Journal of Experimental Psychology:
Learning, Memory, and Cognition
2003, Vol. 29, No. 6, 1194–1209

0278-7393/03/$12.00 DOI: 10.1037/0278-7393.29.6.1194

A Hybrid-Similarity Exemplar Model for Predicting Distinctiveness Effects in Perceptual Old–New Recognition

Robert M. Nosofsky
Indiana University Bloomington

Safa R. Zaki
Williams College

In 2 sets of experiments, the authors investigated the basis for old-item distinctiveness effects in perceptual recognition, whereby distinctive old items are recognized with higher probability than are typical old items. In Experiment 1, *distinctive* old items were defined as those lying in isolated regions of a continuous-dimension similarity space. In this case, any beneficial effects of distinctiveness were absent or small, regardless of the structure of the test list used to assess recognition memory. In Experiment 2, distinctive items were defined as those objects containing certain discrete, individuating features. In this case, large old-item distinctive effects were observed, with the nature of the effects being modulated by the structure of the test lists. A hybrid-similarity exemplar model, combining elements of continuous-dimension distance and discrete-feature matching, was used to account for these distinctiveness effects in the recognition data.

According to global-activation models of old–new recognition memory, observers make recognition judgments on the basis of the overall familiarity of test objects. Presentation of a test object is assumed to give rise to a global activation of information stored in memory. The greater this global activation, the greater is the sense of familiarity associated with the object, and so the greater is the probability with which an observer judges the object to be *old* (Gillund & Shiffrin, 1984; Hintzman, 1988; Metcalfe Eich, 1982; Murdock, 1982).

In the present research, we are primarily concerned with forms of perceptual recognition, involving stimuli such as colors, faces, dot patterns, and geometric forms. In the domain of perceptual recognition, one member of the class of global-activation models is the exemplar-based generalized context model (GCM; Nosofsky, 1988, 1991). According to the GCM, people store individual exemplars of study items in memory. The exemplars are represented as points in a multidimensional perceptual space, and similarity between exemplars is a decreasing function of distance in the space. According to the model, the observer makes old–new recognition judgments by summing the similarity of test items to the individual exemplars stored in memory. This summed similarity provides the measure of overall familiarity that serves as a basis for the observer's old–new decisions. Previous research has demonstrated that, at least in fairly simple perceptual domains, the GCM can successfully predict fine-grained differences in old–new recognition probabilities as a function of fine-grained differences in summed similarities among items (Nosofsky, 1988, 1991; Shin & Nosofsky, 1992; Zaki & Nosofsky, 2001; for closely related work, see Brockdorff & Lamberts, 2000; Estes, 1994; Lamberts, Brockdorff, & Heit, 2002; K. J. Lee, Byatt, & Rhodes, 2000).

Most of the success of the GCM has involved predicting how false-alarm rates associated with new items vary with their similarity to old items. A robust finding from the perceptual recognition literature is that as the similarity of new items to old study items increases, false-alarm rates to these new items increase. The GCM successfully predicts, in quantitative detail, this pattern of false-alarm rates.

A potential problem associated with the GCM, however, concerns results involving hit rates to old items. In the present research, our key question concerns how hit rates to old items may vary with the distinctiveness of the old items. *Distinctiveness* is a fundamental construct in memory research, but the construct is open-ended and multifaceted (for a review, see Schmidt, 1991). In general, it seems to entail the extent to which an item stands out from other items in the set. In the initial part of the present article, we operationalize the construct in terms of how similar individual items are to other items in the study set. Specifically, we define *typical* items as those that are highly similar to other experienced items, whereas *distinctive* items are highly dissimilar to other experienced items.

In various experimental paradigms, the effects of distinctiveness on memory for old items are fairly clear-cut. For example, in tasks of free recall, a distinctive item embedded in a list of otherwise homogeneous items has a significant memorial advantage (Hunt, 1995). Likewise, consider tasks of perceptual identification, in which observers need to learn to assign unique labels to each member of a set. In such tasks, identification accuracy is superior for distinct items lying in isolated regions of a similarity space than for nondistinct items lying in dense regions (Lockhead, 1970). Such results are consonant with the predictions from exemplar-based models of memory and perceptual classification. For exam-

Robert M. Nosofsky, Department of Psychology, Indiana University Bloomington; Safa R. Zaki, Department of Psychology, Williams College.

This work was supported by National Institute of Mental Health Grant R01 MH48494. We would like to thank Michael Lee and three anonymous reviewers for their helpful criticisms of a draft of this article.

Correspondence concerning this article should be addressed to Robert M. Nosofsky, Department of Psychology, Indiana University, Bloomington, Indiana 47405. E-mail: nosofsky@indiana.edu

ple, according to the search of associative memory (SAM) model (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981), recall takes place by using cues to sample individual item traces from memory. Because distinctive items receive far less competition than do nondistinct items in this sampling process, the model naturally predicts a distinctiveness advantage in free recall. Likewise, according to the GCM (Nosofsky, 1986), perceptual identification accuracy is determined by the similarity of an item to its own memory trace compared with its similarity to all competing memory traces. Because distinctive items receive far less competition in this similarity-comparison process, their identification accuracy is enhanced.

In tasks such as recall and identification, the goal involves gaining access to unique memory traces associated with the individual objects. The analysis of old–new recognition is different, however, because performing this task does not require such access. Instead, according to the global-familiarity models, recognition judgments are mediated by a global activation of all traces stored in memory, not by the sampling and recovery of individual traces.

Indeed, on the basis of its summed similarity rule, the GCM predicts that hit rates associated with typical old items will be *higher* than hit rates associated with distinctive old items, not the reverse. According to our present definition, typical old items have greater summed similarity than do distinctive old items, so typical items will give rise to a greater sense of familiarity. There is evidence, however, that the reverse result may sometimes hold, whereby distinctive old items have higher hit rates than do typical ones.

Much of this evidence arises from the face-recognition literature, where it is often reported that distinctive faces are recognized with higher probability than are typical faces (e.g., Bartlett, Hurry, & Thorley, 1984; Light, Kayra-Stuart, & Hollander, 1979; Valentine and Ferrara, 1991; Vokey & Read, 1992). A good example is found in a recent face-recognition study reported by Busey and Tunnicliff (1999). These researchers had subjects study a large set of pictures of faces. Extensive similarity-scaling work was performed to locate these faces as points in a multidimensional similarity space. An intriguing aspect of Busey and Tunnicliff's results was that distinctive old faces (i.e., those located in isolated regions of the multidimensional-similarity space) had higher hit rates than did typical old faces. Furthermore, as would be expected on the basis of this result, the GCM failed to provide a good quantitative account of these face-recognition data.

However, as acknowledged by Busey and Tunnicliff (1999), other aspects of the faces could have influenced these results besides their degree of isolation in the scaled space. For example, in their study, faces located in isolated regions of the similarity space all tended to have beards, whereas faces located in dense regions did not. Thus, degree of isolation in the space was confounded with various other properties of the individual objects themselves.

Therefore, to study more systematically the potential effects of typicality and distinctiveness on both false-alarm rates to new items and hit rates to old items, Zaki and Nosofsky (2001) conducted a study in which these variables were manipulated experimentally across conditions. The design of one of their experiments is illustrated schematically in Figure 1. (The stimuli were Munsell colors varying in hue, saturation, and brightness.) There were three critical regions (A, B, and C) of old study items, represented by the dotted triangles in the figure. All items within a region were highly similar to one another, whereas items from different regions were dissimilar. In each condition, one, two, or three items from a region were presented during study. For example, in Condition 1, one item was presented from Region A, two items were presented from Region B, and all three items were presented from Region C. Thus, in this condition, the item from Region A was a distinctive old item, whereas the items from Region C were typical. The number of items presented from each region was rotated across conditions so as not to confound the variables of typicality and distinctiveness with other properties of the individual objects themselves. Finally, at time of test, all items illustrated in the figure, both old and new, were presented for recognition judgments.

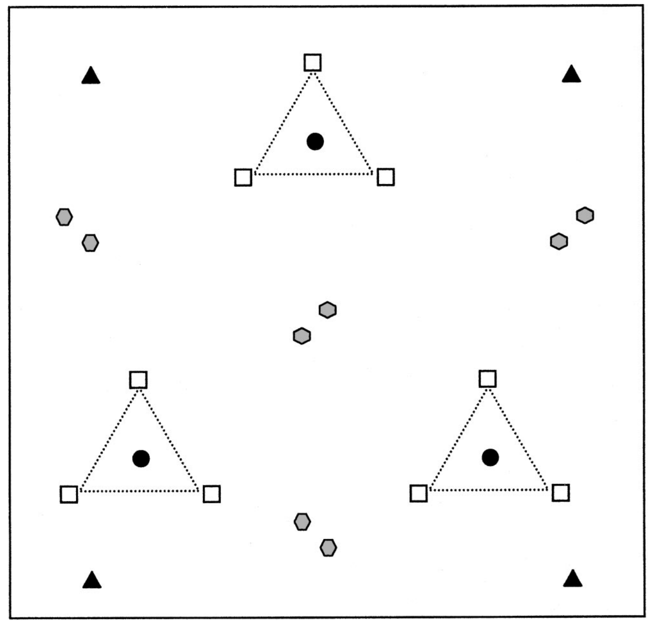

Figure 1. Schematic design of the *distinctiveness* experiment tested by Zaki and Nosofsky (2001). Open squares represent potential old items; solid circles represent prototypes; shaded hexagons represent related items; and solid triangles represent distractor items.

