

## Category typicality effects in episodic memory: Testing models of distinctiveness

STEPHEN R. SCHMIDT

Middle Tennessee State University, Murfreesboro, Tennessee

Category typicality effects were investigated within the context of three models of distinctiveness: a univariate model, a fixed-multifeature model, and a weighted-multifeature model. High-typical, medium-typical, and atypical targets were embedded in lists containing a background set of medium- to high-typicality items. Atypical items were more poorly recalled than were medium- and high-typical items independently of list structure. In recognition, subjects who studied high-typical items had difficulty discriminating between high-typical items that were and were not presented as part of the list. However, item typicality had little effect on the recognition performance of subjects who did not study high-typical items. These findings were consistent with the weighted-multifeature model of distinctiveness.

Unusual, atypical, or distinctive events are generally believed to be better retained than more typical everyday phenomena. This basic idea has been the subject of a great deal of research in learning and memory, and several new theories or frameworks have been developed to explain the relation between distinctiveness and memory (Hunt & McDaniel, 1993; Neath, 1993a, 1993b; Schmidt, 1991). Unfortunately, distinctiveness is often defined intuitively and employed post hoc to explain the effects of some variable on memory. In the research presented below, I attempted to define distinctiveness more formally within several different theoretical frameworks. This analysis led to clear distinctions among theories of distinctiveness. The theories were then applied to the domains of category distinctiveness and typicality effects on memory. The result was a better understanding of theories of distinctiveness, and an empirical evaluation of typicality effects on memory.

In attempts to define distinctiveness, a useful analogy can be drawn between the perceptual salience of an object and the mnemonic salience of an event. Such an analogy was drawn by Koffka (1935) in his analysis of von Restorff's research. Koffka (following von Restorff) employed geometric shapes. For the present purpose, consider a collection of marbles including nine red marbles and one yellow marble. The yellow marble will stand out perceptually from the group. It "emerges at the first

glance, while the others form a fairly uniform aggregate in which no special member stands out by itself" (Koffka, 1935, p. 485). Koffka was interested in whether the laws governing perceptual organization applied to memory traces, and whether mnemonic salience could emerge in the absence of perceptual salience. These and other questions led memory researchers to vary the structure of experiences in attempts to discover the laws underlying mnemonic salience. One means of studying mnemonic salience in the absence of physical isolation is to isolate items semantically from the surrounding items (Hunt & Mitchell, 1982; Schmidt, 1985). Schmidt made items distinctive by embedding the names of several musical instruments in a list of city names. Memory for the musical instruments was then compared with memory for the same items contained in a list of all musical instruments. The conceptually isolated items were better recalled and recognized than the same items from homogeneous lists. The musical instruments, by analogy, were yellow marbles embedded in the city names, the red marbles.

Explanations for these effects of distinctiveness vary greatly, and often concern whether distinctiveness improves memory as a result of encoding processes, retrieval processes, or both (see Hunt & McDaniel, 1993; Riefer & Rouder, 1992; Schmidt, 1985, 1991). In this paper I am less concerned with such explanations than with the more fundamental issue of how one should define and operationalize *distinctiveness*. As we will see, different definitions of distinctiveness lead to different predictions prior to any consideration of memory mechanisms. The theoretical landscape becomes considerably more complex when varied definitions of distinctiveness are teamed with varied mechanisms by which distinctiveness influences memory.

One way to define the distinctiveness of an item was developed by Murdock (1960) and applied to a magnitude estimation task and the serial position curve in serial learning. This method was recently employed by Neath

---

Experiments 2 and 5 were presented in part at the Midwestern Psychological Association annual meeting, May 1992. Experiment 3 was presented in part at the annual meeting of the Psychonomic Society, November 1994. I wish to thank Constance R. Schmidt for her comments on early drafts of this paper. I also wish to thank the following students for their assistance in this research: Kelly Gilbertson, Carolyn F. Hunter, and Angela L. Donegan. Correspondence should be addressed to S. R. Schmidt, Psychology Department, Middle Tennessee State University, Murfreesboro, TN 37132 (e-mail: sschmidt@frank.mtsu.edu).

(1993a, 1993b; Neath & Knodler, 1994) in analyses of the serial position curves in free recall and recognition. In both the Murdock and Neath investigations, stimuli were employed that varied along the univariate and quantitative dimensions of magnitude and time. For future reference I will call this the “univariate model” of distinctiveness.

Within the univariate model, the distinctiveness of item  $k$  ( $\delta_k$ ) is defined as follows (from Neath, 1993b):

$$\delta_k \sum_{j=1}^n \sqrt{(d_k - d_j)^2}, \quad (1)$$

where  $d_j$  represents the values of the items along some dimension. To apply this analysis to our marble example, the log values of wavelength would serve as  $d$ s. The contrast of the yellow marble ( $d_k$ ) to the red marbles ( $d_j$ s), summed over the group of marbles, would yield a high distinctiveness score. Neath (1993b, p. 693) has argued that “this model could be applied to any stimuli lying along any ordered dimension (e.g., physical or semantic dimensions).” Category membership may be treated as a one-dimensional variable and mapped onto a scale of distinctiveness. Many researchers have argued that categories have fuzzy boundaries, making category membership a matter of degree (e.g., Rosch, 1975). Within this framework, one can construct a scale of “goodness of category membership,” with typical items at the conceptual core of the category, and atypical items near the fuzzy frontier of the category boundary.

The application of the Murdock model to category membership is illustrated in Table 1, using five members of the “bird” category. Murdock (1960) transformed raw physical values to a logarithmic scale, noting that the perception of physical values often follows a logarithmic function as described by the Weber–Fechner Law. In the

present case, ratings of typicality are a direct measure of perceived category membership rather than a physical quantity that we wish to map onto a perceptual dimension. Thus, the typicality ratings were used directly in Formula 1 as values of  $d$ . Percent distinctiveness scores were then calculated (following Murdock, 1960) by dividing  $\delta_k$  for each item by the sum of the  $\delta_k$ s across other items in the list. From these values one can see that the univariate model clearly predicts that if the word *turkey* were presented in the context of a list of typical birds, it should be distinctive. In other words, atypical category members should be salient within a memory representation containing primarily typical category members.

Further assumptions may be needed to map mnemonic salience onto memory performance. For example, Neath (1993a) argued that recall performance was a function of two factors: the distinctiveness of an item and the retrievability of the item given a specific retrieval cue. These two factors combined to provide an explanation of the transient effects of recency on free recall. Recency items were thought to be temporally distinctive. But, the recency effect depended on whether or not recency items shared contextual information with other items in the list. If we apply this analysis to category recall, recall should be a combined function of distinctiveness (as calculated from category typicality) and the extent to which category and contextual information are useful as retrieval cues. Recognition performance, in contrast, should be a direct function of category typicality, with recognition increasing as items become more atypical.

Another way to conceptualize distinctiveness is in terms of feature overlap (Eysenck, 1979). Each item may be represented by a fixed set of features, and a distinctive item is one that shares few features with surrounding items. Only a subset of features is encoded in a particular study or test presentation. Memory performance depends on the overlap between features encoded at study and features encoded on a recognition test. Distinctive items share few features with other items, and thus are easily identified on the test. In addition, a greater number of features may be encoded for distinctive items than for common items, further aiding recognition performance. Although the number of encoded features is not fixed in this model, the contribution of features to the definition of an item is fixed. For this reason, I will call this the “fixed-multifeature” approach to distinctiveness.

