results.

Context Specificity

According to the present theory, concepts do not have unitary representations. The information retrieved from SM by a word, for example, is a function of the particular subset of episodic traces activated by the probe, and the probe consists not only of the word but also of its context. The crucial dependence of a word's meaning on context has proved difficult for theories of semantic memory positing unitary concept representations to explain.

A recent investigation by Roth and Shoben (1983, Experiment 2a) illustrates the problem in terms of the study of natural categories. Roth and Shoben had subjects rate the typicalities of category members, for example, the beverages coffee, milk, and tea. In the absence of a biasing context, subjects rated coffee as a particularly good example of a beverage, tea as less so, and

milk even less. This ranking might be explained by assuming that coffee is a nearly prototypical beverage and that tea is seen as a better example of a beverage than milk because it is more similar to the prototype. But Roth and Shoben (1983) showed that by manipulating sentential context, one can reorder typicality rankings in ways that the prototype model cannot easily explain. For example, after reading about two secretaries gossiping and drinking a beverage during their midmorning break, subjects ranked the three beverages just as they did without a biasing context: Coffee was most typical, followed by tea and then by milk. But after reading about a truck driver starting his day by having a donut and beverage at his favorite truck stop, they ranked them in the order of coffee, then milk, and then tea. The latter ranking is inconsistent with the unbiased ordering, and it also violates similarity relations, because tea is rated as more similar to coffee than is milk.

To demonstrate an analogous effect in MINERVA 2, classification percentages rather than typicality ratings were used. The model was essentially unmodified from the one employed in the previously described simulation runs, except that four contextual features were added to each memory trace, in positions $j = 24 \ldots 27$. Two contexts were defined, with $T(i, 24) \ldots T(i, 27)$ alternating $+1, -1, +1, -1$ for context A and $-1, +1, -1, +1$ for context B. Five category names and prototypes were generated. For each category, two SM traces of the prototype pattern (analogous to coffee) were combined with the category name (analogous to the label *beverage*) and stored in SM. One trace included features of context A and the other features of context B. Next, a low-level distortion (analogous to tea) was generated by changing two randomly chosen features of the prototype. This pattern was combined with the category name and encoded in a single trace that incorporated the features of context A. Finally, a high-level distortion was generated by changing four features of the prototype, combined with the category name and encoded in a single trace that included the features of context B. Thus there were four SM traces for each of five categories: one of the prototype and one of the low-level distortion in category A, and one of the prototype and one of the high-level distortion in category B.

In classification testing, 18 probe types—all with name features set at zero—were used for each category. Nine were generated by combining the prototype and the original low- and high-level distortions with either no context (i.e., the four context features set at zero), context A, or context B. The remaining 9 probes were generated by producing low-level distortions of each of the three original patterns (i.e., similar to prototype, similar to low-level distortion, and similar to high-level distortion) and combining them with no context, context A, and context B. After a cycle of 50% forgetting, testing was repeated. The entire procedure just described was repeated 80 times, for 80 simulated subjects. Correct classification percentages from the first test are presented in Table 1. (Because the same general pattern across conditions was found after forgetting, data from the second test are not shown.) Each entry is based on 400 observations.

The result was the same whether testing was done with the stored patterns (top half) or with distortions of the stored patterns (bottom half), although in the former case the outcome is somewhat obscured by ceiling effects. When there were no

Table 1

*Correct Classification Percentages for Probes
as a Function of Stimulus Type and Context*

Stimulus type	Context		
	None	A	B
Prototype	99.5	99.3	99.5
Low distortion	98.5	98.8	94.8
High distortion	98.3	76.0	99.0
Similar to prototype	94.5	94.0	90.8
Similar to low distortion	88.0	89.8	68.3
Similar to high distortion	79.5	45.0	87.5

context cues, classification was ordered according to similarity to the prototype. In context A, which favors the same ordering (e.g., tea being more common in that context than milk), the ranking was preserved. But in context B, where the test item that is less similar to the prototype is more common, the rankings of the two nonprototypical patterns were reversed. This is analogous to the context specificity demonstrated in the Roth and Shoben (1983) study. It should be noted that classification was not determined solely by traces having the appropriate contextual features, if it were, then there would be no generalization across contexts, and correct classification of items tested in the inappropriate context would be near 20%, or chance—substantially lower than the values that were obtained.

Similarity Among Exemplars

As a further challenge to MINERVA 2, we consider a finding that Elio and Anderson (1981) interpret as inconsistent with exemplar theories of classification learning. Reduced to its bare essentials, the result can be illustrated by a comparison of two conditions. In the control (no-similarity) condition, a probe is to be classified whose stimulus features can be represented as AB. Two exemplars of the target category have been learned: AC and DB. The probe matches both exemplars equally well (on A and on B), but the exemplars bear no similarity to each other. In the experimental (generalize) condition, the probe that must be classified is EF, and the two stored exemplars are EG and EH. Again, the probe matches both exemplars equally well (on E in both cases), but the two exemplars also match each other on the same feature(s) that they share with the probe.

