Journal of Experimental Psychology:
Learning, Memory, and Cognition
2001, Vol. 27, No. 4, 1022–1041


0278-7393/01/$5.00 DOI: 10.1037//0278-7393.27.4.1022

# Exemplar Accounts of Blending and Distinctiveness Effects in Perceptual Old–New Recognition

Safa R. Zaki and Robert M. Nosofsky
Indiana University, Bloomington

Recent findings from the perceptual old–new recognition literature indicate that observers have extremely high false-alarm rates to new items that are "blends" of old ones. In addition, evidence suggests that "distinctive" old items—that is, those located in isolated regions of the similarity space—are recognized with higher probability than are typical old items. Both types of phenomena challenge the predictions of global-familiarity exemplar models of perceptual old–new recognition, which posit that the probability that an observer judges an item as old is based on its summed similarity to previously presented exemplars. In the present research the authors pursued these blending and distinctiveness effects by testing paradigms in which similarity relations among objects are highly controlled and in which the variables of blending and distinctiveness are not confounded with other properties associated with the individual objects themselves. In contrast to previous results, the authors found effects of blending and distinctiveness that are compatible with the predictions of a pure summed-similarity exemplar model.

An interesting effect in recognition memory research is sometimes observed when a prototypical item, not seen during training, is falsely recognized during test (e.g., Busey & Tunnicliff, 1999; Franks & Bransford, 1971; Metcalfe & Fisher, 1986; Neumann, 1977; Solso & McCarthy, 1981). In certain cases, the false recognition of the prototype is high and sometimes even exceeds the hit rates of the old items. This effect is robust in the sense that it occurs across a wide variety of stimulus domains such as words (Roediger & McDermott, 1995), faces (Solso & McCarthy, 1981), and colors of vehicles in eyewitness testimony (Loftus, 1977). For example, in a well-known study by Roediger and McDermott (1995; see also Shiffrin, Huber, & Marinelli, 1995), participants heard a list of words that were all associated to a particular theme word. In a subsequent test, participants judged with high probability that they recognized the theme word, even though the word was new.

To account for these effects, researchers have considered whether novel traces that correspond to the prototypes are created in memory. Specific theories of recognition attribute this false alarming at the central tendency to a creation of a "blend" or prototype in memory (e.g., Busey & Tunnicliff, 1999; Homa, Goldhardt, Burruel-Homa, & Smith, 1993; Loftus, 1977; Metcalfe, 1990). For example, in Metcalfe's composite holographic associative recall model (CHARM), the traces of events that are experienced are explicitly blended or superimposed in memory. This blend in memory is accessed when the participant is probed with the prototype of the old items, and hence the high false-alarm rates.

However, another description is compatible with these results. Exemplar-based models account for recognition data without recourse to an explicit blending mechanism (e.g., Estes & Maddox, 1995; Gillund & Shiffrin, 1984; Heit, 1993; Hintzman, 1988; Nosofsky, 1988). These models assume that when a new item is presented, it is compared with the old items in memory. The familiarity of an item is a function of the similarity of that item to all of the old items. The greater the summed similarity to the old items, the greater the feeling of familiarity and therefore the greater likelihood the items are called "old." One can intuit that this sort of model would predict high false recognition of the prototypes because the prototypes are extremely similar to numerous items in memory. In fact, formal versions of exemplar models, such as Nosofsky's generalized context model (GCM), have been shown to yield impressive quantitative fits to perceptual recognition data (e.g., Nosofsky, 1988, 1991; Shin & Nosofsky, 1992) and have been shown to predict classic effects of false recognition of the prototype. More recently, Busey and Tunnicliff (1999) conducted simulations of various data sets (e.g., Kroll et al., 1996) from experiments that used parts of studied faces to form conjunction stimuli. Data previously taken as evidence of a blending or binding mechanism in the memory-conjunction literature could qualitatively be accounted for by the GCM.

Another important factor that may influence recognition memory is the "distinctiveness" of an item. The term *distinctiveness* carries varied connotations in alternative research settings. In the present research we used the term in a highly specific sense to refer to the degree to which an object lies in an isolated region of the similarity space among previously studied items. Thus, distinctive items lie in isolated regions of the similarity space, whereas

Safa R. Zaki and Robert M. Nosofsky, Department of Psychology, Indiana University, Bloomington.