Returning to the marble analogy: Red and yellow marbles may be described by a list of features, with hue being the one feature that distinguishes them. Similarly, musical instruments share few features with cities, but share many features with other musical instruments. Thus, a musical instrument presented in the context of a list of cities is more distinctive than the same musical instrument presented in a list of other musical instruments. Typical category members share a great number of features as illustrated by the similarity among the typical birds *robin*, *sparrow*, and *blue jay*. Atypical items share few features with the conceptual core, as illustrated by the lack of similarity between *penguins* and the items in

**Table 1**  
Percent Item Distinctiveness for Each Item in a List  
Containing Three Background Items and a Target Item  
That Is Either Typical or Atypical

Item	Typicality	Percent Distinctiveness
List Containing the Typical Item <i>Sparrow</i> in a List of Typical Items		
robin	1.20	19.6
blue jay	1.31	22.2
canary	1.42	36.1
sparrow	1.18	22.2
List Containing the Atypical Item <i>Turkey</i> in a List of Typical Items		
robin	1.20	18.3
blue jay	1.31	17.1
canary	1.42	17.1
turkey	4.09	47.5
List Containing the Typical Item <i>Sparrow</i> in a List of Medium-Typical Items		
raven	2.01	17.4
parrot	2.07	17.4
goldfinch	2.06	17.0
sparrow	1.18	48.1

the list of typical birds above. Thus the words *penguin* or *turkey* should be distinctive and well retained in the context of a list containing *robin*, *blue jay*, *sparrow*...

Eysenck's (1979) model was specifically designed to predict recognition performance. However, one might assume that recall is a combined function of item accessibility and item distinctiveness (following Neath, 1993a; and Marschark & Hunt, 1989). Thus embellished, Eysenck's model may provide a reasonable account of the effects of item typicality on recall.

The univariate and the fixed-multifeature approaches lead to different predictions in many situations. In the univariate approach, a very typical item (e.g., *sparrow*, with a score of 1.18) would be just as distinctive as an atypical item within a list containing mostly moderately typical items (typicality scores above 2.00; see the lower portion of Table 1). That is, both ends of the distribution of items along the univariate scale should be distinctive relative to the rest of the items. In a multifeature approach, the items can be represented in multivariate space. A group of items from the same conceptual category should be thought of as a cluster or cloud of points within that space. Typical category members should reside in the center, or denser part of the cloud. Moderately typical items would share some features with the highly typical items and reside well within the cloud-like structure. Typical items would always be most central or similar to the group as a whole. Such typical items would not be distinctive, independently of list structure. Atypical items should stand apart from the rest of the cloud. Thus, the univariate approach allows that very typical and/or very atypical items may be distinctive, depending on list structure. In the fixed-multifeature model, it is always predicted that atypical items will be more distinctive than moderately typical items, which will be more distinctive than highly typical items.

Recently, several researchers have rejected the fixed-multifeature approach in favor of what I will call "weighted-multifeature" approaches (Hunt & McDaniel, 1993; Schmidt, 1991). These theories were motivated by recent advances in research concerning similarity judgments that challenged the idea that the features comprising an item make a fixed contribution to item similarity (see, e.g., Gati & Tversky, 1984; Medin, Goldstone, & Gentner, 1993; Murphy & Medin, 1985). In the Hunt and McDaniel and the Schmidt models, the distinctive item is defined relative to a qualitative set of weighted attributes. Hunt and McDaniel (1993) employed the concept of alignment to capture the notion of weighted attributes. According to Hunt and McDaniel, the attributes of an item are given more or less credit depending on the overall structure of an experience. For example, "What comes to mind about CAT is different when it is preceded by MOUSE than when it is preceded by DOG" (p. 427). A similar concept was proposed by Schmidt (1991). The features of recent items were thought to be maintained in working memory. "In the absence of specific strategies that highlight certain features . . . , the weight given to a particular feature may be a direct function of the number

of times that feature has been recently processed" (p. 537). As in Eysenck's (1979) model, the distinctiveness of an item depends on feature overlap, but the weight attached to a feature is determined by context. Features that are shared by a number of items in an experience are given a lot of weight, as are features that segment an experience into clear conceptual groups.

In terms of the marble analogy, red and yellow marbles can be described with a long list of features, including size, shape, smoothness of the surface, material, and so on. All of the perceptually salient features shared by the items should be given a lot of weight, including color. However, color provides a means for organizing the experience when one item differs from others in terms of this feature. The yellow marble thus stands out against the background of other marbles. Slight differences along one or more of the other physical dimensions are given little weight. Had the marbles been of uniform size and color except for one rather large marble, size would serve as a discriminating feature, and the large marble would have been distinctive.

The weighted-multifeature approach provides yet another set of predictions concerning the effects of item typicality on memory. Consider the items in Table 1. The common items *robin*, *blue jay*, and *sparrow* share a great number of bird features, such as feathers, beak, laying eggs, and so on. According to Schmidt (1991), these shared features should be given a lot of weight in working memory. The item *turkey* shares a number of these features, and is thus not distinctive in this context. In the language of Hunt and McDaniel (1993), the attributes of turkey may be aligned to fit the context of a list of birds.

Both Schmidt (1991) and Hunt and McDaniel (1993) assumed that memory was a combined function of item distinctiveness and retrieval factors. Schmidt argued that retrieval strategies could either increase or decrease the effects of distinctiveness. According to Hunt and McDaniel, retrieval of the items is guided by the aligned features. From both points of view, it is reasonable to assume that subjects employ category information when retrieving a categorized list of words. Success at retrieving items should be a function of goodness of category membership. Because distinctiveness is not related to category typicality, these theories predict that recall should decline as items become less typical. To the extent that recognition performance is unaffected by retrieval strategies, recognition and item typicality should be unrelated.

The models of distinctiveness reviewed above yield three different predicted effects of typicality on memory. In the univariate approach, it is predicted that item distinctiveness should be a function of relative typicality, with either atypical or highly typical items becoming distinctive, depending on the rest of the list. In the fixed-multifeature approach, it is predicted that items should generally become more distinctive as they become more atypical. In the weighted-multifeature approach, it is predicted that typicality and distinctiveness will be unrelated in a list comprised of items from the same conceptual class. Within each framework, additional assump-

tions convert values of distinctiveness into effects on memory performance. Given the complexity of the theoretical landscape, it is not surprising that the review of typicality effects on memory presented below uncovered mixed results. Research concerning memory for atypical faces and actions atypical of a scripted activity provided support for the hypothesized monotonic relation between retention and typicality predicted by the fixed-multifeature model. However, research with semantic categories supports the weighted-multifeature approach. These three areas of research are reviewed briefly below.

Light, Kayra-Stuart, and Hollander (1979) reported one of the most comprehensive studies on the effects of face typicality on memory. Subjects rated pictures of faces on a scale from 1 (*very usual*) to 5 (*very unusual*). Twenty-four hours later, the subjects were given a recognition test containing half old and half new pictures. Hit rates were positively correlated with typicality ratings (where a high score indicated an atypical face), whereas false-alarm rates were negatively correlated with typicality. These correlations suggest that typicality is monotonically related to recognition performance. In subsequent experiments, faces were separated into those rated as typical and those rated as unusual, and direct memory comparisons were made (Experiments 2–4). False alarms to typical faces consistently exceeded false alarms to unusual faces. However, the researchers failed to find consistent effects of typicality on hit rates.

Enhanced memory for unusual faces has been amply demonstrated (see Bartlett, Hurry, & Thorley, 1984; Cohen & Carr, 1975; Going & Reed, 1974; Valentine & Bruce, 1986; Vokey & Read, 1992). However, within the framework of a general theory of distinctiveness, these demonstrations have been less than satisfactory. The most troubling limitation is that although the effects of face typicality have been studied in mixed-list designs, no attempt has been made to determine if, or how, the effects depend on list structure. As noted, the three theories of distinctiveness yield predictions of different effects of list structure on memory. Another shortcoming is that the effects of face typicality on memory may be limited to false-alarm rates in recognition. The effects of typicality on hit rates are weak at best (e.g., Light et al., 1979; Vokey & Read, 1992), and researchers have not investigated the effects of face typicality on recall or reconstruction. If the effects of typicality are limited to false alarms, these effects are open to several interpretations (see the discussion following Experiment 5).