The empirical question is whether classification of the probe is better in the generalize condition than in the no-similarity condition. If it is, Elio and Anderson (1981) argued, then performance in the generalize condition must be based on some additional information not present in the no-similarity control, presumably a representation of the generalization "stimuli of the form E__ belong in category 1." Elio and Anderson (1981) performed three appropriate experiments using elaborately constructed stimulus sets (brief descriptions of individual members of two fictional social clubs). In all three experiments, performance in the generalize condition exceeded that in the control.

Rather than attempting to represent the Elio and Anderson experimental materials in a form suitable for MINERVA 2, I have tried to aim at what seems to be the basic issue. Eight random

prototypes and their category names were generated, using features $j = 1 \ldots 10$ for names and $j = 11 \ldots 22$ for stimulus patterns. Two exemplars were then generated from each prototype by multiplying 3 of the 12 stimulus features by -1. In the generalize condition, both exemplars were created by randomly selecting the 3 features to be changed from the same set, $j = 11 \ldots 16$. In the no-similarity, or control, condition, one exemplar was generated by changing 3 features from the set $j = 11 \ldots 16$, and the other by changing 3 in the set $j = 17 \ldots 22$. Thus, in the control condition the two exemplars were orthogonal, matching on exactly 6 of the 12 stimulus attributes. In the generalize condition, the two exemplars matched on between 6 and 12 of the features, the expected number of matches being 9.

Half of the eight categories were assigned to the generalize condition and half to the control. With two exemplars per category, there were 16 SM traces altogether, each an encoding of a name–exemplar pair. The eight prototypes were then used as probes, and as before, category assignments were based on correlations of the name features of the echo, $C(1) \ldots C(10)$, with the eight category names. The echo–probe conversion routine, described earlier, was then invoked twice. Echo 1, as a probe, yielded Echo 2, and Echo 2 in turn became a probe for Echo 3. A total of 300 simulated subjects were run, for 1,200 observations in the generalize condition and 1,200 in the control.

Correct response percentages on the three successive echoes are shown in the top half of Table 2. For Echo 1, classification accuracy does not differ reliably between the generalize and control conditions ($z = 1.42, p > .15$). This result is just as one would expect, because the degree of activation of each trace depends solely on its similarity to the probe, and similarity to the probe was exactly the same (9 of 12 features) for every trace in both conditions. Thus far, the simulation confirms Elio and Anderson's (1981) conclusion that exemplar theories do not predict their generalization result.

A different picture emerges, however, when the echo–probe conversion process is allowed to operate. On both Echo 2 and Echo 3, classification in the generalize condition was better than in the control ($zs = 5.53$ and 7.95, respectively). The reason is that echo–probe conversion allows activated traces to influence one another, and they do so according to their similarity. Traces in the same category match each other on all 10 name features and half of the 12 stimulus features if they are in the control condition (16/22 altogether), and on all 10 name features and an average of 9/12 of the stimulus features if they are in the

Table 2

*Correct Classification Percentages Over Three Successive
Echoes for Two Models of the Generalization Result*

Model and condition	Echo		
	1	2	3
Conversion			
Generalize	70.9	72.4	72.1
Control	68.2	61.9	57.0
Resonance			
Generalize	70.0	70.3	70.0
Control	68.9	64.7	60.3

generalize condition (an average of 19/22). The cumulative effect of this difference increases across successive echoes, as Table 2 shows.

There is another way in which traces might interact, and it deserves some attention because it is suggested by the resonance metaphor that was introduced earlier. The idea is that activation among similar traces might be mutually reinforcing, much as is activity among tuning forks of the same pitch. To explore the implications of such a process, the echo–probe conversion procedure was deleted from the program, and a procedure was added that allowed the activation produced by the probe to spread among traces directly, according to their pairwise similarities.

Specifically, let $A(k)$ represent the activation of trace k by the probe (at Time 1), and let $B(i)$ be the activation of trace i by all SM traces (at Time 2). Then,

$$B(i) = \left\{ (1/n) \sum_{k=1}^{m} \left[A(k) \sum_{j=1}^{n} T(k,j)T(i,j) \right] \right\}^3. \quad (5)$$

In Equation 3, $B(i)$ can now be substituted for $A(i)$, to determine the content of the echo resulting from trace interaction.

With this procedure added to the program, the generalization task was simulated for 150 subjects, or 600 observations per condition. Again there were three classification attempts following each probe: one using the echo produced by the probe itself, one using the echo from the ensuing trace interaction described by Equation 5, and a third obtained by applying the interaction procedure yet again. Correct classification percentages are shown in the bottom half of Table 2. Data from this resonance model followed the same pattern as those from the conversion model: the generalize and control conditions did not differ reliably on Echo 1 but did on Echo 2 ($z = 2.07$) and Echo 3 ($z = 3.59$). Further, data from the two models were quantitatively alike: None of the six pairs of corresponding values from the two models differed significantly (all zs < 1.43).

It is no accident that the generalization result emerges from both the echo–probe conversion model and the intertrace resonance model, or that the two are quantitatively indistinguishable. In the former model, Echo 1 represents the summed effects of all traces in SM, and it is the similarity of each trace to this aggregate pattern that determines how much activation it feeds into Echo 2. In the latter model, the activation that each trace feeds into the second echo is determined by that trace's similarity to all activated SM traces (including itself). Although the second method of computing the contribution of a trace to the echo requires many more computations than the first, the two produce the same results and are, in fact, formally equivalent (see Appendix).