As is almost always observed in tests of perceptual old–new recognition, Zaki and Nosofsky (2001) found that false-alarm rates associated with new items increased as their typicality increased. For example, in Condition 1 described above, false-alarm rates associated with new objects from Region C were higher than false-alarm rates associated with new objects from Region A. The more critical finding concerned the results for old items. In apparent contrast to the results observed from the face-recognition literature, hit rates associated with distinctive old items were *not* higher than hit rates associated with typical old items. Although the hit rates for typical old items were only slightly higher than those for distinctive ones, Zaki and Nosofsky found that the GCM provided good quantitative fits to the complete sets of data. That is, the model predicted well the magnitude of both the false-alarm and hit rates as a function of the typicality manipulations. Therefore, Zaki and Nosofsky concluded that when properties of individual objects were held fixed, the GCM predicted correctly the effects of typicality and distinctiveness on false-alarm and hit rates.

Shiffrin, Huber, and Marinelli (1995) reported an old–new recognition study with results that were similar to those observed by Zaki and Nosofsky (2001; albeit in a study involving memory for words rather than perceptual stimuli). In their experiment, subjects studied a single long list of words, with the words coming from various categories that varied in size (small: two words per category; medium: six words; and large: nine words). At time of test, Shiffrin et al. presented subjects with two old words from each category, two new words, and the category prototype. Parallel to the result found by Zaki and Nosofsky, Shiffrin et al. observed that false-alarm rates to new items and to the prototype were substantially higher for the large-size categories than for the small-size ones. In addition, parallel to the result from Zaki and Nosofsky, Shiffrin et al. observed that hit rates for old items from small-size categories (i.e., distinctive old items) were *not* higher than hit rates to old items from large-size categories. Furthermore, the global-activation SAM model (Gillund & Shiffrin, 1984) provided a good quantitative account of these old–new recognition results.

Comparing the results of Shiffrin et al. (1995) and Zaki and Nosofsky (2001) with those from the face-recognition literature therefore leads to an interesting question: Why are hit rates for distinctive faces often greater than hit rates for typical ones, whereas the reverse result was obtained in the above cited experiments involving colors and words?

In the present research, we investigate two main hypotheses to shed light on this question. The first hypothesis, which we pursue in Experiment 1, is that the contrasting results may reflect procedural differences in the structure of the test situations used in these types of experiments (Zaki & Nosofsky, 2001, p. 1036). The second hypothesis, which we pursue in Experiment 2, is that distinctiveness effects in recognition arise not simply when items lie in an isolated region of the similarity space but rather when they possess unique, individuating properties as well. To anticipate, we find that some combination of these factors seems to be involved, and we develop and test an extended version of the GCM to account for these distinctiveness effects.

Experiment 1

In standard face-recognition studies, such as the one reported by Busey and Tunnicliff (1999), a sample of faces with differing properties is generated, and some faces are randomly selected as old and others as new. Furthermore, some of the faces happen to lie in dense regions of the similarity space (i.e., they are typical faces), whereas others lie in isolated regions (i.e., they are distinctive). At time of test, all faces are then presented for old–new recognition judgments. Note that by using this procedure, there are likely to be many more foils that are highly similar to target faces from typical regions than to target faces from distinctive regions. The reason is that the same naturally occurring distribution that produces the old faces is also being used to produce the new ones. Thus, in many face-recognition studies, the nature of the test situation tends to be correlated with that of the study situation. If a highly distinctive face is presented at time of study, there are unlikely to be any foils that are similar to that face at time of test; whereas typical faces presented during study are likely to have highly similar foils at time of test.

By contrast, in the designs conducted by Shiffrin et al. (1995) and Zaki and Nosofsky (2001), different testing situations were used. In Zaki and Nosofsky's experiment, subjects were tested in a uniform manner with all colors from all regions, regardless of which particular colors were presented at time of study. Likewise, in Shiffrin et al.'s experiment, subjects were always tested on two old words, two new ones, and the prototype, regardless of the size of the category that was experienced during the study phase.

Conceivably, in the standard face-recognition studies, subjects may gain some metaknowledge of the structure of the test situation and use this knowledge to influence their recognition judgments. For example, an observer may come to realize that he or she is never being tested on foils that are similar to the distinctive target faces presented during the study phase. Thus, an observer would optimize performance by responding "old" to any test face that seemed at all similar to a distinctive study face. The participant would need to set a higher criterion, however, for responding "old" to test faces that seemed similar to typical study faces.

The purpose of Experiment 1 was to pursue this hypothesis by systematically manipulating the structure of the test situation used to assess old–new recognition. In all conditions, subjects studied a list of objects in which the typicality and distinctiveness of individual objects were varied. At the time of the test, subjects were presented with some of the old objects as well as new ones. In one condition, the test situation was designed to mimic the standard face-recognition paradigm: Highly similar stimuli were never presented as foils for distinctive study items, only for typical ones. The other two conditions paralleled the test situations used by Zaki and Nosofsky (2001) and Shiffrin et al. (1995). To obtain generality of the results, the same basic design was used in three different stimulus domains: colors, faces, and words. We hypothesized that a distinctiveness effect (i.e., higher hit rates for distinctive objects than for typical ones) would be observed only in the test situation in which similar items never served as foils for distinctive targets.

Design

The general experimental design was similar for the three stimulus domains. The sets of objects were organized into categories, in which objects within categories were similar to one another, whereas objects between categories were relatively dissimilar. For example, in the domain of colors, the categories were different Munsell hues. Within each hue, the colors varied slightly in brightness and saturation, thereby producing a set in which within-category objects were similar and between-category objects were dissimilar. During the study phase, half the categories were designated as *dense*, and half were designated as *sparse.* Numerous old study items were presented from each dense category (typical old items), whereas a single old study item was presented from each sparse category (distinctive old items).

The structure of the test situation was manipulated as a between-subjects variable. In the complete-set testing condition (analogous to the method used by Zaki & Nosofsky, 2001), subjects were tested on all items from all categories. In the fixed-subset testing condition (analogous to the method used by Shiffrin et al., 1995), subjects were tested on a single old item and either one or two new items from each category, regardless of whether the study category was dense or sparse. And in the correlated-subset testing condition, subjects were tested on all items from the dense categories (including both old and new items), but on only the single old item from the sparse categories. Thus, in this latter condition, similar foils were presented only for typical target items, not distinctive ones.

Method

Subjects

There were 78 subjects in the color experiment, 94 in the word experiment, and 97 in the face experiment. All subjects were undergraduate students at Indiana University. A roughly equal number of subjects in each stimulus-domain experiment participated in the complete-set, fixed-subset, and correlated-subset test conditions (details are provided in Table 1). Subjects in the word and face experiments received partial credit toward an introductory psychology course requirement, whereas subjects in the color experiment were paid $6 for their participation.

Stimuli

In the color experiment, the stimuli were seven colors from each of six hue regions defined by the Munsell system. The six hues were blue, red-purple, yellow-red, green-yellow, green, and purple. The colors within each hue region varied in brightness and saturation. The configuration of the colors in each individual hue region is shown in Figure 2. The color at the centroid of the configuration is defined as the prototype of the hue category. The stimuli were produced by scanning the Munsell color chips into a computer. This scanning procedure has been shown in previous work to roughly maintain the relative perceptual distances among colors in the Munsell space (Zaki & Nosofsky, 2001). The complete list of hue, brightness, and saturation values for the 42 colors, together with the corresponding red–green–blue (RGB) values resulting from the computer scans, is provided in Appendix A. Each stimulus was presented as a 2-in. square on a white background.

In the word experiment, there were 15 words from each of 12 categories. The categories were fish, flower, gem, furniture, vehicle, weapon, tool, bird, sport, toy, clothing, and vegetable. A complete list of the words from each category is provided in Appendix B. The lists of words were adapted from the stimuli used by Rosch (1975). In particular, 1 of Rosch's original categories was dropped because it had high similarity to another category, and 3 new categories were added. From each of the categories, we chose the 15 most frequent items to help ensure that the words would be familiar to subjects in the experiment. The prototypes of the categories from Rosch were provided in her article. For each of the three newly created categories, we chose a single word that we deemed to be the most typical, although this choice of category prototype is not critical to the goals of the experiment. The prototype of each category is identified in Appendix B.

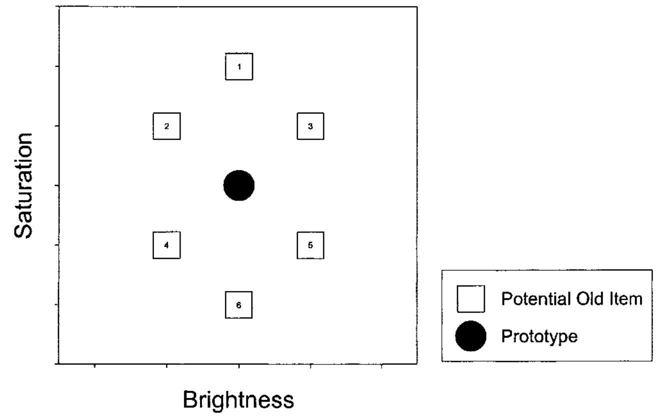

Figure 2. Configuration of Munsell colors within each hue region tested in the color-domain condition of Experiment 1 and in each of the conditions of Experiment 2. Within each hue region, the colors varied only in brightness and saturation, with stimulus spacings as illustrated in the figure. Open squares represent potential old items; solid circles represent prototypes.