Numerous researchers have demonstrated that actions atypical of a script are retained better than typical actions (Bower, Black, & Turner, 1979; Graesser, Gordon, & Sawyer, 1979; Graesser, Woll, Kowalski, & Smith, 1980). For example, Graesser et al. (1980) studied memory for scripted activities such as eating at a restaurant. Atypical actions included “put a pen in his pocket,” “bought some mints,” and “picked up a napkin off the floor.” Typical actions included “paid the bill” and “sat down at the table.” These researchers demonstrated enhanced recall and recognition of atypical actions relative to typical ac-

tions. More recent research supports a two-factor model of distinctiveness in which script typicality effects on recall are a combined function of item distinctiveness (or typicality) and item accessibility (Smith & Graesser, 1981). According to this hypothesis, atypical script actions are well recalled on an immediate test because item accessibility is relatively easy following a short delay. On a delayed memory test, item accessibility is more difficult, and thus typical items are easily recalled whereas atypical items are poorly recalled. However, script typicality is not the only factor affecting the accessibility of scripted activities. Davidson (1994) demonstrated that the effects of delay on recall depend on whether the atypical script actions are irrelevant to the script or interruptions of the scripted activity. Interruptions were better recalled than script-typical actions, independently of retention interval. In addition, vivid script-irrelevant actions were better recalled than were typical actions on delayed memory tests. These results demonstrated the importance of causal links and sentence vividness in the recall of scripts. The recognition data did support a fixed-multifeature model of distinctiveness. Script-atypical actions of all types led to higher hit rates, and lower false-alarm rates, than did script-typical actions. Unfortunately, as with research with atypical faces, no attention has been given to the role of list structure in producing good memory for atypical script actions.

Research with faces and scripts has not provided definitive tests for the theories of distinctiveness reviewed above. Perhaps semantic categories can provide a more suitable domain for studying typicality effects and their relation to models of distinctiveness. Rosch and her associates have studied the nature of semantic category structures and have provided norms of category typicality (Rosch, 1975; Rosch & Mervis, 1975). As a result, category typicality can be easily manipulated, and extraneous variables (e.g., word frequency) can be easily controlled. Also, recall and recognition tests of memory can be employed so that the effects of typicality on both memory access and discrimination processes can be observed.

Greenberg and Bjorklund (1981) demonstrated that typical category members were better recalled than were atypical category members. Thus, typical items were better remembered than were atypical items in what appears to be an ideal domain for studying typicality effects (see also Bjorklund & Bernholtz, 1986; Bjorklund & Thompson, 1983). Bjorklund’s studies are limited in two ways. First, whereas Greenberg and Bjorklund (1981) manipulated item typicality within subjects, the memory lists contained typical items from one category and atypical items from another category. Thus, the atypical items were not presented in the context of typical items from the same category. In terms of the models reviewed above, only the fixed-multifeature model, unembellished by a retrieval process, would predict better memory for the atypical items with this list structure. However, other list structures were not tested. A second limitation of Bjorklund’s research was that only recall measures of memory were employed. Other memory measures must be em-

ployed to tease apart effects of distinctiveness from effects of item accessibility.

In the research presented below, comparisons were made between memory for typical and atypical items presented in the context of typical items from the same category. Numerous studies have revealed the effects of distinctiveness on recall measures of memory, including memory for items from distinctive categories (Schmidt, 1985), script-atypical items (Graesser et al., 1980), orthographically distinctive items (Hunt & Mitchell, 1982), and perceptually salient items (McLaughlin, 1968). For this reason, the effects of category typicality were first observed on recall. However, in the studies presented below, typical items were consistently better retained than were atypical items, challenging simple models relating typicality, distinctiveness, and recall. Thus, the first series of experiments represents a succession of failed attempts to create conditions under which atypical items are well recalled. Experiments 4 and 5 tested the more elaborate or complete models of distinctiveness in which typicality effects are mediated by item accessibility. The results of these five experiments will be discussed in terms of the univariate, fixed-multifeature, and weighted-multifeature models presented above.

## EXPERIMENT 1

In Experiment 1, four categories were selected from the Rosch (1975) norms. Half of the subjects viewed lists containing highly typical targets, whereas the remainder of the subjects viewed lists that contained atypical targets. Both types of targets were embedded in a background of moderately typical category members. According to the univariate model and the fixed-multifeature model, the atypical targets should be more distinctive than the typical targets. At this point it was unclear how these effects of distinctiveness would map onto recall performance.

### Method

**Subjects.** The subjects were 80 students from the psychology subject pool, which consists of students who participate in partial fulfillment of a general psychology course requirement and students from a variety of other psychology courses who receive extra course credit for participation.

**Materials.** The following four categories were selected from the Rosch (1975) category norms: clothing, toys, vehicles, and weapons. These norms provide typicality ratings on a scale from 1 (*highly typical*) to 7 (*very atypical*). From each category, four atypical category members (mean typicality rating = 4.63, range = 3.71–5.91) were selected that matched in word frequency four items rated as highly typical ( $M = 1.65$ , range = 1.03–2.78). The mean Thorndike and Lorge (1944) ratings were 68.00 and 58.19 for the atypical and typical items, respectively. In this manner, four pairs of items were selected from each category that held word frequency relatively constant while varying item typicality. For example, the items *coat* and *hat* were paired. Both items are AA Thorndike and Lorge words, and *coat* has a typicality rating of 1.88 whereas *hat* has a typicality rating of 4.20.

Two lists were constructed for each category. Each list contained the 16 category members not selected as targets plus 4 target items. The background items had a mean typicality rating of 2.64. In one

list, the typical pair members served as targets, and appeared in serial positions 6, 8, 11, and 15. In the other list, the atypical pair members served as targets and appeared in these same serial positions. The serial positions of the target items were randomly selected with the constraint that the items were isolated from potential primacy or recency effects, and that the items did not appear contiguously.

With a few additional assumptions, the univariate model can be used to calculate values of distinctiveness for the typical and atypical items in this experiment. First, typicality ratings were used as a direct estimate of  $d$ . Second, because serial position was held constant across the two types of items, the contribution of serial position to distinctiveness was ignored. All the items in the list were included in the computations. With these assumptions, the application of Formula 1 led to percent distinctiveness scores of 8.5% (range = 6.3%–11.7%) for atypical items and 6.5% (range = 3.7%–9.4%) for typical items.

**Design.** Two factors were manipulated between subjects. The first factor was type of target (typical vs. atypical). The second factor was category (clothing, toys, vehicles, and weapons). Ten subjects served in each condition in a between-subjects design.

**Procedure.** The lists were videotaped from a computer screen. List items were individually presented at a 1.5-sec rate. The tape was played on a large-screen TV in a small classroom. The subjects were told to study the list in preparation for a free-recall test. Following list presentation, the subjects performed 2 min of arithmetic and then spent 3 min attempting to recall the words.

### Results and Discussion

The probabilities of recalling target and background items were calculated, and the number of intrusions was counted. Two types of intrusions were evaluated. Target intrusions occurred when subjects recalled targets that had not appeared on the list that they had viewed. Extralist intrusions occurred when subjects recalled category members not included in the experiment. Target intrusions provided a rough index of guessing rates for typical and atypical targets. Analyses of variance (ANOVAs) were calculated on these dependent variables. In each analysis, target typicality and category served as between-subjects factors. For all reported effects, a  $p < .05$  level of significance was employed.

Typical category targets were more likely to be recalled ( $M = .71$ ) than were atypical targets ( $M = .38$ ) [ $F(1,72) = 62.08$ ,  $MS_e = .0340$ ]. There was an interaction between typicality and category [ $F(3,72) = 6.86$ ,  $MS_e = .0340$ ]. Nonetheless, the probability of recalling typical targets exceeded the probability of recalling atypical targets for each category. Newman-Keuls tests revealed that the effect of target typicality was significant for every category but one (clothing). Typicality did not affect recall of the background list, with  $M_s = .49$  and  $.53$  for lists containing atypical and typical targets, respectively [ $F(1,72) = 1.61$ ,  $MS_e = .0146$ ]. However, typical target items were more likely to be intruded ( $M = .16$ ) than were atypical items ( $M = .01$ ) [ $F(1,72) = 21.07$ ,  $MS_e = .0214$ ].