One implication of this equivalence is that the ability of MINERVA 2 to bootstrap its way out of the ambiguous recall problem (see Figure 4) does not rest solely on the echo–probe conversion procedure. Although it has not been demonstrated, the intertrace resonance version of the model should accomplish the same result.

Discussion

The aim of this work was to explore the possibility that an episodic memory theory might be able to explain how abstract concepts are acquired from examples and how the concepts are represented, without appealing to a separate generic memory system operating according to special rules. In applying MINERVA 2 to the schema-abstraction task, we have seen that a number of the most salient findings from the experimental literature are easily accounted for by the model. Among these findings are several that have been interpreted as evidence against an exemplar approach to categories, even though, as a multiple-trace model, MINERVA 2 represents an extreme form of the exemplar view.

It is important to note that, although traces of abstractions were not involved in these simulations, the present theory does not deny that abstract representations can be stored. Echoes, as well as external stimuli, can contribute to the contents of PM, and as Figure 4 shows, the echo triggered by a category name can be an abstraction. If such an abstraction in PM were made the object of conscious reflection, a trace of the resulting experience would itself be stored in SM. Although such a process was not employed here, nothing in the theory forbids it. Indeed, it would be highly implausible to assume that abstract ideas, as retrieved in the echo, are never encoded in SM. Thus, an ironic side benefit of an explanation of concept learning that does not require abstract memory representations is an explanation of how such representations could be learned.

Even though the present theory allows abstract memory representations, there are good reasons for attempting to simulate schema-abstraction results without such representations playing a role. First, abstractions derived from one's direct experience may be encoded in memory only rarely—for example, in a situation in which one attempts to define a term or to describe or imagine a typical member of a category. Because the purpose of most schema-abstraction experiments has been to investigate a presumed automatic abstraction process, the instructions are typically intended to discourage such attempts. Thus, it has been assumed here that the sort of deliberate reflection that would be necessary to formulate and store an abstraction in SM is not common in these experiments.

Second, even if abstract traces were formed during training, there is reason to doubt that they would play an important role in classification. According to the present theoretical perspective, a trace of an abstraction would be just another episodic memory trace (e.g., a record of an attempt to define a concept). Such a trace would be strongly activated by a future attempt at the same activity and could aid the attempt considerably, but it would be unlikely to be strongly activated by a category member requiring classification. The exemplar would be classified primarily by analogy with past instances—"nonanalytically," to use the term of Jacoby and Brooks (1984)—whether a trace of the abstraction episode existed or not.

Third, even a theorist who believes that abstract representations play a central role in the schema-abstraction task must be concerned to know which phenomena require them for their explanation. This may best be determined by a theoretical exercise in which one attempts to get along without abstract traces. From the residual phenomena not explained by the exercise, one should get a hint as to what kind of abstraction process needs to be added. A conclusion of the present study is that no clear examples of such residual phenomena have been uncov-

ered so far, even among very likely candidates, and so their number may be surprisingly small.

Defined more broadly than a particular experimental task, abstraction has been the focus of considerable recent attention in cognitive science. Semantic memory theorists in particular have encountered two difficult problems: One is to explain how the abstract knowledge that is assumed to be stored explicitly in semantic memory was learned originally and how it is modified by experience. The other is to explain the flexibility and context dependency that characterize the human use of concepts.

Rumelhart and Norman (1978) confronted the learning problem by proposing three different kinds of learning, each appropriate to a different relationship between new input and the abstract structures (schemata) with which the input is processed. In accretion, the new experience can be handled by existing schemata without change; in schema tuning, only minor modifications or adjustments of the existing structures are required; and in restructuring, new schemata must be specifically designed to deal with new discrepant input. This proposal vaguely suggests a powerful executive routine that is able to identify and keep track of processing failures, diagnose their causes, and infer the nature and amount of tinkering with existing structures that will be necessary to insure that the failures do not happen again. Rumelhart and Norman (1978) do not specify how all this might be done.

In a more ambitious vein, Schank (1982) has attempted to spell out in some detail how processing failures might trigger particular kinds of memory modifications. The key to understanding memory organization, Schank argued, is understanding how a memory stored in one structure (e.g, a structure concerning airplanes) can remind us of another memory that should be stored in a different structure (e.g., one concerning restaurants). He proposed that memory is continually reorganizing itself on demand, by creating, modifying, and abandoning abstract structures that serve to cross-classify experiences in various ways.

The difficulty of explaining how abstract knowledge is acquired, apparent in the efforts of Rumelhart and Norman (1978) and Schank (1982), suggests that it might be worthwhile to try a different approach. The view offered here is that, to the extent that abstract knowledge as such is stored in memory, it has no special status or function. All experiences to which one attends are encoded as episodic traces, whether they violate one's expectations or not (although a suprising event may command more attention than one that is routine). A new experience never modifies an old memory trace. It modifies memory as a whole, in the sense that the addition of a trace can alter the behavior of the system. The more the new trace differs from previous ones, the more obvious the behavior change will be, but even dramatic differences in the degree to which behavior has changed do not imply different kinds of structural change or learning. Because information is abstracted from concrete experiences at the time of retrieval rather than during learning, no sophisticated executive routine is needed to decide when and how to tune, reorganize, or abandon memory structures. Reminding is not confined to predetermined structures, and changes in behavior follow automatically from the indiscriminate accumulation of new episodic traces in memory.