This research was supported by Grant PHS RO1 MH48494-09 from the National Institute of Mental Health. We thank Thomas Busey, Henry L. Roediger III, and two anonymous reviewers for their helpful criticisms of an earlier version of this article. We also thank Thomas Busey for providing the stimulus sets from the Busey and Tunnicliff (1999) experiments. In addition, we are grateful to Elizabeth Arnott, Tony Mascari, Brent Meyers, and Ryan Williams for their assistance in testing the participants.

Correspondence concerning this article should be addressed to Safa R. Zaki, Department of Psychology, Indiana University, Bloomington, Indiana 47405-1301. Electronic mail may be sent to szaki@indiana.edu.

"typical" items lie in dense regions. A robust finding from the perceptual recognition literature is that false-alarm rates associated with new typical items are greater than those associated with new distinctive items (e.g., Metcalfe & Fisher, 1986; Nosofsky, 1991; Nosofsky, Clark, & Shin, 1989; Omohundro, 1981; Shin & Nosofsky, 1992). This result follows readily from global-familiarity models such as the GCM. According to the GCM, because the summed similarity of new items to old objects is higher for typical items than for distinctive items, the false-alarm rates to typical items should exceed false-alarm rates to distinctive ones. Indeed, the summed-similarity mechanism in the GCM has been shown in past work to accurately predict fine grained differences in recognition probabilities as a function of fine grained differences in similarities among items (Nosofsky, 1988, 1991; Nosofsky & Zaki, 1998; Shin & Nosofsky, 1992).

In summary, the GCM has fared well in most past research in accounting for prototype effects and distinctiveness effects in perceptual old–new recognition memory. However, in a recent study, Busey and Tunnicliff (1999) reported a rich set of old–new recognition data that severely challenged the predictions of the model. In the Busey and Tunnicliff experiments, participants trained with a large set of pictures of faces. The stimuli in the training phase included similar and dissimilar pairs of faces, called "parents," as well as other targets. In the test phase, participants were shown the parents, the target items, and various new foils. Included among the foils were morphs, which were faces created by averaging each pair of parent faces. There were two aspects of the results that were potentially challenging to the exemplar model accounts of recognition. The first was that the false-alarm rates to the morphs of similar parents were significantly higher than the hit rates of the parents. As discussed previously, the GCM can predict high false-alarm rates to these types of stimulus blends; however, as we discuss later, the model predicts certain constraints on the magnitude of the false alarms relative to the hit rates of the parents, and these constraints were violated in Busey and Tunnicliff's data. A second challenge from their data set was a recognition advantage displayed by old distinctive items. That is, old distinctive items had high hit rates relative to old typical items. By contrast, the summed similarity mechanism of the GCM predicts the opposite pattern of responding. Indeed, the GCM could not fit these two sets of results simultaneously. Although the model could generally predict that the false recognition of the prototypical new objects would be high, it was unable to predict the magnitude of the false recognition effect. In addition, the GCM systematically underestimated the hit rates to the old distinctive items.

Nevertheless, there are some important limitations of Busey and Tunnicliff's (1999) experiments that need to be considered. The first is that the face space is likely highly complex, and, as acknowledged by Busey and Tunnicliff, the derived six-dimensional multidimensional scaling (MDS) solution that they used for representing similarity relations among the faces was likely inadequate. It is plausible that observers use more than six dimensions when recognizing and classifying faces. Obviously, if the faces are not adequately represented in the MDS solution, then the exemplar model may not provide the correct predictions. A second potential limitation is that the morphing process itself may have caused artifacts in the stimuli. Previous research indicates that morphing may indeed produce artifacts of the sort that may influence recognition judgments. For example, morphed faces tend to appear smoother, younger, and even more attractive than the faces from which they were created (e.g., Busey, 1998; Galton, 1878). These factors may be at least partly responsible for the high false-alarm rate to the prototypical items.

Third, it is important to point out that in Busey and Tunnicliff's (1999) study, as well as in other studies of perceptual old–new recognition, although the compared items varied in their typicality and distinctiveness, this variation was confounded with properties of the individual objects themselves. For example, as acknowledged by Busey and Tunnicliff, in their design the faces that occupied distinctive regions of the similarity space all tended to have beards, whereas the faces that occupied typical regions did not. Furthermore, a face that was distinctive with respect to the other faces studied may also have been distinctive with respect to all other faces that an observer had experienced during his or her lifetime. It is important to carefully separate the contributions of typicality and distinctiveness in a particular experimental setting from these other properties associated with the individual objects themselves.