In a nutshell, these results provide no support for the hypothesis that atypical items are better remembered than typical items. Quite the contrary, typical items were much more likely to be recalled than were atypical items. Researchers investigating the effects of category distinctiveness on memory have repeatedly reported that dis-

tinctive targets suppress recall of background items (e.g., Schmidt, 1985). There was no evidence for such an effect in this experiment. Thus these results challenged the joint assumptions that (1) typicality manipulations can be mapped onto a metric of distinctiveness, and (2) distinctiveness directly supports good recall performance. Further research was needed in order to determine which assumption was incorrect.

## EXPERIMENT 2

Perhaps the effects of typicality on recall are curvilinear. That is, very typical items, and very atypical items, may appear distinctive relative to a set of moderately typical items. In Experiment 1, the selected typical items may have been more distinctive in the test lists than the selected atypical items (calculated values of distinctiveness notwithstanding). To address this issue, three levels of typicality were tested in Experiment 2. Across experimental conditions, highly typical, moderately typical, and atypical targets were embedded in the list of moderately typical background items. The univariate model predicted that both the highly typical and the atypical extremes should be distinctive in this list structure. In the fixed-multifactor model, it is predicted that only the atypical items should be distinctive.

A second issue addressed in this experiment concerned the particular categories tested. The categories used in Experiment 1 are at best fuzzy, nontaxonomic categories. Perhaps a different pattern of results would emerge with clearly defined taxonomic categories. In Experiment 2, memory for items from the categories birds and vegetables was tested.

### Method

**Subjects and Design.** The subjects were 90 students from the same source as those in Experiment 1. There were 15 subjects in each of six experimental groups. Each subject was tested on two lists of words. Type of target (high, medium, and atypical) was crossed with two orders of list presentation (birds first vs. vegetables first) as between-subjects factors. List served as the only within-subjects factor.

**Materials.** Eight highly typical, eight moderately typical, and eight atypical birds and vegetables were selected from the Rosch (1975) norms. From these eight items, four were selected that held the average Kücera and Francis (1967) and Thorndike and Lorge (1944) frequencies constant across typicality. Sixteen additional

category items were selected to fill the list. Table 2 summarizes the typicality and word frequency ratings for the target items.

Three lists were constructed for each category. All three lists contained 16 background items and four targets. The targets always appeared in serial positions 6, 8, 11, and 15. One list contained high-typical targets (mean typicality = 1.44, range = 1.18–1.83), the second list contained medium-typical targets (mean typicality = 2.52, range = 1.89–3.32), and the third list contained atypical targets (mean typicality = 4.75, range = 4.09–5.56). The background items had a mean typicality rating of 2.62 (range = 1.75–3.67). Following the assumption made in Experiment 1, percent distinctiveness scores were calculated for each type of item within the context of the fixed set of background items. These scores were 5.85%, 4.40%, and 7.65% for typical, moderately typical, and atypical items, respectively. In terms of a univariate scale of distinctiveness, the atypical items should be best retained, followed by the typical items, and then the moderately typical items.

**Procedure.** With one exception, the procedure for Experiment 2 was identical to that for Experiment 1. Each subject independently studied and was tested on two lists of words (see the design description above). Each list was followed by a 2-min arithmetic task and a 3-min recall period.

### Results and Discussion

Analyses similar to those used in Experiment 1 were conducted. Typicality had a significant effect on target recall [ $F(2,84) = 5.03$ ,  $MS_e = .0677$ ]. The probabilities of recalling high-typical, medium-typical, and atypical items were .53, .47, and .38, respectively. There were no effects of target typicality on recall of the background items [ $F(2,84) = 2.32$ ,  $MS_e = .0190$ ]. The only other significant effects were the effects of list on target and background item recall [ $F(1,84) = 33.09$ ,  $MS_e = .0390$ , and  $F = 8.05$ ,  $MS_e = .0114$ , respectively]. Target items from the list of birds were better retained than were targets from the list of vegetables. Background vegetables were better recalled than background birds. There were no interactions involving list.

These results demonstrated that as items became less typical they became less memorable, replicating the effects reported in Experiment 1. The reported effects of typicality were not confined to a restricted range of typicality, nor to a particular type of category. These results challenged the univariate approach to distinctiveness in which both typical and atypical items should be distinctive relative to a background of moderately typical items. This experiment also provided further evidence contradicting the hypothesis that when distinctiveness is de-

Table 2  
Mean Typicality Ratings, Word Frequencies, and Recall and Recognition Performance for  
Targets Used in Experiments 2, 4, and 5

Measure	Item					
	High Typical		Medium Typical		Atypical	
	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>
Typicality	1.44	.076	2.52	.190	4.75	.193
Word frequency	6.65	2.46	6.75	2.47	6.25	2.10
Recall (Experiment 2)	.53	.046	.47	.042	.38	.046
Simulated recall (Experiment 4)	.59	.035	.20	.028	.16	.025
Confidence old/old (Experiment 5)	4.13	.233	4.45	.103	4.55	.093
Confidence old/new (Experiment 5)	2.29	.227	2.11	.156	1.92	.128

fined in terms of a fixed set of category features (i.e., the fixed-multifeature model), distinctive items are better remembered than are common items. Clearly something is amiss, but once again it is unclear whether the error is in mapping typicality to a scale of distinctiveness, or in relating values of distinctiveness to recall performance.

### EXPERIMENT 3

The background items in Experiments 1 and 2 represented a range of category typicality. One could argue that this range of items diluted the contrast between atypical category members and the rest of the list. For example, the atypical bird *penguin* may be quite distinctive within the context of very typical birds like *robin* and *sparrow*, but not distinctive within a list containing the moderately typical birds *stork* and *flamingo*. The lists in Experiment 3 were designed to provide a stronger contrast by presenting a single atypical item in the context of seven very typical category members. Perhaps within such a strong contrast, atypical category members will be better recalled than typical category members.

#### Method

**Subjects and Design.** The subjects were selected from the same source as those from Experiments 1 and 2. There were 10 subjects in each cell of the 2 (high vs. low typicality)  $\times$  4 (category) mixed design. Type of target was manipulated between subjects. Each subject was tested on two lists of words, and each list contained two categories, so that data on all four categories were obtained from each subject. The order of the lists, and the order of the categories within the list, were counterbalanced across subjects, creating eight experimental groups.

**Materials.** Once again, items were selected from the Rosch (1975) norms. The four categories of vehicles, clothing, birds, and vegetables were selected. Whenever possible, the top eight typical items were selected, and one of these items was matched in word frequency with an extremely atypical category member. These two items served as target words for this category. For example, the top eight selected items from the category of vehicles ranged in typicality from 1.02 to 1.65. The item *car*, with a typicality rating of 1.24, was matched with the item *feet*, with a typicality rating of 5.34. Both items are AA Thorndike and Lorge (1944) words. Across categories, the mean typicality ratings for background items, typical targets, and atypical targets were 1.37, 1.21, and 5.42, respectively. Lists were constructed from these items so that each list contained eight items from one category in a block, followed by eight items from a second category. Each list contained 16 items and two targets. Target items appeared in the middle of the block of eight items (position 5). Half the lists contained typical targets and the remaining half contained atypical targets. For example, one group read a list containing the items *truck*, *taxi*, *motorcycle*, *ambulance*, *feet*, *bus*, *jeep*, *automobile*. A second group read the same items with the word *car* substituted for the word *feet*. On the basis of univariate and fixed-multifeature models of distinctiveness, the item *feet* should be more distinctive than the item *car*.

Additional assumptions were needed before percent distinctiveness scores could be calculated for this set of materials. First, it was assumed that distinctiveness should be defined relative to other items in the category. Only the seven items surrounding the target, and from the same category as the target, were included in the calculations. Second, it was assumed that serial position would not impact item distinctiveness. The percent distinctiveness values averaged 10.78 (range = 8.9–11.9) for typical items and 41.8

(range = 40.6–43.5) for atypical items. Thus these materials provided for a sizable manipulation of distinctiveness defined along the univariate scale.