The second difficult problem for current theories of semantic memory has been to explain the apparent flexibility of retrieval. Concepts appear to be too fluid and context-dependent to be represented by fixed structures in fixed hierarchies, as is commonly assumed. The Roth and Shoben (1983) study discussed earlier showed that the typicalities of category members can be reordered by a change in sentential context. The flexibility with which context can alter a word's interpretation led R. Anderson and Ortony (1975) to argue that "only in a vague, abstract sense could words be said to have fixed meanings" (p. 168). Even a common verb such as eat, though appearing to have a simple core meaning, has very different implications depending on the subject (e.g., termite, princess, snake) and object (crumb, pencil, melon) with which it appears. Understanding a sentence cannot be simply a matter of combining the dictionary meanings of the words; somehow, knowledge about the world is used to invoke the specific meanings that are appropriate in a particular context.

The MINERVA 2 model obviously cannot simulate sentence understanding, but it does suggest a partial answer to the problem of polysemy. If there are no fixed concept representatives in memory, but only traces of episodes in which the concept name has been used, then it is understandable that the meaning retrieved by a word should be highly context dependent. The word eat could be said to be represented by a very large number of traces that can be activated in parallel, but not all of these traces will be strongly activated in any one encounter with the word. Any particular active subset of these traces is a biased sample from the population, and the active subset will vary in both size and content from one occasion to another, depending on the context in which the word appears.

Several other problems relating to language use take on a different perspective when viewed in this way. Homonymy is seen as an extreme case of polysemy, in which a word has two or more dissociated populations of traces, separated by a boundary that active-trace samples seldom or never cross. It can be seen how essentially the same meaning could be expressed in several different ways, depending on context, and how two different languages might carve up experience along somewhat different lines. And idiomatic expressions need not seem anomalous if one assumes that the retrieved meaning of a word, being highly context dependent, can be constrained by other words with which it co-occurs.

Evidence that a context-specific meaning of a word is sometimes retrieved directly during comprehension, rather than having to be constructed, was provided by Potter and Faulconer (1979). Their subjects listened to sentences and were occasionally interrupted to confirm whether or not drawings depicted objects that the sentences named. For example, immediately following a noun phrase such as "burning house," a picture appeared that represented either the noun alone (house) or the noun phrase (burning house). The picture matching the noun phrase was verified more quickly than the one matching the unmodified noun, even though it was less typical of the general category (e.g., house). Most interesting for our purposes is that the size of this effect was a function of the familiarity of the noun phrase: It was much greater when the phrase was familiar (e.g., roasted turkey) than when it was unfamiliar (e.g., broken screwdriver). This is what should occur if familiar noun phrases

directly "look up" their meanings by activating appropriate memory traces, making it unnecessary to construct their meanings anew each time they are understood.

Much current thinking about how abstract information is represented and retrieved has been influenced by schema theory, which was originally applied to memory by Bartlett (1932). Although the notion is notoriously difficult to pin down, a schema is generally held to be a kind of holistic, coherent knowledge structure that determines expectancies, organizes encoding, and systematically distorts retrieval in the direction of internal consistency. Two excellent reviews of schema theory are available, by Alba and Hasher (1983) and by Brewer and Nakamura (1984). Most modern schema theorists appear to assume that a schema is a data structure that has its own unitary representation in memory. However, it is not clear that this is what Bartlett (1932) had in mind in describing schemas as "masses of organised past experiences" (pp. 197–198). Although he did argue against enduring traces of individual events, his emphasis on the active, momentary nature of schemas suggests something like the echo of MINERVA 2.

Because even advocates of schema theory (e.g., Brewer & Nakamura, 1984) admit that Bartlett was too hasty in dismissing trace theory, perhaps it would not be too heretical to suggest that schemas are based on traces. Thus, rather than an abstraction that is stored in memory, a schema may be a temporary, dynamic structure that springs into being when a retrieval cue occurs. It could then summarize information in the memory traces activated by the cue, fill in uncertain details, and dictate expectancies—and it would continually change as the experience unfolds. A modification of schema theory along these lines would have the advantage of explaining the relationship between traces and schemas and would avoid the vexing question of how schemas are learned.[4]

Comparison With Other Exemplar Models

Exemplar theories of concept representation range from the moderate position that abstract concepts such as *bird* might be represented by somewhat less abstract concepts such as *bluejay, robin,* and so forth (Smith & Medin, 1981), through the intermediate view that concepts are represented by individuals (Brooks, 1978), to the present view that individuals as well as more abstract concepts are represented by traces of many episodes. Jacoby's arguments that repetition effects in perceptual enhancement are based on episodic traces rather than threshold changes in logogens (e.g., Jacoby, 1983a, 1983b; Jacoby & Brooks, 1984) are generally consistent with the present position, in that logogens are assumed to be generic representations of individual words (Morton, 1969).