The general goal of the present research was to follow up on Busey and Tunnicliff's (1999) study of blending and distinctiveness effects in old–new perceptual recognition. First, we tested for the possibility that the morphing process used by Busey and Tunnicliff to produce blends might indeed have produced artifacts that contributed in a substantial way to the extremely high false-alarm rates observed for the prototypes. Second, rather than using complex faces as a stimulus set, we decided to use colors varying in hue, brightness, and saturation. Numerous similarity-scaling and classification studies indicate that the color stimuli are well represented in this multidimensional space. Thus, use of these simpler stimuli might allow for fairer and more rigorous tests of the exemplar model's predictions of old–new recognition judgments. Furthermore, it is straightforward to test blends in the color space that are unlikely to have the same artifactual properties associated with the morphed faces. Finally, to separate effects of typicality and distinctiveness from other factors associated with the individual objects, we tested designs in which such individual object properties were held constant while rotating which objects resided in typical or distinctive regions of the similarity space across conditions.

In this article, we have grouped the experiments into three sections. In Experiment 1, we investigated the possibility that high false-alarm rates to the morphs in Busey and Tunnicliff's (1999) experiments may be attributed in part to artifacts of the morphing process. In Experiments 2–4, we tried to reproduce several of these blending effects with color stimuli and tested the extent to which these effects can be accounted for by a pure summed-similarity-to-exemplars model and without recourse to an explicit blending mechanism. In Experiments 5–6, we investigated the roles of typicality and distinctiveness on old–new recognition and tested the extent to which the summed-similarity mechanism in the GCM can account for the data.

## False Recognition of Morphed Faces

### *Experiment 1*

As a preliminary study, Experiment 1 was geared at investigating whether stimulus artifacts could have produced the high levels of false alarms to the morphed faces in Busey and Tunnicliff's (1999) design. It has long been known that composite or morphed stimuli are perceived as being younger and more attractive than the parent faces from which they are generated (e.g., Busey, 1998; Galton, 1878). Conceivably, such factors could strongly influence people's old–new recognition judgments.[1]

Experiment 1 was designed to investigate the idea that, even without previous exposure to the parents, people would be more likely to judge the morphed faces as old. To test this hypothesis, we measured the rate of false alarms to all the stimuli used in Busey and Tunnicliff's (1999) experiments in a paradigm in which participants were told that they had seen these stimuli subliminally but in fact had seen no stimuli at all.

### Method

*Participants.* Fifty Indiana University undergraduate students participated in the research to fulfill a requirement of their introductory psychology class.

*Stimuli.* The stimuli were digitized pictures of bald men (Kayser, 1985) used by Busey and Tunnicliff (1999). There were 104 pictures that Busey and Tunnicliff had designated as four different stimulus types. More specifically, the stimuli consisted of 32 parents, 36 targets, 20 foils, and 16 morphs. The parents were organized into 16 pairs of stimuli, half of them similar pairs, and half of them dissimilar. The morphs were the averages of each parent pair. Busey and Tunnicliff created the morphed stimuli by first manually placing at least 150 control points on the images of the two parents. Using these control points as landmarks on the two faces, they then digitally warped the pictures so that the various features were maximally aligned. Finally, corresponding pixels in the two parents were averaged. The faces were shown on Sony Trinitron 15-in. monitors.

*Procedure.* There were 68 trials in the mock subliminal training phase, the same number of trials as in the training phase of the Busey and Tunnicliff (1999) experiment. On each trial, participants saw a fixation point followed by a black rectangle. The black rectangle remained on the screen for approximately 50 ms. Therefore, what the participants essentially saw was a quick flash on the screen. Participants were asked to attend to the stimuli as best they could. They were told that we were testing their ability to subliminally perceive faces. The instructions read in part as follows:

> In the first phase, you will be shown each face very briefly. The display of the face will be so quick that you will think that you have not seen anything. You will simply see something flash on the screen. In the phase that follows, we will test your memory of the faces in the first phase.

In the test phase that followed, participants received the following instructions:

> In the next phase you will be shown a series of 88 faces. Half of these are faces that were subliminally flashed during the first phase. The remaining faces are new. We are interested if your memory system was able to record any information whatsoever regarding the nature of the subliminally flashed faces. For each face, we will ask how likely it is that it was one of the faces in the first phase. The ratings range from 1 (*extremely unlikely*) to 9 (*extremely likely*). Obviously, you will have no conscious recollection of the faces, so you will need to rely on various forms of unconscious information to perform the task—gut instinct, a sense of familiarity, and so forth.