**Procedure.** Except for the fact that each list contained two categories, the procedure for this experiment was identical to that in Experiment 2.

### Results and Discussion

The probability of recalling typical target items ( $M = .73$ ) exceeded the probability of recalling atypical targets ( $M = .45$ ) [ $F(1,78) = 22.14$ ,  $MS_e = .2858$ ]. Although the effect of typicality interacted with category [ $F(3,234) = 2.51$ ,  $MS_e = .1899$ ,  $p < .06$ ], typical items were better remembered than were atypical items in every category, and this effect was significant for every category but one (vegetables). There were no effects of target typicality on the recall of background items, with  $M_s = .63$  in each condition [ $F(1,78) = .05$ ,  $MS_e = .0571$ ]. Typical targets were more likely to be erroneously recalled from lists that contained atypical targets ( $M = .20$ ) than vice versa ( $M = .00$ ) [ $F(1,78) = 30.81$ ]. The latter effect also interacted with category [ $F(3,234) = 5.17$ ,  $MS_e = .0660$ ].

Under the conditions created in this experiment, one might expect the atypical items to stand out or be distinctive. The contrasts between the categories and the atypical items were pushed to the limit. Nonetheless, the typical items were better recalled than were the atypical items. At this juncture, one must either abandon the univariate and fixed-feature definitions of distinctiveness or embed these definitions in more elaborate models to explain why atypical items are poorly recalled.

### EXPERIMENT 4

As noted, it is asserted in several theories that recall of distinctive events is mediated by a two-stage process. In the first stage, items must be accessed, whereas during the second stage, item occurrence information is evaluated to determine what items to recall (Marschark & Hunt, 1989; Neath, 1993a). Atypical category members may be difficult to access. Once accessed, it may be relatively easy to decide which atypical items were presented. The recall tests employed in Experiments 1–3 may have emphasized the role of item access and thus obscured the effects of distinctiveness. Support for this interpretation was found in the intrusion rates reported in the first three experiments. Typical targets were frequently intruded in the recall of lists not containing those targets. Intrusions of atypical targets were very low to nonexistent. The two-stage hypothesis leads to two additional predictions. First, measures of item accessibility should predict recall performance. This prediction is tested in Experiment 4. Second, atypical items should be better recognized than typical items. Recognition performance is evaluated in Experiment 5.

In Experiment 4, simulated recall was used to provide a measurement of item accessibility. Subjects were asked to simulate recall from the categories of birds and vegetables. This procedure is similar to that employed in the



collection of category production norms, such as the Battig and Montague (1969) norms. Simulated recall, item typicality, and calculated distinctiveness (from Formula 1) were then used to predict the actual recall performance from Experiment 2. A two-stage version of the univariate approach predicted that item recall should be a combined function of simulated recall and calculated distinctiveness. The fixed-multifeature model predicted that actual recall should be a combined function of simulated recall and item typicality. According to a weighted-multifeature approach, subjects should use the shared features of the conceptual core of the category to guide reconstruction. Items that fit this core will be better remembered than items that do not fit this core. As a result, high-typical items should be better recalled than medium-typical items, and medium-typical items should be better recalled than atypical items. This pattern of results should be found in both simulated recall and actual recall. As a result, typicality and simulated recall are essentially measures of the same thing: goodness of category membership. These two measures should be highly correlated, and each should be correlated with item recall. There should be no relation between relative calculated distinctiveness (in which both very typical and very atypical items are distinctive) and recall.

## Method

**Subjects and Design.** The subjects were 30 students who participated for extra course credit. Half of the students were asked to recall the birds first and vegetables second. The order of testing was reversed for the other half of the students.

**Materials.** The categories of birds and vegetables were once again employed. Items defined as targets in Experiment 2 served as targets in this experiment. The recall protocols from subjects were scored for the recall of high-typical, medium-typical, and atypical targets.

**Procedure.** The students were told about the difference between an actual experiment and a simulated experiment. They were told that the purpose of a simulated experiment was to provide a baseline for comparison to the actual experiment. The procedure for Experiment 2 was then briefly described. They were told, "I want you to try to recall the same list, without having seen the words." They were then given the category name and asked to try to recall the 20 items from the list. No mention was made of the structure of the list, or that the experiment had anything to do with category typicality. Subjects were given 3 min to recall items from the first category. They were then told the second category name and were given 3 min to recall items from that category.

## Results and Discussion

An ANOVA was conducted using simulated target recall as the dependent variable. Type of target (high-typical, medium-typical, and atypical), and list (birds vs. vegetables) served as within-subjects factors and order served as a between-subjects factor.

There were two main effects, type of target [ $F(1,28) = 63.47, MS_e = .0530$ ] and list [ $F(1,28) = 6.36, MS_e = .0241$ ]. Recall declined with category typicality ( $M_s = .59, .20, \text{ and } .16$  for high-typical, medium-typical, and atypical targets, respectively). Newman-Keuls analyses revealed that high-typical targets were recalled better than were

medium-typical and atypical targets. The difference between medium-typical and atypical targets was not significant. These results support the prediction that item accessibility declines as typicality declines. However, further tests are needed to determine if simulated recall is a reliable predictor of recall, or if a combination of simulated recall, typicality, and/or distinctiveness predicts recall.

Modeling was approached from two directions. First, item recall (from Experiment 2) was modeled collapsing across subjects, and employing item typicality, simulated recall (from Experiment 4), and calculated distinctiveness as predictor variables. Second, subject recall was modeled, employing average item typicality, item-simulated recall, and item distinctiveness as predictor variables. Separate measures of simulated recall from each list, in each list order, were used in the appropriate cells. In each of these analyses, recall was assumed to be a simple additive function [i.e.,  $P(\text{recall}) = f(\text{distinctiveness} + \text{simulated recall})$ ]. These two types of analyses were then repeated with a multiplicative model of recall [i.e.,  $P(\text{recall}) = f(\text{distinctiveness} \times P(\text{simulated recall}))$ ].

The correlations among typicality, calculated distinctiveness, simulated recall, and actual recall are presented in Table 3. In the item analysis, typicality and simulated recall were highly correlated [ $r(23) = -.66$ ], indicating that as items became less typical (higher scores), they were less likely to be recalled. Also, item recall was significantly correlated with typicality [ $r(23) = -.43$ ], and marginally correlated with simulated recall [ $r(23) = .36, p < .088$ ]. However, the correlation between recall and calculated distinctiveness was not significant [ $r(23) = -.28$ ]. A multiple regression combining typicality and simulated recall to predict actual recall yielded  $r(2,21) = .44$ . The multiple  $r$  did not account for any more variance than did the simple correlation between typicality and recall. A multiple regression combining calculated distinctiveness and simulated recall was not significant [ $r(2,21) = .43$ ].

In the additive model of subject recall, recall was significantly correlated with typicality [ $r(89) = -.32$ ] and

**Table 3**  
Correlations Between Typicality, Simulated Recall, Calculated Distinctiveness, and Recall Collapsed Over Subjects and Collapsed Over Items

Measure	Typicality*	Simulated Recall	Calculated Distinctiveness
Collapsed Over Subjects			
Typicality			
Simulated recall	-.66†		
Calculated distinctiveness	.61†	-.16	
Recall	-.43†	.36	-.28
Collapsed Over Items			
Typicality			
Simulated recall	-.79†		
Calculated distinctiveness	.71†	-.15	
Recall	-.32†	.25†	-.20

\*Higher typicality scores reflect more atypical items. † $p < .05$ .



with simulated recall [ $r(89) = .25$ ], and marginally correlated with calculated distinctiveness [ $r(89) = -.20, p < .06$ ]. The latter correlation indicated that recall declined as items became more distinctive! The multiple regression combining typicality and simulated recall yielded  $r(2,87) = .32$ , again failing to show a significant improvement in predicting recall by adding simulated recall to the variance accounted for by typicality. The multiple  $r$  combining calculated distinctiveness and simulated recall to predict recall was  $r(2,87) = .30$ . Adding calculated distinctiveness to simulated recall did not significantly improve the prediction of recall [ $F(2,87) = 2.74$ ].