Two other models of the classification process that have taken an extreme multiple-trace view were presented by Restle (1961) and Hintzman and Ludlam (1980). Restle assumed that each training trial leaves behind its own trace and that classification is done by sampling randomly from the set of traces until one matching the test item is found. The test item is assigned to the same category as the matching trace. The retrieval assumptions of this model obviously differ from the parallel, similarity-based mechanism of MINERVA 2.

The predecessor of MINERVA 2 was MINERVA 1 (Hintzman

& Ludlam, 1980). That model represented stimuli as complex propositional structures rather than as feature lists, and it had nothing corresponding to the echo of MINERVA 2. It did not combine information from activated traces, but simply classified stimuli according to a nearest-neighbor rule. The MINERVA 1 model was incapable of retrieving from memory a category name or an abstract representation that had not explicitly been stored. As was discussed earlier, MINERVA 1 and MINERVA 2 both predict the differential forgetting of prototypes and old instances. Apparently the result does not require a strict nearest-neighbor classification rule: The differential weighting of traces according to similarity to the probe approximates the rule closely enough to produce the same effect.

The context model of Medin and Schaffer (1978) is not presented as a multiple-trace model, although the authors do suggest that it could be modified to deal with repetition in this way (Medin & Schaffer, 1978, p. 232). In the context model, as in MINERVA 2, classification is determined by all stored exemplars of each category, each weighted according to its similarity to the probe. Similarity is computed across features by multiplying values that represent the degree of congruence between probe and exemplar on each feature. The values range from 1 (identical) to 0 (completely different). To take a simple three-feature example, if the probe and exemplar are identical on the first feature and have similarity .3 on the second feature and .5 on the third, the similarity of the probe to the exemplar is $1 \times .3 \times .5 = .15$. The effect of this multiplication rule is that similarity is a positively accelerated function of the proportion of features that match. In MINERVA 2, the similarity function $S(i)$ is linear with the proportion of matching features, but the activation function $A(i)$ is positively accelerated. Thus both models imply generalization gradients that are concave upward, so that the influence of an exemplar drops off rapidly with decreasing similarity to the probe. Further, in both models, classification is determined by summing the effects of all exemplars. The two models differ in many particulars, including the way that features are defined, but as applied to classification tasks their predictions may be practically the same.

McClelland (1981) described an exemplar model that, although not a multiple-trace model, resembles MINERVA 2 in several ways. First, classification is determined by the exemplars in the neighborhood of the test item, acting in parallel. Activation of a trace is again a positively accelerated function of similarity to the probe, although in this case the underlying mechanism is lateral inhibition. Second, the retrieved information can include not only a category name but also other properties of the stored exemplars. The model can retrieve a description of a typical category member, for example, by summing information across traces, each weighted according to its activation level. Third, any combination of features can be used as a probe, to retrieve the other features with which the combina-

---

[4] Of course, the simple feature lists used in MINERVA 2 do not capture the kinds of complex structure that the representation of everyday knowledge requires, and so the model cannot be applied directly to the specific problems with which semantic memory theorists and schema theorists have been concerned. It remains to be demonstrated that abstraction at retrieval is possible with structured information (e.g., information coded as in MINERVA 1 [Hintzman & Ludlam, 1980]).

tion was most commonly associated. McClelland's model thus shares with MINERVA 2 the ability to create concepts at the time of retrieval by combining traces in novel ways.

Comparison With Nonexemplar Models

A nonexemplar model that has been successfully applied to classification learning is the property-set model of Hayes-Roth and Hayes-Roth (1977). Like earlier models of Neumann (1974) and Reitman and Bower (1973), it proposes that what is stored is a frequency count of each property in each category, where the set of properties is the powerset of the features (i.e., the set of all possible sets of features). Thus, if an exemplar of Category A is presented that includes Features 1, 2, and 3, seven Category-A counts are incremented: one each for Properties 1, 2, 3, 1 + 2, 1 + 3, 2 + 3, and 1 + 2 + 3. Classification of a test item is performed by summing, over each of the item's properties, the number of times the property occurred in each category, and then choosing the category giving the highest total.

Like the context model and MINERVA 2, the property-set model produces a generalization gradient that is a positively accelerated function of the proportion of matching features. The reason is that the size of the powerset roughly doubles for every unit increase in number of features. By emphasizing the unity of the individual trace, MINERVA 2 and the context model weigh conjunctions of features heavily, and this is done explicitly by the Hayes-Roth model. Medin and Schaffer (1978) commented that, as applied to the classification task, property-set models make many of the same qualitative predictions as the context model, and as we have seen the context model is very similar to MINERVA 2. Among these competitors, then, the primary advantage of MINERVA 2 may lie in its applicability to nonclassification tasks.