In the face experiment, the stimuli were black-and-white yearbook photographs adapted from a set of stimuli used by Criss and Shiffrin (2000). There were 12 categories of faces, with 12 items per category. The categories differed primarily in terms of race, gender, hair color, hair length, or other salient features. In the face experiment, we did not identify any single item as corresponding to the category prototype.

Procedure

In all three stimulus-domain experiments, there was a study phase followed by a test phase. During the study phase, the old items were

Table 1
Procedure in the Color, Word, and Face Versions of Experiment 1

	Color			Word			Face		
Design variable	Comp	Fixed	Corr	Comp	Fixed	Corr	Comp	Fix	Corr
Subjects	26	27	25	30	32	32	32	32	33
Study blocks	4	4	4	1	1	1	1	1	1
Regions	6	6	6	12	12	12	12	12	12
Total stimuli per region	7	7	7	15	15	15	12	12	21
Study items per region									
Dense regions	4	4	4	10	10	10	7	7	7
Sparse regions	1	1	1	1	1	1	1	1	1
Test items per region									
Dense region old items	4	1	4	10	1	10	7	1	7
Dense region foils	2	1	2	4	1	4	5	1	5
Dense region prototype	1	1	1	1	1	1	0	0	0
Sparse region old items	1	1	1	1	1	1	1	1	1
Sparse region foils	5	1	0	13	1	0	11	1	0
Sparse region prototype	1	1	0	1	1	0	0	0	0

Note. Comp = complete-set test condition; Fixed = fixed-subset test condition; Corr = correlated-subset test condition.

presented on a computer screen in random order, and subjects were instructed that their memory for the items would be tested. During the subsequent test phase, both old and new items were presented on the computer screen. Subjects judged whether each item was old or new and supplied a confidence judgment on a 5-point scale, ranging from 1 *(guessing)* to 5 *(extremely confident).* The order of presentation of the test items was randomized for each subject.

In the following section, we describe the procedure for the color experiment in detail. Because the procedures for the word and face experiments were directly analogous to the color experiment, we only briefly summarize them. Details of the procedure for all three experiments, however, are provided in Table 1. One difference among the three stimulus-domain experiments is that a differing number of study items was used. Preliminary work indicated that the colors were the hardest stimuli to hold in memory, so this experiment used the smallest number of study items; whereas the words were the easiest stimuli to memorize, so this experiment used the largest number of study items.

Color experiment. For each subject, three hue categories were randomly selected as dense, and the other three hue categories were designated as sparse. The old study items consisted of four colors from each dense hue category and one study item from each sparse hue category, for a total of 15 old study items. The old study items from each hue category were randomly selected for each subject, except that the prototype never served as an old study item. In the complete-set test condition, subjects made old–new recognition judgments for all items from all categories. Thus, for each dense category, there were 4 old study items, 2 foils, and 1 prototype; whereas for each sparse category, there were 1 old study item, 5 foils, and 1 prototype. Therefore, this condition had a total of 15 old items and 27 new ones. In the fixed-subset test condition, there was 1 old study item, 1 foil, and 1 prototype from each category, regardless of whether the study category was dense or sparse. The old study items and foils from each category were randomly selected for each subject. Thus, this test condition had a total of 6 old study items and 12 new ones. In the correlated-subset test condition, all items were presented from the dense categories, but only the single study item was presented from each sparse category. Thus, there were 4 old items, 2 foils, and 1 prototype from each dense category, and a single old item from each sparse category, for a total of 15 old items and 9 new ones.

Word experiment. For each subject, six word categories were randomly selected as dense, and the other six word categories were designated as sparse. The old study items consisted of 10 randomly selected words from each dense category and 1 randomly selected word from each sparse category, for a total of 66 old study items. The prototype never served as an old study item. The structure of the complete-set, fixed-subset, and correlated-subset test conditions is reported in Table 1.

Face experiment. For each subject, six face categories were randomly selected as dense and the other six were designated as sparse. The old study items consisted of seven randomly selected faces from each dense category and one randomly selected face from each sparse category, for a total of 48 old study items. The structure of the complete-set, fixed-subset, and correlated-subset test conditions is reported in Table 1.

Results

The mean probability with which subjects judged the different types of items to be old across the three stimulus-domain experiments and conditions of testing is reported in Table 2. The patterns of results were similar across the three stimulus domains, so we describe them in general terms rather than for each stimulus domain separately.

We start by describing the results for the new items. First, inspection of the results from the fixed-subset and complete-set testing conditions reveals that false-alarm rates to new items increased as their typicality increased. That is, false-alarm rates were greater for new items drawn from dense categories than for new items drawn from sparse categories. (Recall that no new items were drawn from sparse categories in the correlated-subset testing condition, so this comparison is not available in that condition.) Second, false-alarm rates to the prototypes were higher than to the other new items. (Prototypes were not presented in the face experiment, however, so this comparison is not available in that experiment.) The increasing false-alarm rates as a function of category density and prototypicality were as predicted by global-activation models of recognition memory. In particular, as category density and prototypicality increase, the summed similarity of test items to the stored exemplars increase, so global-familiarity models predict increased rates of responding “old.”

The critical results pertain to the old items. First, averaged across all three experiments and conditions of testing, hit rates to old items from dense categories (typical old items) slightly exceeded hit rates to old items from sparse categories (distinctive old items; .780 vs. .756). Thus, as was found in the experiments reported by Zaki and Nosofsky (2001) and Shiffrin et al. (1995),

Table 2
Mean Recognition Probabilities for the Different Item Types as a Function of Test Condition and Category Density

	Test condition					
	Complete set		Fixed subset		Correlated subset	
Experiment and item type	Sparse	Dense	Sparse	Dense	Sparse	Dense
Color experiment						
New	.35	.59	.22	.48		.50
Prototype	.45	.74	.37	.69		.65
Old	.76	.79	.86	.88	.84	.79
Word experiment						
New	.13	.23	.05	.17		.19
Prototype	.16	.26	.08	.23		.24
Old	.68	.79	.81	.82	.75	.78
Face experiment						
New	.18	.28	.12	.24		.19
Old	.74	.75	.74	.76	.62	.66

there was no overall evidence of a distinctiveness effect in which distinctive old items had higher hit rates than typical old items.

The most important question concerned whether the relative hit rates to typical and distinctive old items depended on the structure of the testing conditions. We found little or no evidence in favor of this hypothesis. Although the results were in the predicted direction in the color experiment, the interaction between testing condition and category size on old-item hit rates was not statistically significant, $F(2, 75) = 1.56$, $p = .22$. Likewise, this key interaction was not statistically significant in the word experiment, $F(2, 91) = 1.88$, $p = .16$; or in the face experiment, $F(2, 94) = 0.32$, $p = .73$. Averaged across the three stimulus domains, mean hit rates to typical-old and distinctive-old items were .78 and .73, respectively, in the complete-set testing condition; .82 and .81 in the fixed-subset testing condition; and .74 and .74 in the correlated-subset testing condition. At least under the present conditions, therefore, any influence of the structure of the test situation on a distinctiveness effect for old items was either absent or fairly small.

One concern that might be raised is that the proportion of old versus new items varied across the three testing conditions (complete set, fixed subset, and correlated subset). It would be reasonable to expect that such a factor would influence the overall bias for responding "old" versus "new" in each condition. However, the critical question concerned the relative rate of responding "old" for distinctive (sparse) versus typical (dense) targets *within* each testing condition. Any global change in response bias would affect hit rates for typical versus distinctive targets in a similar manner. Thus, if there were a distinctiveness advantage in any given testing condition, it would have appeared regardless of the overall level of response bias operating in that condition.

Another concern is that the lack of a distinctiveness advantage in any given testing condition might simply have reflected a lack of statistical power. First, note that averaged across the three stimulus domains, the results never even went in the direction of a distinctiveness advantage. In addition, we conducted a power analysis based on the standard error of the difference scores (hit rates for sparse vs. dense items) computed from the correlated-subset testing condition. We focused on the correlated-subset condition because there we hypothesized that a distinctiveness advantage would most likely be observed. Pooling across the three stimulus domains and assuming an effect size (i.e., a true distinctiveness advantage) of .10, the power of the test was .99; for an assumed effect size of .05, the power of the test was .60. Thus, our finding that the overall hit rates for distinctive versus typical items were virtually identical in the correlated-subset testing condition did not stem from a lack of statistical power.

Most important, in Experiment 2, we show that with a somewhat different manipulation but using the same population of subjects, stimulus materials, and an extremely similar design, large and significant effects of distinctiveness are observed. We do not interpret the null effects in the present experiment to mean that the structure of the test situation has no influence on whether old-item distinctiveness effects are observed. Rather, the results from the combined experiments will suggest that there is a powerful joint influence of multiple factors that give rise to old-item distinctiveness effects.

Discussion

In Experiment 1, we pursued the hypothesis that the structure of the test situation exerts an impact on whether a hit-rate advantage will be observed for distinctive old items. Specifically, we hypothesized that subjects may gain some metaknowledge of the structure of the test situation. In situations in which similar stimuli never serve as foils for distinctive old items, subjects may learn to set a low criterion for responding "old" to such items. However, across three separate stimulus-domain experiments, we failed to find any strong evidence for such an effect. Although this null result does not indicate that the structure of the test situation is irrelevant, it does suggest that other factors may also be at work when substantial effects of distinctiveness on old-item recognition are observed. The purpose of Experiments 2A–2C was to investigate one such potential factor.