In the test of the multiplicative model of item recall, the product of typicality and simulated recall was not significantly correlated with item recall [ $r(23) = .28$ ]. The product of calculated distinctiveness and simulated recall was not significantly correlated with recall [ $r(23) = .32$ ]. The correlation between subject recall and the product of typicality and simulated recall was even less reliable [ $r(89) = -.05$ ]. However, the correlation between recall and the product of calculated distinctiveness and simulated recall was marginally reliable [ $r(89) = .21, p < .06$ ]. Each of these correlations accounted for less variance than did the simple correlation between typicality and recall.

In yet another approach to modeling recall, the recall scores of each subject in Experiment 2 were transformed. The average simulated recall score from Experiment 4 was subtracted from the recall scores from the same condition in Experiment 2. One can view these difference scores as a measure of the extent to which recall of items from a category is improved by explicit presentation of the items. An ANOVA was conducted on the resulting data set with type of target (high-typical, medium-typical, or atypical) and order serving as between-subjects factors, and list serving as a within-subjects factor. There was a significant effect of type of target [ $F(2,84) = 27.42, MS_e = .0677$ ]. The difference scores were  $-.05, .27$ , and  $.23$  for high-typical, medium-typical, and atypical targets. These means can be obtained (within rounding error) by simply subtracting simulated recall in Experiment 4 from recall in Experiment 2 (see Table 2). One interpretation of these results is that recall of medium-typical and atypical items benefited more from presentation than did presentation of typical items. This was the first finding in any of the experiments that supports the hypothesis that memory for atypical items is in any way better than memory for typical items. Even so, these data were inconsistent with both the univariate and fixed-multifeature models developed above. According to the univariate model, both the typical and atypical targets should have been distinctive within the list structure used in this experiment, and the medium-typical items should have been the least well retained. Clearly the pattern of difference scores does not fit this prediction. The fixed-multifeature model predicted increasing distinctiveness with decreasing typicality. Nonetheless, atypical items benefited slightly less from presentation than did medium-typical items.

In summary, whereas recall was indeed correlated with both typicality and simulated recall, each attempt to combine typicality, simulated recall, and distinctiveness to improve prediction of recall failed. Recall was not related to calculated distinctiveness. In other words, the only reliable and independent predictor of recall was typicality, with recall declining as items became more atypical. These results suggest that the mappings of typicality to a metric of distinctiveness are in error in both the univariate and the fixed-multivariate models.

## EXPERIMENT 5

The relation between item typicality and recognition was tested in Experiment 5. The materials were the same as those employed in Experiments 2 and 4. A recognition test was constructed that contained the full range of item typicality, and confidence judgments served as the dependent measure. Recognition performance should be influenced less by item accessibility, providing a clearer picture of the effects of distinctiveness on the quality of the memory trace. In the univariate model, it was predicted that atypical items would be better recognized than typical items, and that typical items would be better recognized than medium-typical items. In the fixed-multifeature model, it was predicted that recognition would improve as items became more atypical. According to the weighted-multifeature model, with these materials item distinctiveness should be unrelated to item typicality. As such, recognition performance should be unrelated to typicality.

### Method

**Subjects and Design.** The subjects were 45 students from the psychology subject pool. Type of target (high-typical, medium-typical, and atypical) was manipulated between subjects. List (birds vs. vegetables) served as a repeated measure.

**Materials.** The materials selected for Experiment 2 were again employed. One recognition test was constructed for each of the categories. The recognition test contained eight items from each typicality group, plus the 16 background items, for a total of 40 items. For each subject, half of these items were "old." In addition, for each subject the "old" targets from their list were matched by an equal number of "new" items of the same typicality level as the old targets. Items that served as targets for one group served as distractors for the other two groups. This test construction enabled us to collect confidence judgments for "new" items across all levels of typicality for all subjects, independently of the type of targets they studied.

**Procedure.** The procedure for list presentation was the same as that employed in Experiment 2. Once again a 2-min retention interval separated presentation and test. On the recognition test subjects were asked to circle a confidence score ranging from 1 to 5 (1 = *sure the item did not appear on their list*; 5 = *sure the item did appear on their list*). The subjects were given 3 min to complete the recognition test. After they completed the test for List 1, List 2 was presented, followed by a second distractor task and the recognition test for List 2.

### Results and Discussion

A two-way ANOVA was conducted on confidence judgments on old targets, with type of target studied (high-

typical, medium-typical, and atypical) and list (1 and 2) serving as the independent variables. Recognition was not significantly affected by target typicality [ $F(2,42) = 1.09$ ,  $MS_e = 1.30$ ,  $Ms = 4.13, 4.45$ , and  $4.55$  for high-typical, medium-typical, and atypical targets, respectively]. Confidence judgments on recognition lures were analyzed with a 3 (type of target studied)  $\times$  2 (list)  $\times$  3 (type of recognition lure: high-typical, medium-typical, and atypical) mixed factorial design. There was a marginally significant effect of type of target [ $F(2,42) = 3.08$ ,  $MS_e = 2.40$ ,  $p < .06$ ], with mean confidence ratings of 2.43, 1.87, and 2.04, for high-typical, medium-typical, and atypical conditions, respectively. More importantly, the effect of type of lure was significant [ $F(2,84) = 8.20$ ,  $MS_e = 0.36$ ;  $Ms = 2.29, 2.11$ , and  $1.93$  for high-typical, medium-typical, and atypical items, respectively]. In other words, subjects became more confident in judging recognition lures as "new" as lure typicality decreased.

These results provide some support for the hypothesis that atypical items are easy to recognize, but as in the research with faces (Light et al., 1979) and scripts (Graesser et al., 1980), the support was found only in ratings of "new" items. Furthermore, there was a significant interaction between type of item presented and type recognition lure rated [ $F(4,84) = 3.49$ ,  $MS_e = 0.36$ ]. That is, recognition judgments were a joint function of the structure of the memory list and the type of item rated on the recognition test. This interaction is depicted in Figure 1. Typicality had a significant effect on recognition judgments when subjects viewed high-typical [ $F(2,84) = 12.14$ ,  $MS_e = 4.43$ ]. The effect of typicality was not significant for subjects who viewed medium-typical [ $F(2,84) = 2.95$ ,  $MS_e = 1.08$ ] or atypical [ $F(2,84) = 0.27$ ,  $MS_e = 0.10$ ] targets. Thus, the effects of category typicality on recognition were limited to rather special circumstances. Those effects were observed only in the ratings of new items by subjects who viewed highly typical category members.

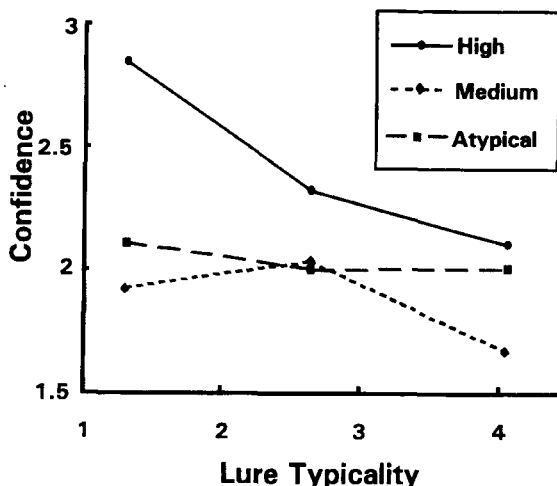


Figure 1. Confidence judgments of recognition lures as a function of type of target item contained in the list. Higher ratings reflect decreased confidence that the "new" items were new.