The MINERVA 2 model may also be compared to composite-trace holographic and associative matrix models, at least three of which have now been applied to the basic schema-abstraction task (Eich, 1982; Knapp & Anderson, 1984; McClelland & Rumelhart, 1985). Eich's model, CHARM, produces the typicality effect and category size effect. An attempt to reproduce the differential forgetting of prototypes and old exemplars, however, failed (Eich, 1982). Another apparent difference between CHARM and MINERVA 2 is seen in the ability of the present model to deal with the ambiguous recall problem (see Figure 4), which Eich (1982) solved only by invoking a separate "semantic" memory system with the ability to identify the ambiguous patterns retrieved from the composite trace. Associative matrix models have likewise scored successes in accounting for some typicality and category size effects (Knapp & Anderson, 1984; McClelland & Rumelhart, 1985) but have not reproduced the range of empirical phenomena reported here. Accounts of differential forgetting are conspicuously absent from these reports. It is difficult to draw broad generalizations about the behavior of these models as a class, because many such models are possible (see Hinton & Anderson, 1981), they differ in important ways (e.g., Pike, 1984), and only a few have been tried on categorization tasks.

The present model is at the opposite extreme from holographic and associative matrix models in one respect: MINERVA 2 assumes that each experience is represented by a separate

memory trace, whereas the others postulate just one composite memory trace in which all experiences are stored. It has been argued that the effect of combining experiences at retrieval—as is done by MINERVA 2—is indistinguishable from that of combining them at encoding (Eich, 1982, pp. 654–655). However, a system that waits until the time of retrieval to combine experiences has more flexibility in how the combination is to be done. The MINERVA 2 model is able to activate traces highly similar to the probe much more strongly than traces moderately similar to it (Equation 2) because it computes the activation of each trace separately.[5]

As was suggested by the previous discussion, nonlinear generalization functions seem to be characteristic of the most successful models of the classification task. In fact, humans have little trouble with categories that systems employing linear generalization functions cannot learn (Medin & Schwanenflugel, 1981).[6] The CHARM model, like most of the matrix models that have been discussed in the literature, generalizes from past experiences in direct proportion to the overlap between each experience and the retrieving probe (Eich, 1982, p. 635).

One way that a composite-trace model can escape this dilemma is to define some features (or properties) in a context-sensitive or relational way—that is, as conjunctions of more elementary features (Hinton, 1981; McClelland & Rumelhart, 1985). Unfortunately, there seems to be no principled way to limit the number of features that are to be so conjoined. The extreme form of such a model would be one that treats every member of the powerset of features as a separate input element, much as was done in the classification model of Hayes-Roth and Hayes-Roth (1977). Such a model would not have to combine traces at retrieval, because an element for every conceivable combination of features would have been predefined. Unfortunately, the number of such elements grows exponentially with the number of features. A less fantastic model can be imagined that makes use of only a few levels (e.g., binary and trinary) of conjunctions, or of some other reasonable-sized subset of the powerset. The behavior of such a model might be quite similar to that of MINERVA 2.

Comparison With Semon's Theory

The present attempt to account for episodic and generic memories with the same system is remarkably similar to the forgotten memory theory of Richard Semon (1909/1923; see also Schacter, Eich, & Tulving, 1978). Semon assumed that each experience leaves behind its own memory trace or "engram." A retrieval cue was assumed to communicate in parallel with all

---

[5] An interesting variation on the present model would be one in which the generalization function is under some degree of intentional control; for example, the exponent in Equation 2 could be set at 1 when broad generalization is desired and 5 when the retrieval task requires sharp discrimination.

[6] A related problem discussed by Hinton (1981) involves the learning of a set of triplets such as ACE, ADF, BDE, and BCF, in which one letter predicts another only in the context of the third. Such item sets cannot be learned by matrix models that encode only simple binary associations (A–C, etc.) and are learnable by MINERVA 2 only if the exponent used in Equation 2 is greater than 1.

traces, activating each according to its similarity to the cue. If several traces having common components were activated simultaneously, they were said to be in a state of *nondifferentiating homophony*—a kind of resonant state in which the common properties of the traces stand out and their distinctive properties are masked, so that what appears in consciousness is an abstraction, rather than the content of a particular memory trace. The term *homophony* suggests music sung or played in unison, and Semon illustrated the notion with a chorus of voices metaphor. Homophony played a central role in Semon's theory of memory, and an entire chapter of his book (Semon, 1909/1923, chap. 16) was devoted to its role in abstraction. Computation of the echo in MINERVA 2 can be seen as an information-processing analog of Semon's notion of homophony.

Among current memory theories, the only ones I know of that account for abstraction in this way are MINERVA 2 and McClelland's (1981) model—inventions that were independent of each other and of Semon's writings. Too kindly, perhaps, Semon (1909/1923, p. 279) gave partial credit to Thomas Huxley for independent invention of the idea. However, in the process described by Huxley (1881, pp. 94–95), abstraction takes place at the time of encoding, not at retrieval. Huxley offers as an analogy the method of composite portraiture, in which a photographic plate that has been partially exposed to a series of faces displays an image of a generic face. He undoubtedly derived this analogy from the photographic experiments of Francis Galton. Interestingly, it is in Galton's writings rather than Huxley's that one finds the germ of the notion of homophony. In a talk on composite portraiture given in April 1879, Galton proposed that abstract ideas may first arise during retrieval. "Whenever a single cause throws different groups of brain elements simultaneously into excitement," he said, "the result must be a blended memory" (Galton, 1883, p. 229).