Experiment 2A

In the previous experiment, we conceptualized *distinctiveness* solely in terms of the degree to which an old item resided in an isolated portion of a similarity space. There are other reasons why an item might stand out from others in the set, however. Recall that in Busey and Tunnicliff's (1999) experiment, for example, objects that were scaled in isolated regions of the similarity space tended to have certain stimulus-specific properties, such as beards, whereas typical objects did not. Conceivably, at time of test, an observer might be remembering the stimulus-specific property and using it as a basis for recognition, which could potentially explain elevated hit rates for distinctive items. Furthermore, as is explained below, the current version of the GCM has no mechanism for capturing the potential influence of discrete, individuating properties associated with objects. This limitation could explain the model's failure to predict the hit-rate advantage for distinctive old items in Busey and Tunnicliff's study.

The main purpose of our remaining experiments was to explicitly manipulate the presence of discrete, individuating features of objects and test for effects of this manipulation on perceptual old–new recognition judgments. We used as stimuli a set of Munsell color patches. As was the case in Experiment 1, the colors were organized into different hue categories. Within each hue, colors varied in brightness and saturation. Colors of the same hue were highly similar to one another, whereas colors of different hues were relatively dissimilar. The key manipulation was that certain discrete features were added to a small number of the color patches. Specifically, one old color patch had a plus sign (+) drawn at its center, a second color patch had the letter M, and a third color patch had the symbol >>>. Thus, in these experiments, distinctive items were defined as those that had one of the discrete, individuating features.

In Experiment 2A, if one of these discrete features appeared on an old object at time of study, then it also appeared on that object at time of test. Furthermore, these discrete features did not appear on any foils. In Experiments 2B and 2C, we conducted manipulations that were similar to those of Experiment 1 by varying the structure of the test list. Specifically, we varied the extent to which these discrete features also appeared on foils that were similar to the distinctive targets. Because the ensuing experiments are identical to one another except for manipulations in the structure of the test list, they can be conceptualized as different conditions within a single experiment, and so we occasionally report some cross-experiment comparisons.[1]

Method

Subjects

The subjects were 61 undergraduates from Indiana University, who participated as part of a course requirement. To increase motivation, we offered a $15 reward to the 2 subjects with the highest accuracy in the experiment.

Stimuli

The stimuli were the same color patches used in the color-domain condition of Experiment 1. Again, there were seven colors from each of six hue regions. The configuration of the colors within each region was the same as shown in Figure 2, that is, it included a prototype item surrounded by a ring of other items. In this experiment, however, each of three color patches (each one from a separate hue region) had a discrete feature drawn at its center (+, M, and >>>).

Procedure

For each subject, three of the six hue regions were chosen at random to serve as *marked* regions. Within each marked region, a single color (not including the prototype) was chosen at random to serve as a *distinctive* item. One of the three discrete features was added to this item. (A different discrete feature was used for each marked region.) We refer to the remaining colors, without the discrete features, as *typical* items. Also, we refer to the three hue regions that were not chosen to have a distinctive item as *unmarked* regions.

For each subject, three colors from each of the six hue regions were randomly chosen to serve as old study items, with the constraint that the distinctive colors were always old. The remaining four colors from each region (including the prototype) were new. Thus, there was a total of 18 old items (3 from each of the 6 hue regions), and 3 of these old items were distinctive. There was a total of 24 new items (3 typical new items as well as the prototype from each hue region).

Before the experiment began, subjects were instructed that some of the color patches would have markings placed on them and that they might be able to use these markings to help remember the items. In all other respects, the procedure was the same as already described for the complete-set testing condition of Experiment 1.

Results

The mean probability with which each of seven main types of items was judged to be old is reported in Table 3. The seven item types are the distinctive old (DO), typical old (TO), typical foil (TF), and prototype (P) items from the marked regions; and the typical old, typical foil, and prototype items from the unmarked regions. As is evident from inspection of Table 3, hit rates for the old items exceeded false-alarm rates for the new items. Also, the prototypes had higher false-alarm rates than did the other new items. The critical result was that the DO items had significantly higher hit rates ($M = .93$) than did the TO items ($M = .77$), $t(60) = 4.73$, $p < .001$. We refer to this difference in hit rates between the DO and TO items (which in the present case equals .16) as the magnitude of the old-item distinctiveness effect.[2]

Table 3
Observed and Predicted Old Recognition Probabilities for the Seven Item Types in Experiment 2A

Regions and item type	Observed	Hybrid-similarity GCM	Standard GCM
Marked regions			
DO	.93	.93	.82
TO	.77	.77	.83
TF	.45	.43	.45
P	.51	.55	.56
Unmarked regions			
TO	.78	.78	.83
TF	.49	.48	.47
P	.63	.62	.60

Note. GCM = generalized context model; DO = distinctive old; TO = typical old; TF = typical foil; P = prototype.

Theoretical Analysis

The finding that the distinctive old items had higher hit rates than did the typical old items is perhaps not very surprising. Nevertheless, this finding provides a severe challenge to the current version of the GCM. In this section, we first describe the formal model as it is applied to the present paradigm and explain its fundamental limitation. We then go on to propose an extended version of the model that may allow it to account for the present types of distinctiveness effects in old–new recognition.

According to the GCM, the global familiarity (F_i) of item i is found by summing its similarity to all of the old study exemplars. The probability with which item i is recognized as old is then given by

$$P(\text{old}|i) = F_i/(F_i + k), \tag{1}$$

where k is a response-criterion parameter. The GCM uses a multidimensional scaling (MDS) approach to modeling similarity. Exemplars are represented as points in a multidimensional space, and similarity between exemplars is an exponentially decreasing

[1] We do not report Experiments 2A–2C as individual conditions within a single experiment because the undergraduate students were tested at different times during the semester. This difference seems like a very minor one, however. Therefore, we believe that the occasional cross-experiment comparisons, which are strongly theoretically motivated, are very reasonable ones to make.

[2] To determine whether the old-item distinctiveness effect might have arisen simply because subjects gave greater attention during study to the colors that had the individuating features, we also tested a control condition with 90 new subjects. The control condition was identical to the experimental condition in all respects, except that during the test, the individuating features were removed from all of the color patches. Subjects were instructed to base their old–new judgments on only the colors themselves, without regard to the presence or absence of the discrete, individuating features. In the control condition, average hit rates to the distinctive old and typical old items were .68 and .73, respectively (i.e., distinctive old items were recognized with slightly *worse* accuracy than were typical old items). This result suggested that the old-item distinctiveness effect in the experimental condition was not simply due to subjects giving greater attention to the distinctive old colors.

function of distance in the space (Shepard, 1987). The similarity of item i to exemplar j is given by

$$s_{ij} = \exp(-\kappa \cdot d_{ij}), \quad (2)$$

where d_{ij} is the distance between objects i and j in the space, and κ is an overall scaling parameter. In the present paradigm, the addition of the discrete features to the distinctive objects can be conceptualized as increasing their distance to the other exemplars in the space.

There are two key reasons why the GCM cannot account for the present old-item distinctiveness effect. First, the model assumes that all stimuli have equal self-similarity. More specifically, in the GCM, the similarity of an object to itself is always equal to 1. The reason is that if two stimuli match exactly on all dimensions, then their distance is equal to 0. Because similarity is an exponentially decreasing function of distance, the self-similarity is therefore equal to 1. Second, note that the summed similarity of the distinctive stimuli (i.e., those with the discrete features added) will therefore be less than the summed similarity of the typical stimuli (i.e., those without the discrete features). Although the self-similarities of the distinctive and typical stimuli are equal, the distinctive stimuli are less similar to the remaining exemplars in the study set. Thus, because summed similarity for the distinctive stimuli is less than for the typical stimuli, the standard version of the GCM fails to predict the present distinctiveness effect on recognition hit rates.

To address this challenge, we propose here an extended version of the GCM that makes use of a richer form of similarity representation. M. D. Lee and Navarro (2002; Navarro & Lee, in press) have proposed closely related models, which we describe in our General Discussion. The idea is to develop a hybrid-similarity representation that combines the MDS approach in the GCM with Tversky's (1977) classic ideas involving discrete feature matching. This extension of the GCM is strongly motivated by a huge converging literature from fields such as similarity judgment, visual search, and classification, which demonstrates the central role of discrete-feature matching in cognitive processing (Gati & Tversky, 1984; Shepard & Arabie, 1979; Treisman & Souther, 1985; Tversky, 1977).

According to Tversky's (1977) feature-contrast model (FCM), the similarity between two objects i and j is given by

$$\mathrm{S}_{\mathrm{FCM}}(i, j) = F[\theta f(I \cap J) - \alpha f(I - J) - \beta f(J - I)], \quad (3)$$

where $f(\mathrm{I} \cap \mathrm{J})$ is a measure of the features that are common to objects i and j; $f(I - J)$ and $f(J - I)$ are measures of the features that are distinctive to objects i and j, respectively; θ, α, and β are decision weights; and F is an increasing function. The aspect of Tversky's model that is key here is that it allows for stimuli to have differing degrees of self-similarity to one another. A stimulus with a highly salient feature will be more similar to itself (i.e., have greater self-match) than a stimulus without such a feature, because matching common features increase similarity (see Equation 3). If the degree of self-match is sufficiently strong, then the summed similarity of a highly distinctive item can exceed the summed similarity of more typical ones.