There were two test sets, each of which included one of the parents and all of the morphs, targets, and foils. The two sets were constructed so that only one parent from each pair would be shown during testing to ensure that the level of false-alarm rates to the morphs could not be attributed to blending of the parents at the time of test. Half of the participants judged each test set.

### Results

The mean recognition ratings for the different types of faces are shown in Table 1. The morphs (5.53) were rated as more likely to be one of the faces shown in the first phase than were the parents (5.07), $t(49) = 3.54$, $p < .001$. Busey (1998) has shown that morphs tend to lie toward the center of the space, and therefore their high false-alarm rate could conceivably be due to the blending of items at time of test. To address this concern, we calculated the false-alarm rates to the morphs during the first and last half of the test. There was no difference between the false-alarm rate to the morphs in the first half (5.65) and the second half (5.40). In addition, there was no increase in the recognition rate of the morphs relative to that of the parents from the first half (difference of .60) to the second half (difference of .31).

### Discussion

The results of this experiment support the idea that the high false-alarm rates of the morphs in Busey and Tunnicliff's (1999) study can be attributed, at least in part, to factors other than blending. That is, the high false-alarm rates to the morphs that were attributed to a blending mechanism by Busey and Tunnicliff were evident even without exposure to the two parents. Therefore, in the following series of experiments, we tested for blending effects in an alternative stimulus domain in which the problems associated with the morphed faces would be unlikely to arise; namely, in the domain of colors varying in hue, brightness, and saturation. Colors have been used extensively as stimuli in many published experiments and have the advantage that the relevant dimensions are known and perceptual distances among the objects have previously been measured. In addition, Metcalfe (1990) has suggested that blending is more likely to occur in domains in which the intermediate points between two stimuli exist or are capable of existing in the real world, as is the case with colors.

## Tests for a Blending Mechanism in the Color Domain

In the experiments presented in this section, we investigated the evidence for a blending mechanism in the domain of colors and tested the extent to which hit rates for parents and false-alarm rates for novel prototypes were compatible with the predictions from the GCM. The GCM is a member of a more general class of global-familiarity models of recognition, including Gillund and Shiffrin's (1984) search of associative memory model of recall (SAM), Hintzman's (1988) MINERVA 2, Metcalfe-Eich's (Eich, 1982) CHARM, Murdock's (1982) theory of distributed associative memory (TODAM), and Shiffrin and Steyvers's (1997) retrieving effectively from memory (REM). These alternative models include more intricate processing machinery than does the GCM and thereby account for a broader range of old–new recognition phe-

[1] We consider in the General Discussion section how such factors might be incorporated within an elaborated exemplar-based model of recognition. The key question here, however, is whether a blending process is involved.

Table 1
*Familiarity Scores for the Various Stimulus Types in Experiment 1*

| Item type | Familiarity |
|---|---|
| Prototype | 5.53 |
| Parent | 5.07 |
| Target | 4.86 |
| Foil | 5.07 |

nomena, including effects of context, interitem associative strength, spacing effects, the slopes of receiver operating characteristic (ROC) curves, and the mirror effect. An advantage of the GCM, however, is that it incorporates a rigorous theory of similarity and thereby makes precise quantitative predictions of old–new recognition performance for individual items based on their locations in multidimensional similarity space. Thus, effects of blending and distinctiveness can be examined in rigorous quantitative fashion from the perspective of the model. To explain the rationale for the core manipulations in the following experiments, in this section we review the specific assumptions of the model and then outline some key predictions.

According to the GCM (Nosofsky, 1988, 1991), recognition judgments are based on the summed similarity of an object to all items in memory. The objects are represented as points in a multidimensional space, and similarity between objects is a decreasing function of their distance in the space. The psychological distance between two items is calculated from their location in the MDS space as follows:

$$d_{i,j} = \sqrt{\sum w_m(x_{i,m} - x_{j,m})^2}, \tag{1}$$

where $x_{i,m}$ is the MDS coordinate of item $i$ on dimension $m$, and $w_m$ ($0 \leq w \leq 1$, $\sum w_m = 1$) is the attentional weight given to dimension $m$. The similarity between items $i$ and $j$ is then given by an exponential decay function of their psychological distance,

$$s_{i,j} = e^{-cd_{i,j}}, \tag{2}$$

where $c$ is a sensitivity parameter (Nosofsky, 1984; Shepard, 1987).