Semon's theory differs from the one offered here in one important respect. Probably because he relied for his observations almost entirely on introspection, Semon seems to have restricted the contents of engrams to sensory features.[7] The present theory, however, assumes that there are many primitive characteristics of experience that are abstract, in that they are not tied to any particular sensations, but that are encoded in episodic traces much as sensory features are. This concession to rationalist philosophy seems necessary if all acquired knowledge is to reside in episodic memory. Schacter (1982) has offered speculations on the reasons that Semon's memory theory was ignored. Whatever the reasons were, the present simulations suggest that, perhaps with a few modifications, Semon's general approach to memory and abstraction deserves to be given a serious second look.

---

[7] Translated literally, the original title of Semon's book is *Mnemic Sensations.*

References

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93,* 203–231.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review, 84,* 413–451.

Anderson, R. C., & Ortony, A. (1975). On putting apples into bottles—a problem of polysemy. *Cognitive Psychology, 7,* 167–180.

Bartlett, F. C. (1932). *Remembering.* London: Cambridge University Press.

Bower, G. H. (1967). A multicomponent theory of the memory trace. In K. W. Spence, J. T. Spence, & N. H. Anderson (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 1, pp. 230–325). New York: Academic Press.

Brewer, W. F., & Nakamura, G. V. (1984). The nature and function of schemas. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (pp. 119–160). Hillsdale, NJ: Erlbaum.

Brooks, L. R. (1978). Non-analytic concept formation and memory for instances. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Erlbaum.

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review, 89,* 627–661.

Elio, R., & Anderson, J. R. (1981). The effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 397–417.

Fodor, J. A. (1981). *Representations.* Cambridge, MA: MIT Press.

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

Goldman, D., & Homa, D. (1977). Integrative and metric properties of abstracted information as a function of category discriminability, instance variability, and experience. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 375–385.

Hartley, J., & Homa, D. (1981). Abstraction of stylistic concepts. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 33–46.

Hayes-Roth, B., & Hayes-Roth, F. (1977). Concept learning and the recognition and classification of exemplars. *Journal of Verbal Learning and Verbal Behavior, 16,* 321–338.

Hinton, G. E. (1981). Implementing semantic networks in parallel hardware. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 161–187). Hillsdale, NJ: Erlbaum.

Hinton, G. E., & Anderson, J. A. (1981). *Parallel models of associative memory.* Hillsdale, NJ: Erlbaum.

Hintzman, D. L. (1976). Repetition and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 10, pp. 47–91). New York: Academic Press.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments and Computers, 16,* 96–101.

Hintzman, D. L. (1986). *Frequency judgments and recognition memory in a multiple-trace memory model* (Tech. Rep. No. 86-11). Eugene: University of Oregon, Cognitive Science Program.

Hintzman, D. L., & Block, R. A. (1971). Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology, 88,* 297–306.

Hintzman, D. L., Block, R. A., & Summers, J. J. (1973). Modality tags and memory for repetitions: Locus of the spacing effect. *Journal of Verbal Learning and Verbal Behavior, 12,* 229–238.

Hintzman, D. L., Grandy, C. A., & Gold, E. (1981). Memory for frequency: A comparison of two multiple-trace theories. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 231–240.

Hintzman, D. L., & Ludlam, G. (1980). Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. *Memory and Cognition, 8,* 378–382.

Hintzman, D. L., Nozawa, G., & Irmscher, M. (1982). Frequency as a nonpropositional attribute of memory. *Journal of Verbal Learning and Verbal Behavior, 21,* 127–141.

Homa, D., Cross, J., Cornell, D., Goldman, D., & Schwartz, S. (1973). Prototype abstraction and classification of new instances as a func-

tion of number of instances defining the prototype. *Journal of Experimental Psychology, 101,* 116–122.

Homa, D., & Cultice, J. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 83–94.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 418–439.

Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 322–330.

Huxley, T. H. (1881). *Hume.* London: Macmillan.

Jacoby, L. L. (1983a). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 21–38.

Jacoby, L. L. (1983b). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior, 22,* 485–508.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 18, pp. 1–47). New York: Academic Press.

James, W. (1890). *The principles of psychology.* New York: Holt.

Knapp, A. G., & Anderson, J. A. (1984). Theory of categorization based on distributed memory storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 616–637.

McClelland, J. L. (1981). Retrieving general and specific information from stored knowledge of specifics. *Proceedings of the Third Annual Conference of the Cognitive Science Society,* pp. 170–172.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General, 114,* 159–188.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 355–368.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76,* 165–178.

Neumann, P. G. (1974). An attribute frequency model for the abstraction of prototypes. *Memory and Cognition, 2,* 241–248.

Omohundro, J. (1981). Recognition vs. classification of ill-defined category exemplars. *Memory and Cognition, 9,* 324–331.

Peterson, M. J., Meagher, R. B., Jr., Chait, H., & Gillie, S. (1973). The abstraction and generalization of dot patterns. *Cognitive Psychology, 4,* 378–398.