In the proposed hybrid model, the similarity between two objects i and j takes the general form

$$s_{\mathrm{H}}(i, j) = C \times D \times s(i, j), \quad (4)$$

where C ($C > 1$) is the boost in similarity provided by common (matching) discrete features, D ($0 < D < 1$) is the reduction in similarity caused by distinctive (mismatching) features, and $s(i, j)$ is the overall similarity between the objects in the continuous-dimension portion of the space (as computed in Equation 2).

For simplicity in denoting the continuous-dimension component of the model, let s_{w} denote the mean similarity of colors within the same hue region, let s_{p} denote the mean similarity of the prototype to other colors in its hue region, and let s_{b} denote the mean similarity among colors of different hue regions. These similarity values can be computed by calculating distances between the colors in the Munsell color space and then by applying Equation 2. (This computation requires the estimation of a single sensitivity parameter κ in Equation 2.) To illustrate the workings of the hybrid model, consider, for example, its application to a distinctive old item. The item has similarity C to itself, average similarity $s_{\mathrm{w}} \times D$ to each of the other old items in its hue region, and average similarity $s_{\mathrm{b}} \times D$ to each of the old items in the other hue regions. (For simplicity, we assume here that a single parameter D represents the similarity between the presence of a distinctive feature and its absence, as well as the similarity between two mismatching distinctive features.) Thus, in the present design, the overall summed similarity of a distinctive old item to the 18 old exemplars is given by

$$F_i = C + (2 \times s_{\mathrm{w}} \times D) + (15 \times s_{\mathrm{b}} \times D). \quad (5)$$

The probability that the item is judged to be old is then found by applying the recognition response rule (Equation 1). It is straightforward to derive analogous prediction equations for each of the other types of test items in this design.

The free parameters in this application of the hybrid model are the common-feature parameter C, the distinctive-feature parameter D, a response-criterion parameter k (from Equation 1), and an overall sensitivity parameter κ for estimating the values of s_{w} and s_{p}.[3] On the basis of preliminary model explorations, we found that the between-category similarity value s_{b} did not contribute in a substantial way to the model fits. Therefore, for simplicity, and to reduce the number of free parameters, we set $s_{\mathrm{b}} = .01$. We conducted a computer search for the values of the free parameters that minimized the sum of squared deviations (SSD) between the predicted and observed recognition probabilities across the seven item types.

The predictions from this hybrid-similarity GCM are shown along with the observed data in Table 3, with the best-fitting parameters reported in Table 4. The model accounts for 99.8% of the variance in the recognition data and captures the observed distinctiveness effect on old-item hit rates. The best-fitting value of the common-feature parameter was $C = 4.88$, indicating a strong role of self-match in this paradigm. To bring out the importance of

[3] The value of s_{p} was computed by calculating the average distance of the prototype to the six surrounding exemplars in each hue region (using the stimulus spacings illustrated in Figure 2) and then by applying Equation 2. Likewise, the value of s_{w} was computed by calculating the average distance of each of the surrounding exemplars to one another and then by applying Equation 2. More fine-grained predictions could be obtained by calculating similarities for each stimulus on the basis of its exact location in the Figure 2 configuration, but the present approach provides a reasonable approximation for purposes of getting started.

Table 4
Best-Fitting Parameters for the Hybrid-Similarity GCM in Experiments 2A–2C

Parameter	Experiment 2A	2B	2C
C	4.878	1.475	2.269
D	0.001	0.048	0.164
κ	2.280	2.069	2.284
k	0.367	0.352	0.322

Note. GCM = generalized context model; C = common-feature match; D = distinctive-feature mismatch; κ = overall sensitivity; k = response criterion.

the common-feature parameter, we also fitted a restricted version of the hybrid-similarity GCM in which C was held fixed at 1.0. The predictions from this version of the model are shown in Table 3. As can be seen, this standard version of the GCM fails to account for the old-item distinctiveness effect. Furthermore, according to the results of a general linear test, the restricted model with $C = 1$ fits the recognition data significantly worse than does the full version of the model (with C free to vary), $F(1, 3) = 31.14$, $p < .05$.

Experiment 2B

In Experiment 2B, we extended the design of Experiment 2A by including the discrete, individuating features on foils as well as old items. We had two main aims in performing this manipulation. First, by including foils with the discrete, individuating features, we obtain more rigorous tests of the quantitative predictions from the hybrid-similarity model. Second, note that this manipulation is similar to the one from Experiment 1. Specifically, the question is whether including foils that are similar to the distinctive targets will influence the magnitude of the old-item distinctiveness effect. To the extent that subjects realize that the test list contains foils with the discrete, individuating features, they may give less weight to this source of information in making their recognition judgments. Such a result would be reflected in a reduced estimate of the common-feature parameter C.

Method

Subjects

The subjects were 72 undergraduates from Indiana University, who participated as part of a course requirement. Once again, we offered a $15 reward to the 2 subjects with the best performance in the experiment.

Stimuli

The stimuli were the same as in Experiment 2A, except that the discrete features were added to various foils in addition to the three distinctive targets.

Procedure

The procedure was the same as in Experiment 2A, with the following exceptions. In each of the three marked regions, a discrete feature was placed on a randomly chosen foil. The discrete feature was the same as the one placed on the distinctive old item in that hue region. We refer to these items as the *high-similarity distinctive foils* (HSDF), because they are highly similar to the distinctive old item from their own hue region. In addition, in each of the three unmarked regions, one of the three discrete features was also placed on a randomly chosen foil. (A different discrete feature was used in each unmarked region.) We refer to these items as the *low-similarity distinctive foils* (LSDF). Although these foils contain a discrete feature that matches one of the distinctive old items, they are dissimilar to the distinctive old items because they come from separate hue regions. In sum, this design was the same as in Experiment 2A, except that among the 18 foils, there were 3 HSDF items, 3 LSDF items, and 12 typical foils (6 from marked regions and 6 from unmarked regions).

Results

The mean probability with which each of nine main types of items was judged to be old is reported in Table 5. The item types were the same as in Experiment 2A, with the addition of the HSDF and LSDF items. The general pattern of results was the same as in Experiment 2A, with higher hit rates to distinctive old (.82) than typical old items (.76), $t(71) = 1.68$, $p = .05$, one-tailed. However, as hypothesized, the magnitude of this old-item distinctiveness effect (.06) was significantly reduced compared with that observed in Experiment 2A (.16), $t(131) = 2.28$, $p < .05$. This result supports the idea that subjects were indeed sensitive to the structure of the test phase in this condition: Because various foils also included the individuating discrete features, this source of information no longer provided a perfectly reliable cue as to whether items were old or new. Thus, it appears that subjects gave less weight to the common-feature match in making their recognition judgments in Experiment 2B than in Experiment 2A.

An intriguing result was that the high-similarity distinctive foils (HSDF) had lower false-alarm rates (.28) than did the high-similarity typical foils (HSTF; .46). In other words, adding the same distinctive feature to a foil that was similar to an old item helped observers to realize that the foil was new! Consider an analogous situation in a face-recognition experiment. Suppose that a particular old face has a distinctive scar. At time of test, a new face that is similar to the old one is presented, but the scar is placed

Table 5
Observed and Predicted Old Recognition Probabilities for the Nine Item Types in Experiment 2B

Regions and item type	Observed	Hybrid-similarity GCM	Standard GCM
Marked regions			
DO	.82	.81	.76
TO	.76	.78	.79
TF	.46	.46	.48
P	.61	.60	.61
HSDF	.28	.28	.24
Unmarked regions			
TO	.80	.79	.80
TF	.54	.52	.52
P	.62	.66	.67
LSDF	.09	.09	.11

Note. GCM = generalized context model; DO = distinctive old; TO = typical old; TF = typical foil; P = prototype; HSDF = high-similarity distinctive foil; LSDF = low-similarity distinctive foil.

on this new face. Adding the scar helps the observer to realize that the face is new. Intuitively, it is as if the observer is thinking, “Wait a minute, that wasn’t the person with the scar, it was the other person!” This reduction in false-alarm rates for distinctive foils was even more dramatic for the LSDF items (.09).

It might seem that such results would present a major challenge to the proposed hybrid-similarity GCM, because adding the salient feature to the new items would greatly increase their match to the old distinctive item. As will be seen, however, the hybrid-similarity exemplar model in fact provided an excellent account of the complete set of data.

Theoretical Analysis

The prediction equations for the hybrid-similarity GCM are the same as in Experiment 2A, and the same free parameters are used. There are two additional item types, however, the HSDF and LSDF items. Consider, for example, an HSDF item. The item has similarity $C \times s_w$ to the old distinctive exemplar from its own hue region; similarity $s_w \times D$ to each of the other old exemplars in its hue region; and similarity $s_b \times D$ to each of the old exemplars in the other hue regions. Its overall summed similarity to the old exemplars is therefore given by

$$F_i = (C \times s_w) + (2 \times s_w \times D) + (15 \times s_b \times D). \quad (6)$$

Likewise, the summed similarity for the LSDF items is given by

$$F_i = (C \times s_b) + (3 \times s_w \times D) + (14 \times s_b \times D). \quad (7)$$

The predicted recognition probabilities for the nine item types are shown along with the observed data in Table 5, with the best-fitting parameters reported in Table 4. Again, the model provides an excellent fit to the data, accounting for 99.3% of the response variance. As in Experiment 2A, it accounts for the elevated hit rate associated with the distinctive old items. More interesting, it also accounts well for the reduced false-alarm rates associated with the distinctive foils (HSDF and LSDF). The reason is as follows. Although these foils match an old distinctive item on its highly salient discrete feature, they mismatch all of the other study items on this feature. The boost provided by the common-feature match is more than offset by the reduction in similarity to all the other items in the set. The LSDF items are predicted correctly to have extremely low false-alarm rates because they are highly dissimilar to all old exemplars, even to the old distinctive items with which they share the salient discrete feature.