Familiarity of test item $i$ is found by summing the similarity between item $i$ and all other items in memory,

$$F_i = \sum M_j s_{i,j}, \tag{3}$$

where $M_j$ gives the memory strength of item $j$. The memory strengths may vary due to factors such as frequency or recency of presentation, individual item salience, and so forth. However, for simplicity, in the initial section of our article, we assume that all memory strengths are equal to one. Note that in this article, objects we refer to as *typical* are those that give rise to high values of summed similarity ($F_i$ in Eq. 3), whereas objects we refer to as *distinctive* give rise to low values of summed similarity.

Finally, the probability that item $i$ is judged as old is given by

$$P(\text{old}|i) = \frac{F_i}{(F_i + k)}, \tag{4}$$

where $k$ is a response-criterion parameter (Clark, 1988; Estes, 1994; Shin & Nosofsky, 1992). For simplicity, throughout this article, we assume that within a given condition, the criterion-parameter $k$ is invariant across different items and regions of the similarity space. In our General Discussion section, however, we suggest reasons why this simplifying assumption may need to be abandoned in some situations.

Although the GCM can predict high levels of false recognition of the prototype, the level of this false-alarm rate relative to the hit rates of the old items is constrained by the structure of the model. To illustrate this point, consider two general scenarios that capture the essence of the manipulations in the following experiments. In the first case, as illustrated in Figure 1A, there are two old items in a particular region (the open squares) as well as a centrally located new item (the solid circle). This case is analogous to the manipulation in Busey and Tunnicliff's (1999) experiments in which the two items were the studied parents and the central item was the morph presented at test. In a second scenario, as depicted in Figure 1B, four old training items surround a new prototype.

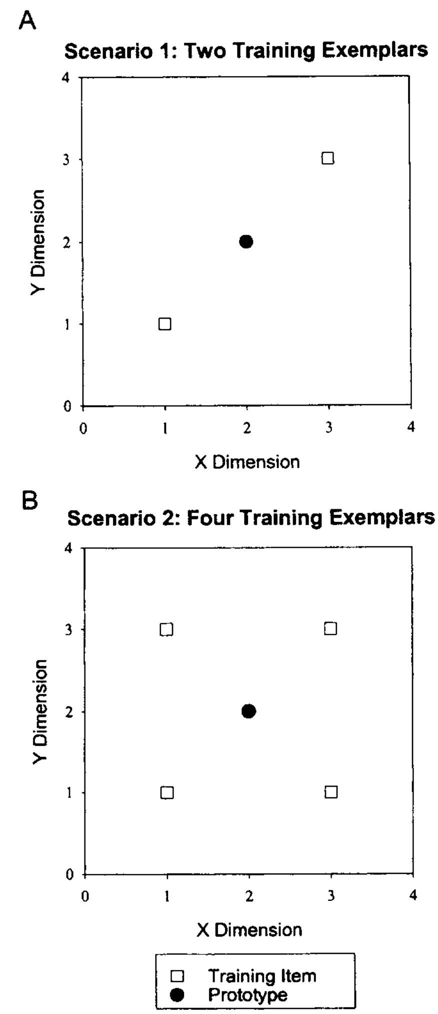


*Figure 1.* A schematic representation of two potential training scenarios: Two training items surround a prototype (A) and four training items surround a prototype (B).

For both of these scenarios, we investigated predictions from the GCM of the prototype false-alarm rates versus the hit rates of the old items by calculating the summed similarity values for the prototypes and old items (Equations 1–3). In performing these computations, we varied the value of the sensitivity parameter, $c$, from 0 to 2 in increments of .1. The coordinates of the objects in MDS space are shown in Figure 1. The results of the computations are plotted in Figure 2. Figure 2A corresponds to the situation illustrated in the first scenario, and Figure 2B corresponds to the second scenario. Critically, for the first scenario, there are no parameter values in the model that allow it to predict that the false-alarm rates to the prototypes would exceed the hit rates of the old items. Furthermore, it is only for values of the sensitivity parameter that are virtually equal to zero that the model predicts that false alarms to the prototypes will be nearly equal to hit rates for the old items. Thus, the design illustrated in Figure 1A is potentially highly diagnostic. If, in fact, one observed false-alarm rates of the prototypes that exceeded the hit rates of the old items,