As noted, the effects of typicality obtained on recognition judgments are open to several interpretations. Perhaps, as suggested by some models of distinctiveness, atypical items are somehow better represented (with a more detailed memory trace, greater activation, increased probability of being tagged, or a higher probability of feature sampling) than are moderately typical or typical items that blend with the background. However, such models predict effects of typicality on the judgments of old items, something not found in this or other experiments (e.g., Graesser et al., 1980; Light et al., 1979). Alternatively, perhaps presentation of a typical item (e.g., *robin*) leads to some activation or marking of other typical category members (e.g., *canary*). In contrast, presentation of an atypical item (e.g., *turkey*) leads to little activation of other atypical category members (e.g., *penguin*). As a result, after viewing the item *robin*, subjects are less confident they did not see *canary* than they are that they did not see *penguin*. This alternative interpretation of typical effects in recognition leads one to predict an interaction between type of target viewed and type of lure rated. The interaction depicted in Figure 1 conforms to this prediction.

## GENERAL DISCUSSION

These experiments make two important contributions. First, they provide a systematic investigation of the effects of item typicality on memory. Second, they bring different ideas about distinctiveness into clearer focus and provide an empirical challenge for several theoretical frameworks.

In four experiments, recall performance declined as items became more atypical. This finding was reported with different categories, with taxonomic and nontaxonomic categories, with different ranges of typicality, and with different magnitudes of contrast. Poor recall of atypical items in these experiments may be explained by two-stage theories of memory positing that the atypical items are difficult to access. This hypothesis was evaluated in Experiments 4 and 5. Experiment 4 provided a direct measure of item accessibility as a function of item typicality. Whereas atypical items were more difficult to access than highly typical items, correlational analyses revealed that the most potent predictor of recall was typicality, with no increase in prediction gained by consideration of calculated distinctiveness and/or accessibility. In addition, when recall was corrected by subtracting item accessibility, the pattern of results was different from that predicted by either the univariate or the fixed-multifactor model. The recognition data from Experiment 4 seemed, at first glance, to support the hypothesis that atypical items were well retained. That is, confidence that a recognition lure was "new" declined with increasing item typicality. However, closer inspection revealed that confidence judgments of lures were greatly affected by the type of target items viewed by the subjects. Poor confidence on judgments concerning typical lures was restricted to subjects who viewed typical tar-

gets. This result was consistent with spreading-activation and shared-feature hypotheses, and not with hypothetical differences in recognition accuracy tied to item typicality. There was no evidence to support a two-stage process in which decreasing typicality adversely affected memory in the first stage but benefited memory through the increased distinctiveness in the second stage.

One potential criticism of these results is that typicality and category accessibility were confounded in each of the experiments. However, given the nature of natural categories, this confound is essential to the experimental design. If natural categories are represented by prototypes (as suggested by Rosch, 1975), then atypical items will, by nature, be more loosely connected to their conceptual category than will typical items. In addition, this structure of semantic memory does not necessarily imply that atypical items are difficult to access in the memory for a specific learning episode. The whole concept of distinctiveness presupposes that unusual or atypical events are often well recalled. In fact, two of the models developed above predicted good retention of atypical items from certain learning contexts, and research concerning script typicality has repeatedly demonstrated good recall of script-atypical actions. Thus the relation between item typicality and item accessibility is an empirical issue addressed in these experiments, not a matter of confounded variables.

These results are consistent with those of other investigations into the effects of typicality on episodic memory. Greenberg and Bjorklund (1981) reported that atypical category members were more poorly recalled than were typical category members. The results of the present experiments are also consistent with much of the research concerning face typicality (e.g., Light et al., 1979) and script typicality (e.g., Graesser et al., 1980). That is, judgments of highly typical lures were less confident than were judgments of atypical lures. However, the results are inconsistent with studies demonstrating good recall of vivid atypical script actions (e.g., Davidson, 1994).

The theoretical contribution of this research is equally compelling. At an intuitive level, it seems that distinctive or unusual stimuli would be well remembered. However, distinctiveness can be formulated in several different ways, as illustrated by the univariate, fixed-multifeature, and weighted-multifeature models reviewed above. These models yield different predictions concerning how typicality and distinctiveness are related. According to the univariate approach, item distinctiveness should be calculated from the distribution of items along a univariate scale of typicality. Either typical or atypical category members may be distinctive, depending on the structure of the list. According to the fixed-multifeature model, item distinctiveness and item typicality should be monotonically related, independently of list structure. According to the weighted-multifeature approach, item typicality and item distinctiveness should be unrelated in a well-structured list.

The results reported above lead one to conclude that the univariate and fixed-multifeature approaches are

flawed. In fact, they appear to have a common problem. When distinctiveness is defined in terms of feature overlap, the concept takes on the properties of a univariate continuous scale, similar to that of the univariate model. That is, as feature overlap declines, items should become more distinctive. Within this approach, similarity and distinctiveness can be represented by distance in multivariate semantic "space." Both frameworks lead to the erroneous prediction that an item that is distant from other items should be distinctive and well retained. The results reported above are inconsistent with this view. Returning to the marble analogy, we can all agree that a yellow marble should stand out in the context of a group of red marbles, but not in the context of a group of marbles varying in hue from red through orange and yellow. By analogy, an atypical bird should stand out in a narrowly defined list of typical birds. This analogy is inappropriate, for it fails to consider that events vary along multiple dimensions, and that the importance that we attach to a given dimension is greatly influenced by context. It appears inappropriate to try to capture distinctiveness in terms of a distance metric, be it along a single dimension or within a fixed-multidimensional semantic space. One cannot employ a distance metric in a space that bends and folds with changes in the context of an experience.

An alternative view is that events are represented by a set of weighted features. Feature weights are determined by the number of recent items that share the feature. If an item contains little feature overlap with other items in the experience, it is represented as a member of a different conceptual group. Additional weight may be attached to the features that set the item apart from other recent items, providing for clearer conceptual groups. As a result, the distinctive item lies on the other side of some boundary, and is thus not represented as a part of the group of other items in an experience. That is, the distinctive item is discontinuous or incongruent with the surrounding context. The "incongruity hypothesis" proposed by Schmidt (1991) incorporates this discontinuous view of distinctiveness. According to this hypothesis, as items are presented they are compared to active cognitive structures. The active structure can be thought of as a collection of weighted features. A physiological "orienting" response is elicited by items that do not fit this active structure. This response is associated with a surge in attentional resources and with enhanced storage of item-specific information concerning the incongruous item. In addition, the distinctive item influences the organizational structure of the experience, which in turn influences retrieval processes.

Atypical members of a conceptual class are unlikely to elicit the orienting response because they share a number of features with other category members. In the context of a list containing members from one conceptual class, the shared features will be weighted heavily. For example, even though penguins are atypical birds, they do have wings and a feather-like covering. In the context of a list of birds, these features have a lot of weight, and thus *penguin* is not incongruous with the background

list. Presentation of the item *whale*, however, should lead to incongruity because of minimal feature overlap with the set of weighted "bird" features. As a result, typicality manipulations do not lead to the pattern of results found for manipulations that create incongruity. Instead, recall declines as typicality declines, reflecting the role of list structure on item retrieval.

A similar concept was developed by Hunt and McDaniel (1993) in their concept of alignment. In this view, distinctiveness is the result of "aligned differences." Such differences are useful for retrieval. Thus, Hunt and McDaniel downplayed the role of encoding processes in producing the effects of distinctiveness; instead they stressed the role of retrieval processes. However, the process of alignment itself is tied to encoding, so this view shares with the incongruity hypothesis the idea that encoded attributes that determine item distinctiveness depend on the context in which the item appears.

Other researchers have also emphasized retrieval factors in producing enhanced memory for distinctive events (McDaniel, Einstein, DeLosh, May, & Brady, 1995; Neath, 1993a; Riefer & Rouder, 1992). Retrieval has been emphasized because the effects of distinctiveness are sometimes obtained in free recall but not in cued recall or recognition. In the present experiments, atypical items were less susceptible to false alarms on a recognition test than were typical items (Experiment 5), even though typical items were consistently recalled better than were atypical items (Experiments 1-3). In other investigations, the effects of distinctiveness were equally robust on recall and recognition tests (see Schmidt, 1991, for a review). Clearly, further research is needed to separate possible effects of encoding and retrieval in producing the effects of distinctiveness.