Pike, R. (1984). Comparison of convolution and matrix distributed memory systems for associative recall and recognition. *Psychological Review, 91,* 281–294.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83,* 304–308.

Potter, M. C., & Faulconer, B. A. (1979). Understanding noun phrases. *Journal of Verbal Learning and Verbal Behavior, 18,* 509–521.

Pylyshyn, Z. W. (1985). *Computation and cognition* (2nd ed.). Cambridge, MA: MIT Press.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Reitman, J. S., & Bower, G. H. (1973). Storage and later recognition of exemplars of concepts. *Cognitive Psychology, 4,* 194–206.

Restle, F. (1961). *Psychology of judgment and choice.* New York: Wiley.

Robbins, D., Barresi, J., Compton, P., Furst, A., Russo, M., & Smith, M. A. (1978). The genesis and use of exemplar vs. prototype knowledge in abstract category learning. *Memory and Cognition, 6,* 473–480.

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

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural basis of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 491–502.

Roth, E. M., & Shoben, E. J. (1983). The effect of context on the structure of categories. *Cognitive Psychology, 15,* 346–378.

Rumelhart, D. E., & Norman, D. A. (1978). Accretion, tuning, and restructuring: Three modes of learning. In J. W. Cotton & R. L. Klatzky (Eds.), *Semantic factors in cognition* (pp. 37–53). Hillsdale, NJ: Erlbaum.

Schacter, D. L. (1982). *Stranger behind the engram.* Hillsdale, NJ: Erlbaum.

Schacter, D. L., Eich, J. E., & Tulving, E. (1978). Richard Semon's theory of memory. *Journal of Verbal Learning and Verbal Behavior, 17,* 721–743.

Schank, R. C. (1982). *Dynamic memory.* Cambridge, England: Cambridge University Press.

Semon, R. (1923). *Mnemic psychology* (B. Duffy, Trans.). London: George Allen & Unwin. (Original work published 1909)

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Strange, W., Kenney, T., Kessel, F. S., & Jenkins, J. J. (1970). Abstraction over time of prototypes from distortions of random dot patterns: A replication. *Journal of Experimental Psychology, 83,* 508–510.

Tulving, E. (1983). *Elements of episodic memory.* Oxford, England: Oxford University Press.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352.

Appendix

Equation 5 (see text) computes the contribution of each trace, i, to the echo resulting from intertrace resonance as the following:

$$B(i) = \left\{(1/n) \sum_{k=1}^{m} [A(k) \sum_{j=1}^{n} T(k,j)T(i,j)]\right\}^3,$$

where $A(k)$ is the activation of trace k caused by the probe. We want to demonstrate that Echo 2, produced by the echo–probe conversion process, yields the same outcome.

According to Equation 3 (see text), given $A(k)$, the value of feature j in Echo 1 will be as follows:

$$C(j) = \sum_{k=1}^{m} A(k)T(k,j).$$

These values are normalized by multiplying by a constant, g, and reentered as the secondary probe.

Activation of trace i by the secondary probe, from Equations 1 and 2 (see text), will be as follows:

$$D(i) = \left[(1/n) \sum_{j=1}^{n} gC(j)T(i,j)\right]^3.$$

Rearranging terms and substituting for $C(j)$ yields

$$D(i) = \left\{g(1/n) \sum_{j=1}^{n} [T(i,j) \sum_{i=1}^{m} A(k)T(k,j)]\right\}^3$$

$$= [g(1/n) \sum_{j=1}^{n} \sum_{k=1}^{m} T(i,j)A(k)T(k,j)]^3$$

$$= \left\{g(1/n) \sum_{k=1}^{m} [A(k) \sum_{j=1}^{n} T(i,j)T(k,j)]\right\}^3$$

$$= g^3 B(i).$$

Correlations with category names are not affected by the constant, g^3; therefore classifications based on $D(i)$ and $B(i)$ are the same.

Integrating Personality and Social Psychology: Call for Papers

The editors of the *Journal of Personality and Social Psychology* intend to publish a special issue devoted to papers demonstrating that social behavior is best understood by integrating the diverse concerns of the three sections of *JPSP*. Papers are invited that deal with particular substantive issues crossing the boundaries of individual differences, social cognition, and interpersonal relations. Papers should represent practical demonstrations that the diverse concerns of this journal belong together in a full understanding of social behavior. We seek previously unpublished contributions, primarily empirical studies, but we are also amenable to syntheses of long-term research programs and to innovative theoretical statements.

Contributions intended for the special issue should be sent to the guest editor:

John F. Kihlstrom, PhD
W. J. Brogden Psychology Building
University of Wisconsin
1202 West Johnson Street
Madison, Wisconsin 53706

Interested authors should send an abstract of their article to the guest editor by September 1, 1986 and plan to submit a completed manuscript by December 1, 1986, at which time the paper will become subject to the usual peer review process.

Journal of Personality and Social Psychology editors:

Attitudes and Social Cognition	*Interpersonal Relations and Group Processes*	*Personality Processes and Individual Differences*
Steven J. Sherman	Harry T. Reis	Irwin G. Sarason
Charles M. Judd	Norbert L. Kerr	Edward F. Diener
		Warren H. Jones