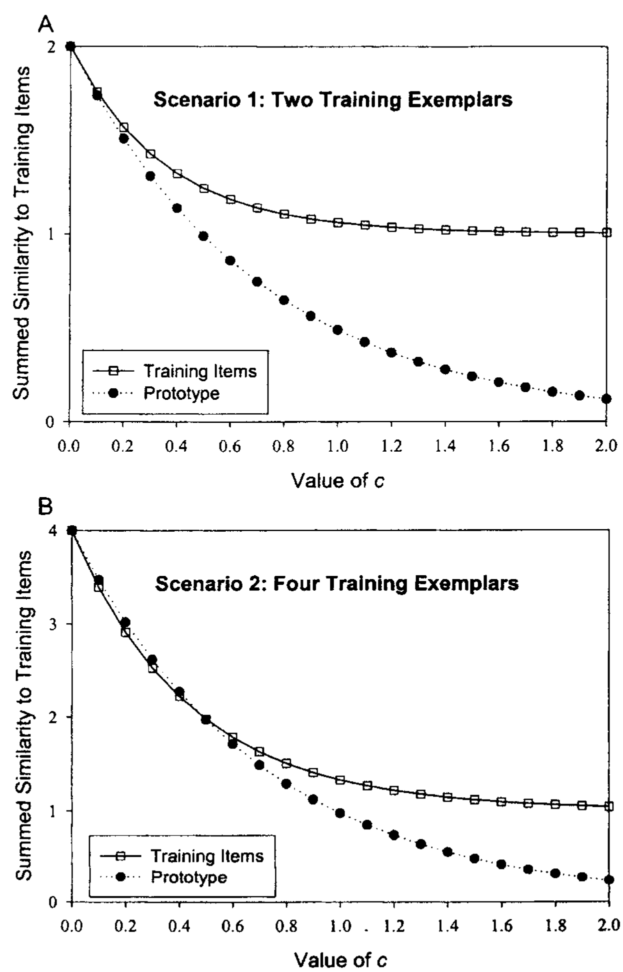


*Figure 2.* Summed similarity of the prototypes and old items to the training items when two items in a particular region are studied (A) and when four items in a particular region are studied (B). The maximum value on the ordinate is the maximum possible summed similarity to the training items. The values on the ordinate are scaled relative to this maximum similarity. $c$ = overall sensitivity.

the exemplar model could simply not account for those results. In the second scenario, on the other hand, the exemplar model predicts that, over a reasonable range of sensitivities, the false recognition rates to the prototypes will at least equal the hit rates of the old items. In other words, as the number of similar items that surround the prototype increases, the rate of false recognition of the prototype relative to the correct recognition of the old items increases. The purpose of the following experiments was to investigate these factors that are predicted by the exemplar model to influence the rate of false recognition of the prototype.

## *Experiment 2*

In Experiment 2, we used a structure similar to that displayed in Figure 1A; the stimuli were colors displayed on a computer screen. In a training phase, participants viewed a list of colors organized into pairs, with each pair being of a constant hue but varying in brightness and saturation. The geometric configuration of each color pair is illustrated in Figure 1A. In a subsequent recognition test phase, participants were tested with the old colors, the prototypes of the training pairs, and the new foils. The prototypes were centrally located between the members of each training pair on both the brightness and saturation dimensions. Previous research has shown that prototype effects may be influenced by category size (Homa, Sterling, & Trepel, 1981; Reed, 1978; Robinson & Roediger, 1997). For this reason, we decided to include two list-length conditions, one in which there were 8 colors on the study list and a second in which there were 16 colors.

This study provides a controlled test of the evidence for blending in the case in which each prototype is closely surrounded by only two items. Although we do not systematically manipulate which item is the prototype and which items are the training items, we believe it is unlikely that item specific effects will occur because the design involves only local variations in the different regions of the color space. (Nevertheless, we checked for the possibility of item-specific effects in Experiment 4.) As discussed previously, computations from the GCM suggest that this design provides a critical test of the exemplar model because under no parameter settings does the exemplar model predict that the false-alarm rates for the new colors would exceed the hit rates of the old colors. The model predicts that, even in the long-list condition, the false recognition of the prototypes should be lower than the hit rates of the old items.

### *Method*

*Participants.* One hundred Indiana University undergraduates participated to complete an introductory psychology course requirement. To motivate the participants, a bonus of $15 was offered to the top two performers in the experiment.

*Stimuli.* Sixteen stimuli were chosen for use in the training phase. These stimuli consisted of pairs of colors from eight distinct hues adapted from the Munsell color system. The eight hues were chosen so that they were maximally separated in the Munsell color space. Within each hue region, a pair of colors varying in both brightness and saturation was selected. In the test phase, the color at the brightness–saturation midpoint of each of these pairs was used as the prototype of the region. In addition, the test phase included eight foils chosen from completely novel hues. Although the distance in terms of hue between each adjacent region was constant, we tried to maximize perceptual distance between the regions by choosing certain values of brightness and saturation. Specifically, adjacent hue regions alternated between ones in which the colors were very bright but not very saturated and ones in which the colors were not bright but were highly saturated. We followed this procedure to minimize any cross-region effects on blending. A complete listing of the hue, brightness, and saturation coordinates of the colors is provided in Appendix A.