Both the incongruity hypothesis (Schmidt, 1991) and the Hunt and McDaniel (1993) framework are in need of elaboration. Both hypotheses fail to provide specific predictions as to when effects of distinctiveness should be expected. For example, the manipulation of category typicality is analogous in many ways to the manipulation of orthographic distinctiveness. The orthographically distinctive item is essentially an atypical member of the category "English words." Using the logic developed above, one could argue that attributes shared across words would be preeminent in the context of a list of common words. The odd features of an unusual word should be given little weight. In the words of Hunt and McDaniel (1993), "Obvious differences among items may not be encoded in well-structured situations" (p. 428). Why, then, are orthographically distinctive words recalled better than are common words (Hunt & Elliott, 1980, Zechmeister, 1972)? Apparently the orthographically distinctive item is different enough from common items for subjects to encode that difference. The challenge for theories of distinctiveness is predicting a priori when differences between items will and will not lead to enhanced memory.

In summary, the relation between category typicality and episodic memory was investigated in the present re-

search. Three models of distinctiveness were applied to category typicality effects, leading to clear differences in predictions. A univariate model related item distinctiveness to relative typicality. A fixed-multifeature model predicted a monotonic function between item typicality and distinctiveness. A weighted-multifeature model predicted that within the context of a single conceptual class, typicality and distinctiveness would be unrelated. These models were evaluated both with and without consideration of the role of item accessibility in memory performance. The results of the recall tests clearly supported the weighted-multifeature approach in that recall declined as items became less typical. The recognition data were more easily explained in terms of poor discriminations between "old" and "new" highly typical category members than in terms of increased memorability of atypical category members. These results challenge many traditional views of the effects of distinctiveness on memory. They are consistent with the view that stimuli are subject to qualitative shifts in feature weights as a function of context. Distinctiveness should be defined with respect to these weighted features, as suggested by Hunt and McDaniel (1993) and Schmidt (1991).

## REFERENCES

- BARTLETT, J. C., HURRY, S., & THORLEY, W. (1984). Typicality and familiarity of faces. *Memory & Cognition*, *12*, 219-228.
- BATTIG, W. F., & MONTAGUE, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs*, *80* (3, Pt. 2).
- BJORKLUND, D. F., & BERNHOLTZ, J. E. (1986). The role of knowledge base in the memory performance of good and poor readers. *Journal of Experimental Child Psychology*, *41*, 367-393.
- BJORKLUND, D. F., & THOMPSON, B. E. (1983). Category typicality effects in children's memory performance: Qualitative and quantitative differences in the processing of category information. *Journal of Experimental Child Psychology*, *35*, 329-344.
- BOWER, G. H., BLACK, J. B., & TURNER, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*, 177-220.
- COHEN, M. E., & CARR, W. J. (1975). Facial recognition and the von Restorff effect. *Bulletin of the Psychonomic Society*, *6*, 383-384.
- DAVIDSON, D. (1994). Recognition and recall of irrelevant and interruptive atypical actions in script-based stories. *Journal of Memory & Language*, *33*, 757-775.
- EYSENCK, M. W. (1979). Depth, elaboration, and distinctiveness. In L. S. Cermak & F. I. M. Craik (Eds.) *Levels of processing in human memory* (pp. 89-118). Hillsdale, NJ: Erlbaum.
- GATI, I., & TVERSKY, A. (1984). Weighting common and distinctive features in perceptual and conceptual judgments. *Cognitive Psychology*, *16*, 341-370.
- GOING, M., & READ, J. D. (1974). Effects of uniqueness, sex of subject, and sex of photograph on facial recognition. *Perceptual & Motor Skills*, *39*, 109-110.
- GRAESSER, A. C., GORDON, S. E., & SAWYER, J. D. (1979). Recognition memory for typical and atypical actions in scripted activities: Tests of a script pointer + tag hypothesis. *Journal of Verbal Learning & Verbal Behavior*, *18*, 319-332.
- GRAESSER, A. C., WOLL, S. B., KOWALSKI, D. J., & SMITH, D. A. (1980). Memory for typical and atypical actions in scripted activities. *Journal of Experimental Psychology: Human Learning & Memory*, *16*, 503-515.
- GREENBERG, M. S., & BJORKLUND, D. F. (1981). Category typicality in free recall: Effects of feature overlap or differential category encod-

- ing? *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 145-147.
- HUNT, R. R., & ELLIOTT, J. M. (1980). The role of nonsemantic information in memory: Orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General*, **109**, 49-74.
- HUNT, R. R., & MCDANIEL, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory & Language*, **32**, 421-445.
- HUNT, R. R., & MITCHELL, D. B. (1982). Independent effects of semantic and nonsemantic distinctiveness. *Journal of Experimental Psychology: Human Learning & Memory*, **8**, 81-87.
- KOFFKA, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace & World.
- KÜCERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- LIGHT, L. L., KAYRA-STUART, F., & HOLLANDER, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning & Memory*, **5**, 212-228.
- MARSCHARK, M., & HUNT, R. R. (1989). A reexamination of the role of imagery in learning and memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 710-720.
- MCDANIEL, M. A., & EINSTEIN, G. O. (1986). Bizarre imagery as an effective memory: The importance of distinctiveness. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 54-65.
- MCDANIEL, M. A., EINSTEIN, G. O., DELOSH, E., MAY, C., & BRADY, P. (1995). The bizarreness effect: It's not surprising, it's complex. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 422-435.
- MC Laughlin, J. P. (1968). Recall and recognition measures of the von Restorff effect in serial learning. *Journal of Experimental Psychology*, **78**, 99-102.
- MEDIN, D. L., GOLDSTONE, R. L., & GENTNER, D. (1993). Respects for similarity. *Psychological Review*, **100**, 254-278.
- MURDOCK, B. B. (1960). The distinctiveness of stimuli. *Psychological Review*, **67**, 16-31.
- MURPHY, G. L., & MEDIN, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, **92**, 289-316.
- NEATH, I. (1993a). Contextual and distinctive processes and the serial position function. *Journal of Memory & Language*, **32**, 820-840.
- NEATH, I. (1993b). Distinctiveness and serial position effects in recognition. *Memory & Cognition*, **21**, 689-698.
- NEATH, I., & KNOEDLER, A. J. (1994). Distinctiveness and serial position effects in recognition and sentence processing. *Journal of Memory & Language*, **33**, 776-795.
- RIEFER, D. M., & ROUDER, J. N. (1992). A multinomial modeling analysis of the mnemonic benefits of bizarre imagery. *Memory & Cognition*, **20**, 601-611.
- ROSCH, E. H. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, **104**, 192-233.
- ROSCH, E. H., & MERVIS, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, **7**, 573-605.
- SCHMIDT, S. R. (1985). Encoding and retrieval processes in the memory for conceptually distinctive events. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 565-578.
- SCHMIDT, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition*, **19**, 523-542.
- SLAMECKA, N. J., & GRAF, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 592-604.
- SMITH, D. A., & GRAESSER, A. C. (1981). Memory for actions in scripted activities as a function of typicality, retention interval, and retrieval task. *Memory & Cognition*, **9**, 550-559.
- THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's word book of 30,000 words*. New York: Teachers College Press, Columbia University.
- VALENTINE, T., & BRUCE, V. (1986). Recognizing familiar faces: The role of distinctiveness and familiarity. *Canadian Journal of Psychology*, **40**, 300-305.
- VOKEY, J. R., & READ, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition*, **20**, 291-302.
- ZECHMEISTER, E. B. (1972). Orthographic distinctiveness as a variable in word recognition. *American Journal of Psychology*, **85**, 425-430.

(Manuscript received August 17, 1994;  
revision accepted for publication January 13, 1995.)