The hybrid-similarity GCM accounts in quantitative detail for a variety of more subtle effects in this data set as well. For example, it predicts that hit and false-alarm rates associated with typical items are slightly larger in the unmarked regions than in the marked regions. The reason is that in the unmarked regions, there are three old exemplars that contribute substantially to an item’s summed similarity (i.e., the three exemplars from any given hue region); whereas in the marked regions, only two old exemplars contribute substantially to the sum. The third exemplar is a distinctive old item and so is dissimilar to all typical test items. The model also predicts the elevated false-alarm rates associated with the prototype items because these items tend to be more similar overall to the old exemplars than are the other foils.

Examination of the best-fitting parameters (Table 4) reveals a similar pattern of results as was observed in Experiment 2A. The interesting difference is that the common-feature parameter takes on a smaller value in the present condition ($C = 1.48$). As discussed previously, this result is as expected if subjects are sensitive to the structure of the foils used in the testing situation. Nevertheless, the common-feature parameter still plays a critical role. A restricted version of the hybrid model with $C = 1$ fits the data significantly worse than does the full model, $F(1, 5) = 8.29$, $p < .05$.

Experiment 2C

Our final experiment in the set was designed to be intermediate between Experiments 2A and 2B. We continued to test foils that included the discrete, individuating features of the distinctive old items. However, in this design, instead of testing three HSDF and LSDF items, we tested only one token of each of these items. If our idea is correct that the weight given to the common feature depends on the structure of the test list, then the magnitude of C will be greater in Experiment 2C than in Experiment 2B (because the discrete features provide more reliable sources of information in the present design). The prediction, therefore, is that relative to Experiment 2B, there should be increases in the hit rates to the distinctive old items and increases in false-alarm rates to the high-similarity and low-similarity distinctive foils.

Method

Subjects

The subjects were 60 undergraduates at Indiana University, who participated as part of a class requirement. We again offered a $15 reward to the 2 top performers.

Stimuli

The stimuli were the same as in Experiments 2A and 2B.

Procedure

The procedure was the same as in Experiment 2B, except there was only a single token of an HSDF and LSDF item. The same randomly selected discrete feature was used for both items.

Results

The mean recognition probability for each of the nine item types is reported in Table 6. The general pattern of results is similar to the one observed in Experiment 2B. The interesting difference is that as predicted, hit rates for the distinctive old items and false-alarm rates for the distinctive foils are systematically larger in the present design. The hit rate for the distinctive old items was significantly larger in Experiment 2C than in Experiment 2B, $t(130) = 2.55$, $p < .05$. Likewise, the average false-alarm rate for

Table 6
Observed and Predicted Old Recognition Probabilities for the Nine Item Types in Experiment 2C

Regions and item type	Observed	Hybrid-similarity GCM	Standard GCM
Marked regions			
DO	.92	.88	.80
TO	.75	.79	.82
TF	.47	.46	.48
P	.63	.60	.62
HSDF	.35	.37	.30
Unmarked regions			
TO	.79	.80	.82
TF	.53	.50	.51
P	.61	.65	.66
LSDF	.20	.19	.24

Note. GCM = generalized context model; DO = distinctive old; TO = typical old; TF = typical foil; P = prototype; HSDF = high-similarity distinctive foil; LSDF = low-similarity distinctive foil.

the distinctive foils was marginally larger in the present design, $t(130) = 1.95$, $p = .05$.[4]

Theoretical Analysis

The predicted recognition probability for each item type is reported along with the observed data in Table 6. The hybrid-similarity GCM accounts for 98.0% of the variance in the data. The main limitation of the model is that it somewhat underpredicts the magnitude of the old-item distinctiveness effect. Nevertheless, the model does provide a reasonable overall account of the increased hit and false-alarm rates associated with the distinctive items (relative to those observed in Experiment 2B). Furthermore, the quantitative fit of the model is again significantly better than that of the standard GCM, $F(1, 5) = 12.25$, $p < .05$. The best-fitting parameters, reported in Table 4, confirm that as hypothesized, the value of the common-feature match parameter (C = 2.27) is intermediate between the estimates obtained in Experiments 2A and 2B.

To test if the variation in the common-feature match parameter across Experiments 2A–2C was statistically significant, we fitted a restricted version of the hybrid-similarity GCM to the data. In this restricted version, we constrained the value of C to be constant across Experiments 2A–2C while allowing all of the other model parameters to vary freely. This restricted model fit the Experiment 2A–2C data significantly worse than did the full version, $F(2, 13) = 4.49$, $p < .05$, providing further evidence that the common-feature match parameter varied systematically across the three test conditions.

General Discussion

The presence of old-item distinctiveness effects provides a potential challenge to global-familiarity models of perceptual old–new recognition. As argued, for example, by Valentine and Ferrara (1991), it would appear that the summed similarity of typical old items would exceed the summed similarity of distinctive old items. Therefore, if summed similarity provides the basis for an observer's recognition judgments, as posited in various global-familiarity models such as the GCM, then typical old items would be recognized with higher probability than distinctive old items. Reported findings that distinctive old items have a recognition advantage therefore provide a clear challenge to such models.

However, the construct of *distinctiveness* is an open-ended one, and it is unclear what are the conditions under which such distinctiveness effects are observed. In the present research, we started by operationalizing *distinctiveness* in terms of the degree of isolation of an object in the similarity space of other studied objects. In agreement with some past work (Shiffrin et al., 1995; Zaki & Nosofsky, 2001), in Experiment 1 of this article, we found that old items located in isolated regions of the similarity space were not recognized with higher probability than were more typical, densely located objects. Furthermore, the absence of such old-item distinctiveness effects was even observed in conditions in which high-similarity items never served as foils for the distinctive targets. Apparently, something more seems to be at work in situations that give rise to substantial old-item distinctiveness effects in perceptual recognition.

Our key hypothesis is that distinctive items stand out from other items in the set not simply because they lie in isolated regions of the similarity space, but because they may possess certain individuating features that take them outside of the similarity space of the other items. Furthermore, global-familiarity models such as the GCM, which embed stimuli as points in a homogeneous, continuous-dimension space, are not well equipped to represent the effects of such discrete, individuating features. For example, in the face-recognition studies reported by Busey and Tunnicliff (1999), a six-dimensional scaling solution was used to represent similarities among a large set of faces. Naturally, faces that possessed highly individuating features, such as scars or beards, would tend to be located in isolated regions of such a continuous-dimension space. However, the presence of these discrete individuating features is not explicitly represented in this type of continuous-dimension scaling solution.

In Experiment 2 of the present research, we explicitly manipulated whether objects contained these types of discrete, individuating properties. Furthermore, we tested an extended version of the GCM with a richer, hybrid-similarity representation that is sensitive to such properties. The key idea in this hybrid representation is that the presence of highly salient, common features across objects can provide a strong boost to the objects' psychological similarity. Across three experimental conditions, we found that the presence of such discrete, individuating features did indeed have a strong impact on hit rates associated with old target items. Furthermore, our preliminary tests demonstrated some strong initial successes for the predictions of the proposed hybrid-similarity GCM.

[4] In a more detailed analysis, we examined whether hit rates to distinctive old items varied depending on which specific HSDF and LSDF items were included in the test list. Note that one of the distinctive old items matches the HSDF and LSDF items on its individuating feature, whereas the other two distinctive old items do not share their individuating features with the foils. The distinctive old item with the matching feature had an average hit rate of .88, whereas the distinctive old items with the mismatching features had an average hit rate of .93, $t(59) = 1.14$, $p > .10$. The results go in the direction of the idea that people are sensitive to the specific matching feature of the distinctive foils, but the effect is a weak one.

Beyond predicting the hit-rate advantage associated with the distinctive targets, the hybrid-similarity model also successfully predicted the decrease in false-alarm rates associated with the distinctive foils. Note that this phenomenon provides an example of the *mirror* effect in recognition memory, in which certain experimental manipulations simultaneously give rise to increased hit rates yet decreased false-alarm rates (Glanzer, Adams, Iverson, & Kim, 1993). The hybrid-similarity GCM can successfully account for the present mirror effect because the discrete, individuating features exert competing influences on summed similarity. The presence of common features across objects results in boosts to summed similarity, whereas the presence of mismatching features results in reductions to summed similarity.

Note, however, that the hybrid-similarity model does not predict that a mirror effect is an automatic consequence of a manipulation of the discrete features. Rather, its occurrence is predicted to depend on a variety of parametric considerations, such as the degree of attention that observers give to the individuating features as well as the level of similarity between the targets and foils.