The computer-displayed stimuli were created by scanning the Munsell color chips into the computer. In Appendix B, we report the results of a scaling study that documents that the geometric relations holding among the scanned colors closely resemble those of the original Munsell color chips.[2] The red, green, blue (RGB) values of the scanned colors are given in Appendix A.

*Procedure.* In the training phase, participants were asked to simply pay attention to the colors as they appeared one at a time on the screen and to try to remember them. They were told that their memory for the colors would be tested in the next phase. Participants were trained on five blocks of the stimuli. Within each block, the order of the colors was randomized for each participant. Each color was displayed on a computer screen as a 2-in. square on a white background and remained on the screen for 5 s.

Participants were randomly assigned to one of two conditions. The two conditions differed only in the number of old items. In the long-list condition, participants trained with 16 items (two colors from each of eight hue regions). In the short-list condition, participants trained with 8 items that were from four contiguous hue regions. In the short-list condition, participants always trained with one of the two halves of the long-list condition so as to not confound interregion similarity with list length.

In the testing phase, participants saw the old stimuli, the prototypes of the training regions, and the foils. In the short-list condition there were four foils, and in the long-list condition there were eight foils. Therefore, half of the test set was novel in both conditions. On a given trial in this phase, the stimulus was presented on the computer screen and the participants judged whether the item was new or old. Participants were told that half of the items would be old and half of the items would be new. Feedback was not provided.

### Results

Table 2 reports the probabilities of judging the different item types as old. There is clearly no evidence for blending in the experiment. The false-alarm rate to the prototypes (.69) was significantly lower than the hit rates to the parents (.83) in both the short-, $t(49) = 5.83$, $p < .001$, and long-list conditions, $t(49) = 2.72$, $p < .05$.

### Discussion

Experiment 2 was an attempt to conceptually replicate the structure of the Busey and Tunnicliff (1999) blending manipulation with stimuli that were more controlled in terms of their similarity relations and in which artifactual properties associated with the blended objects were unlikely to arise. This design is a particularly diagnostic one because only when sensitivity is zero does the exemplar model predict that the false-alarm rates for the prototype colors would equal the hit rates for the old colors. Additionally, in this design, the model never predicts that the false-alarm rates of the prototypes would exceed the hit rates of the parents. However, unlike the results in Busey and Tunnicliff's study, we found under these highly controlled conditions that the blending effect was not observed. The next question we address is whether a blending effect is observed under conditions in which the exemplar model does predict that a blending effect is likely to occur.

Table 2
*Probability of "Old" Responses for Each Type of Stimulus in Experiment 2*

| | Item type | | |
|---|---|---|---|
| List length | Old | Prototype | Foil |
| 8 | .85 | .66 | .12 |
| 16 | .82 | .73 | .14 |

## *Experiment 3*

Although the false-alarm rates of the prototypes did not meet or exceed the hit rates of the parents in Experiment 2, the GCM does make predictions about the conditions under which higher false-alarm rates to the prototype are likely to occur. As discussed previously, for example, the GCM predicts that under certain stimulus conditions in which four items are presented in a given region, as in Figure 1B, the false-alarm rates to the prototypes can meet or slightly exceed the hit rates to old items. Experiment 3 was designed to test this prediction made by the model.

As in the previous experiment, participants viewed the items that surrounded a prototype in the training phase. However, in this experiment, in each color region, there were four studied items, as illustrated in Figure 1B. During the test phase, participants were shown the old items and the prototypes, along with various foils. The prediction that the GCM makes is straightforward. Given the same level of sensitivity across experiments, this design should yield higher false recognition of the prototype items relative to the hit rates of the old items, and the prototype false-alarm rate could conceivably equal or slightly exceed the old-item hit rate.[3]

### Method

*Participants.* Fifty-five Indiana University undergraduates participated to complete an introductory psychology course requirement. Once again, to ensure high motivation, we offered a $15 bonus to the top two participants.