For example, if foils with a distinctive feature are made sufficiently similar to the distinctive targets, then by necessity, the model would predict elevated false-alarm rates for those items as well. This reasoning led us to conduct the following more fine-grained analysis of our data. We divided the high-similarity distinctive foils into two groups. The first group consisted of *very*-high-similarity distinctive foils (VHSDF), which were the closest possible distance from the distinctive old item in their hue region. (In particular, in Figure 2, Item Pairs 1–2, 1–3, 4–6, and 5–6 satisfy this closest-possible-distance relation.) The second group consisted of all remaining distinctive foils in the same hue region, which we termed *moderate*-similarity distinctive foils (MSDF). We performed an analogous breakdown of the typical foils in each hue region, thereby creating a group of very-high-similarity typical foils (VHSTF) and a group of moderate-similarity typical foils (MSTF). Thus, a VHSTF item is a typical foil that is the closest possible distance to the distinctive old item in its hue region.

The mean false-alarm rates for these item types are reported in Table 7. (We analyzed only the Experiment 2B data because it was the only experiment with a sufficient number of observations to conduct the breakdown of the HSDF items.) The table also reports the predictions from the hybrid-similarity GCM, with all parameters held fixed from the previous modeling analyses. The key qualitative result that is correctly predicted by the model is that whereas the MSDF items have lower false-alarm rates than do the MSTF items, the VHSDF items do not have lower false-alarm rates than do the VHSTF items. In other words, when a distinctive foil becomes sufficiently similar to a distinctive target, the mirror effect disappears. Future research is needed to test more rigorously the hybrid model's prediction of a complete reversal of the mirror effect with increases in similarity of the distinctive foils to the distinctive targets.

Finally, note that the hybrid-similarity GCM provides an explanation for why strong old-item distinctiveness effects were observed in Experiments 2A–2C but were weak or absent in the conditions in Experiment 1. In Experiment 1, and previous experiments like it (Shiffrin et al., 1995; Zaki & Nosofsky, 2001), *distinctive* objects were defined simply as those objects located in isolated regions of the similarity space of studied items. However, such objects did not tend to possess individuating discrete properties that were not also possessed by typical objects located in dense regions. Thus, the absence of old-item distinctiveness effects in such experiments is also in accord with the predictions from the hybrid-similarity GCM.

Table 7
Observed and Predicted False-Alarm Rates for High-Similarity Distinctive and Typical Foils in Experiment 2B

Item type	Observed	Predicted
MSDF	.21	.17
MSTF	.46	.48
VHSDF	.49	.50
VHSTF	.46	.41

Note. MSDF = moderate-similarity distinctive foils; MSTF = moderate-similarity typical foils; VHSDF = very-high-similarity distinctive foils; VHSTF = very-high-similarity typical foils.

Relations to Previous Exemplar Models

The idea of extending exemplar-based models of categorization with a discrete-feature similarity representation was proposed previously by M. D. Lee and Navarro (2002). Specifically, in an elegant study, these researchers showed that when a discrete-feature representation was used in conjunction with Kruschke's (1992) exemplar-based connectionist model, known as ALCOVE, it provided a significantly better account of some classification learning data than when a continuous-dimension similarity representation was used. M. D. Lee and Navarro derived the feature-based similarity representation by fitting an additive clustering model (M. D. Lee, 2002; Shepard & Arabie, 1979) to a set of similarity-ratings data. In additive clustering, similarity is modeled in terms of the number of common features shared by objects. However, M. D. Lee and Navarro made no use of this common-feature structure in using ALCOVE to derive its predictions of classification learning. Rather, their classification-modeling application made use of only those features that were distinctive among objects. The present research complements the earlier contribution made by M. D. Lee and Navarro because it is the common-feature structure that is critical to allowing the model to account for the enhanced recognition of distinctive targets.

One question that arises is whether, instead of extending the exemplar model with the assumption of common-feature matching, alternative exemplar-based approaches may suffice to account for these data. In the GCM, the degree to which a stored exemplar is activated is determined jointly by its memory strength and by its similarity to the test item (Nosofsky, 1991; Nosofsky & Palmeri, 1997). In past applications of the GCM, the memory-strength parameters were used to represent effects of differential presentation frequency on exemplar storage and retrieval. A straightforward idea is that distinctive old items may give rise to greater memory strengths compared with typical old items. Indeed, for the present paradigms, versions of the GCM that make allowance for individual-item memory-strength parameters (with values identical to those of the common-feature match parameter) give rise to essentially the same predictions as does the hybrid-similarity model. Our view, however, is that the hybrid-similarity model provides a more natural psychological interpretation of the present data. As documented in our theoretical analyses, the value of the common-feature match parameter varied systematically across conditions of testing. Greater weight was accorded common-feature matches in situations in which they were highly diagnostic that an item was old. By contrast, the psychological interpretation of the memory-strength parameters is one of encoding strength at time of study. It is unclear why the memory strength of an encoded exemplar would vary depending on the nature of the foils included on a test list.

Another question that arises is whether the modification of the similarity rule in terms of common-feature matches would do harm to the GCM's successful predictions of classification in previous work. In particular, in numerous previous studies, the model has been applied successfully in designs in which people learn to classify sets of stimuli varying along a set of binary-valued dimensions (e.g., Medin & Schaffer, 1978). A key aspect of such designs, however, is that all stimuli are defined along the same number of dimensions. In such situations, the use of common-feature match parameters yields predictions that are formally identical to those in the standard version of the GCM (we provide a proof in Appendix C). It is only in cases in which additional features are placed on a stimulus, which make it stand out from others in the set, that the extended similarity rule would result in modified predictions from the model.

Directions of Future Research

The firmest conclusion based on the present research is that there are indeed cases of distinctiveness effects on perceptual old–new recognition that go beyond the ability of standard global-familiarity models, such as the GCM, to explain. At the same time, the extended hybrid-similarity model, which makes allowance for the idea that distinctive old items may have a greater degree of "self-match" than do typical old items, has achieved some preliminary success.

An important avenue for future work might be to contrast the predictions of the hybrid-similarity GCM with those of some alternative models of old–new recognition. One important class to consider is the set of models positing multiple processes in old–new recognition (e.g., Brainerd, Reyna, & Mojardin, 1999; Dosher, 1984; Hintzman & Curran, 1994; Jacoby, 1991; Ratcliff & McKoon, 1989; Rotello, 2000; Yonelinas, 1999). For example, according to some models, although global familiarity serves as one basis for making old–new recognition judgments, people rely on search and retrieval processes as well. Perhaps the enhanced recognition of distinctive old targets arises because it is easier to recall such items than it is to recall typical old items. Furthermore, people might use recall-to-reject processes (e.g., Brainerd, Reyna, & Kneer, 1995; Rotello & Heit, 2000) to help determine that foils containing the discrete, individuating features are new. It is an open question whether the types of recall and verbatim memory processes that are posited to operate in verbal domains operate in the present types of continuous-dimension perceptual domains as well. It is also an open question whether extant familiarity-plus-recall models would yield good quantitative accounts of the present data. In any case, our view is that the present contribution is an interesting one, because it demonstrates the possibility that a hybrid-similarity exemplar model based on a single process of global familiarity may be sufficient to account for a variety of distinctiveness effects in old–new recognition memory.

Finally, in recent work, Navarro and Lee (in press) have developed a similarity-scaling algorithm that combines continuous-dimension and discrete-feature structure in a hybrid representation. In our present research, the continuous-dimension and discrete-feature structure of the objects was obvious and was explicitly manipulated as part of the experimental design. An aim of future work might be to apply the present hybrid-similarity GCM to a more complex set of stimuli, such as naturalistic faces. In that case, it might be possible to apply the hybrid-similarity scaling algorithm developed by Navarro and Lee and use the derived representation in combination with the GCM to improve its predictions of face recognition.

References

Bartlett, J. C., Hurry, S., & Thorley, W. (1984). Typicality and familiarity of faces. *Memory & Cognition, 12,* 219–228.

Brainerd, C. J., Reyna, V. F., & Kneer, R. (1995). False-recognition reversal: When similarity is distinctive. *Journal of Memory and Language, 34,* 157–185.

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review, 106,* 160–179.

Brockdorff, N., & Lamberts, K. (2000). A feature-sampling account of the time course of old–new recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 77–102.

Busey, T. A., & Tunnicliff, J. L. (1999). Accounts of blending, distinctiveness, and typicality in the false recognition of faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1210–1235.

Criss, A. H., & Shiffrin, R. M. (2000, January). *The source of confusion in recognition memory.* Paper presented at the 25th Annual Interdisciplinary Conference, Jackson Hole, WY.

Dosher, B. A. (1984). Discriminating preexperimental (semantic) from learned (episodic) associations: A speed–accuracy study. *Cognitive Psychology, 16,* 519–555.

Estes, W. K. (1994). *Classification and cognition.* New York: Oxford University Press.

Gati, I., & Tversky, A. (1984). Weighting common and distinctive features in perceptual and conceptual judgments. *Cognitive Psychology, 16,* 341–370.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review, 100,* 546–567.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language, 33,* 1–18.

Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review, 2,* 105–112.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513–541.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44.

Lamberts, K., Brockdorff, N., & Heit, E. (2002). *Feature-sampling and random-walk models of individual-stimulus recognition.* Manuscript submitted for publication.

Lee, K. J., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: Testing the face–space framework. *Psychological Science, 11,* 379–385.

Lee, M. D. (2002). A simple method for generating additive clustering models with limited complexity. *Machine Learning, 49,* 39–58.

Lee, M. D., & Navarro, D. J. (2002). Extending the ALCOVE model of category learning to featural stimulus domains. *Psychonomic Bulletin & Review, 9,* 43–58.

Light, L. L, Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 212–228.

Lockhead, G. R. (1970). Identification and the form of multidimensional discrimination space. *Journal of Experimental Psychology, 85,* 1–10.

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

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

Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Navarro, D. J., & Lee, M. D. (in press). Combining dimensions and features in similarity-based representations. *Neural Information Processing Systems.*

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 700–708.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology*: *Human Perception and Performance, 17,* 3–27.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random-walk model of speeded classification. *Psychological Review, 104,* 266–300.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

Ratcliff, R., & McKoon, G. (1989). Similarity information versus relational information: Differences in the time course of retrieval. *Cognitive Psychology, 21,* 139–155.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104,* 192–233.

Rotello, C. M. (2000). Recall processes in recognition memory. *The Psychology of Learning and Motivation, 40,* 183–221.

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition, 28,* 907–922.

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition, 19,* 523–541.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323.

Shepard, R. N., & Arabie, P. (1979). Additive clustering representations of similarities as combinations of discrete overlapping properties. *Psychological Review, 86,* 87–123.

Shiffrin, R. M., Huber, D. E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 267–287.

Shin, H. J., & Nosofsky, R. M. (1992). Similarity scaling studies of dot pattern classification and recognition. *Journal of Experimental Psychology: General, 121,* 278–304.

Treisman, A. M., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General,* 114, 285–310.

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

Valentine, T., & Ferrara, A. (1991). Typicality in categorization, recognition and identification: Evidence from face recognition. *British Journal of Psychology, 82,* 87–102.

Vokey, J. R., & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition, 20,* 291–302.

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1415–1434.

Zaki, S. R., & Nosofsky, R. M. (2001). Exemplar accounts of blending and distinctiveness effects in perceptual old–new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1022–1041.

(*Appendixes follow*)

Appendix A

Munsell Specifications and RGB Values for the Stimuli in the Color Versions of Experiments 1, 2A, 2B, and 2C

Region	Hue	Saturation	Brightness	R	G	B
1	10B	7	6	161	209	242
1	10B	6	4	144	172	189
1	10B	4	4	84	108	121
1	10B	3	6	40	80	104
1	10B	4	8	56	111	152
1	10B	6	8	115	177	221
1[a]	10B	5	6	99	141	170
2	10RP	7	8	255	164	174
2	10RP	6	6	212	138	146
2	10RP	4	6	151	76	74
2	10RP	3	8	136	17	12
2	10RP	4	10	181	60	72
2	10RP	6	10	249	127	141
2[a]	10RP	5	8	197	98	107
3	7.5YR	8	8	255	209	127
3	7.5YR	7	10	254	165	61
3	7.5YR	5	10	176	92	0
3	7.5YR	4	8	142	74	0
3	7.5YR	5	6	167	107	42
3	7.5YR	7	6	233	174	119
3[a]	7.5YR	6	8	204	130	50
4	2.5GY	8	6	224	234	128
4	2.5GY	7	4	199	197	128
4	2.5GY	5	8	122	143	0
4	2.5GY	4	6	95	107	0
4	2.5GY	5	4	126	135	58
4	2.5GY	7	8	193	193	55
4[a]	2.5GY	6	6	151	170	58
5	10G	8	6	167	247	213
5	10G	7	4	154	204	179
5	10G	5	4	102	139	113
5	10G	4	6	60	112	72
5	10G	5	8	75	145	105
5	10G	7	8	124	215	174
5[a]	10G	6	6	113	174	140
6	5P	7	6	213	190	232
6	5P	6	4	172	154	179
6	5P	4	8	129	94	143
6	5P	3	6	97	64	98
6	5P	4	4	119	97	119
6	5P	6	8	185	158	198
6[a]	5P	5	6	154	125	165

Note. R = red; G = green; B = blue; RP = red-purple; YR = yellow-red; GY = green-yellow; P = purple.
[a] prototype.

Appendix B

Stimuli Used in the Word Version of Experiment 1

Fish	Flower	Precious stone/gem	Furniture	Vehicle	Weapon	Tool	Bird	Sport	Toy	Clothing	Vegetable
barracuda	tulip	sapphire	sofa	station wagon	pistol	hammer	sparrow	baseball	top	shirt	green bean
tuna	daffodil	ruby	couch	truck	revolver	ruler	bluejay	basketball	jack-in-the-box	dress	string bean
stingray	daisy	emerald	table	car	rifle	screwdriver	bluebird	tennis	toy	skirt	spinach
swordfish	iris	pearl	dresser	bus	switchblade	drill	canary	softball	yo-yo	blouse	broccoli
sailfish	carnation	gold	rocker	taxi	knife	nails	blackbird	canoeing	block	suit	asparagus
goldfish	lily	silver	loveseat	jeep	dagger	tape	dove	handball	marbles	slacks	corn
marlin	orchid	platinum	chest of drawers	ambulance	sword	sandpaper	lark	rugby	rattle	jacket	cauliflower
flounder	dandelion	jade	desk	motorcycle	bomb	sander	swallow	hockey	stuffed animal	coat	brussel sprouts
trout	sunflower	turquoise	bed	van	hand	level	parakeet	swimming	water pistol	sweater	squash
piranha	mum	onyx	bureau	train	bayonet	plane	oriole	track	teddy	sweatshirt	lettuce
catfish	marigold	opal	davenport	trolley	spear	file	mockingbird	boxing	rocking horse	jumper	celery
salmon	petunia	quartz	divan	bicycle	bazooka	toolbox	redbird	volleyball	ball	socks	cucumber
mackerel	violet	topaz	vanity	carriage	cannon	T-square	wren	lacrosse	jacks	parka	beets
cod	geranium	amethyst	bookcase	airplane	bow	chisel	finch	skiing	erector	pajamas	pea
shark	rose	diamond	chair	automobile	gun	saw	robin	football	doll	pants	carrot

Note. The last word in each category was designated the category prototype.

Appendix C

Proof of Model Equivalence

In this appendix we consider a paradigm in which the stimuli vary along a fixed set of M binary-valued, separable dimensions and in which people are classifying the stimuli into one of two categories, A and B (e.g., Medin and Schaffer, 1978). In application to such paradigms, the GCM reduces to the well-known context model of Medin and Schaffer (1978) (for extensive discussion, see Nosofsky, 1986). The probability that item i is classified into Category A is found by summing the similarity of i to all exemplars of Category A and then dividing by the summed similarity to both categories. The similarity between item i and exemplar a is given by an interdimensional multiplicative rule. Specifically,

$$P(A|i) = \frac{\sum_{a \in A} \prod_{m=1}^{M} s_m^{\delta_m(i,a)}}{\sum_{a \in A} \prod_{m=1}^{M} s_m^{\delta_m(i,a)} + \sum_{b \in B} \prod_{m=1}^{M} s_m^{\delta_m(i,b)}}. \qquad \text{(C1)}$$

In Equation C1, $\delta_m(i, a)$ is an indicator variable set equal to one if item i and exemplar a mismatch on dimension m; and set equal to zero if they match. The parameter s_m ($0 \leq s_m \leq 1$) measures the similarity between mismatching features on dimension m.

In the application of the hybrid-similarity GCM to such a paradigm, one would make allowance for the idea that the common (matching) features along each dimension provide a boost to similarity. Specifically, according to the hybrid-similarity model,

$$P(A|i) = \frac{\sum_{a \in A} \prod_{m=1}^{M} c_m^{1-\delta_m(i,a)} d_m^{\delta_m(i,a)}}{\sum_{a \in A} \prod_{m=1}^{M} c_m^{1-\delta_m(i,a)} d_m^{\delta_m(i,a)} + \sum_{b \in B} \prod_{m=1}^{M} c_m^{1-\delta_m(i,b)} d_m^{\delta_m(i,b)}}. \qquad \text{(C2)}$$

In Equation C2, δ_m (i, a) is the same indicator variable as defined above; c_m ($c_m \geq 1$) is the common-feature match parameter on dimension m; and d_m ($0 \leq d_m \leq 1$) is the distinctive-feature mismatch parameter on dimension m. Equation C2 can be rewritten as

$$P(A|i) = \frac{\sum_{a \in A} \prod_{m=1}^{M} c_m \left(\frac{d_m}{c_m}\right)^{\delta_m(i,a)}}{\sum_{a \in A} \prod_{m=1}^{M} c_m \left(\frac{d_m}{c_m}\right)^{\delta_m(i,a)} + \sum_{b \in B} \prod_{m=1}^{M} c_m \left(\frac{d_m}{c_m}\right)^{\delta_m(i,b)}}$$

$$= \frac{\sum_{a \in A} \prod_{m=1}^{M} \left(\frac{d_m}{c_m}\right)^{\delta_m(i,a)}}{\sum_{a \in A} \prod_{m=1}^{M} \left(\frac{d_m}{c_m}\right)^{\delta_m(i,a)} + \sum_{b \in B} \prod_{m=1}^{M} \left(\frac{d_m}{c_m}\right)^{\delta_m(i,b)}}. \qquad \text{(C3)}$$

Equations C3 and C1 are formally identical with $s_m = (d_m/c_m)$. Thus, in the classification paradigm under current consideration, extending the standard context model with common-feature match parameters does not change the formal predictions from the model. The formal predictions would be changed, however, if individual stimuli were composed of differing numbers of binary-valued dimensions.

Received February 6, 2003
Revision received April 21, 2003
Accepted April 28, 2003 ■