*Stimuli.* The stimuli for this experiment were drawn from four Munsell hues that were as separated as the hues in the previous experiment. For each of these hue regions, five stimuli varying in brightness and saturation were chosen such that four of them formed the vertices of a square with the fifth stimulus in the center. For each of the four regions, the four nonprototypical colors were used in the training phase and the middle stimulus (the prototype) was shown only during test. Figure 1B provides a schematic display of one of these regions. Finally, 12 foils were included in the test set. Four of these items were from the same hues as the four critical regions, but had brightness–saturation values that were considerably different than those of the old items studied. The other 8 foils had distinct hues

[2] In fact, it is not critical that the prototype be perfectly centrally located. The general pattern of results obtained in the simulations holds as long as the prototype is some intermediate point between the training items.

[3] More specifically, fits of the exemplar model to the old–new recognition data for Experiment 2 yielded a sensitivity parameter estimate of $c = .78$ in the long-list condition. (Experiment 3 also includes a long list of 16 study items.) Inspection of Figure 3B reveals that with $c = .78$, the predicted summed similarity for the old items and the prototype are nearly identical.

and were distant from all of the training items. The Munsell chips were scanned using the procedure described in Appendix B. The Munsell color values and RGB values for the 32 items are given in Appendix A.

*Procedure.* In the training phase, participants were asked to simply pay attention to the colors as they appeared, one at a time, on the screen. Participants were instructed that their memory for the colors would be tested in the next phase of the experiment. Sixteen colors were shown in this phase. These colors were drawn from the four hue regions described previously. As in the previous experiment, participants were trained on five blocks of the stimuli. Each color was displayed in a random order within each block and remained on the screen for 5 s.

In a test phase, participants were shown the 16 old items, the four new prototypes, and the 12 foils. On each trial the participants indicated whether the item was old or new. They were told that half of the items were new and half were old. The test items appeared in a different random order for each participant.

### Results

Table 3 shows the levels of responding "old" to the various stimulus types in Experiment 3. In general, participants appeared to be performing similarly on the old colors as they did in Experiment 2. However, in this experiment the levels of false recognition of the prototype (.85) and the hit rates to old colors (.83) were statistically equivalent, $t(54) = 0.56, p > .10$.

### Discussion

Experiment 3 was designed to test a prediction of the GCM that, for a particular geometric configuration of objects, the false-alarm rate of the (new) prototype will approach the hit rate of the old objects. This predicted blending effect was indeed observed. Thus, combining the results from Experiments 2 and 3, the GCM can predict fairly well the conditions in which blending will and will not be observed.

The separation of hues was constant across Experiment 2 and Experiment 3. However, because of constraints of picking four items in a region, the different hue regions were closer together in saturation and brightness in this experiment than in the previous one. Therefore, although the most plausible explanation for the relatively higher false recognition of the prototype rates in Experiment 3 is the high density of the surrounding old colors within each hue region, the greater interregion similarity could also have had some effect.

## Experiment 4

There were two main goals for Experiment 4. The first goal was to provide experimental support for the assumption that the local variations within color regions that we have used in these studies would not give rise to stimulus specific effects. That is, we wanted to make sure that the prototypes and parent items in these experiments would not give rise to different baseline levels of false alarms, as did the stimuli used by Busey and Tunnicliff (1999). Thus, in a mock subliminal condition similar to Experiment 1, participants were told that they had seen a number of colors but in fact had been shown no stimuli at all. The participants then rated the familiarity of the colors. We used this procedure to obtain baseline false-alarm rates to the prototypes and parent items.

The second and more important goal of Experiment 4 was to control interregion hue similarity while manipulating only the configuration of training items within regions in a single experiment. Participants were trained on the same stimulus set as in Experiment 3, except the participants were trained on only half of the items. Participants were randomly assigned to one of two conditions in which they trained on either four items from each of two of the hue regions (the dense condition, illustrated in Figure 1B) or two items from each of the four regions (the sparse condition, illustrated in Figure 1A). In a subsequent recognition phase, the participants were tested on these old colors, the prototypes, and a number of foils. If region density plays an important role in influencing false recognition levels of the prototype, as the GCM predicts, a relatively higher false-alarm rate to the prototype should be observed in the dense condition than in the sparse condition. In the sparse condition, false alarms to the prototype should be clearly lower than hits to the old training items, whereas in the dense condition the false alarms to the prototype should approach the old-item hit rates.

### Method

*Participants.* One hundred Indiana University introductory psychology students participated in the recognition conditions of the experiment to fulfill a course requirement. As an incentive for good performance, we offered $15 to the top two participants. In addition, 56 Indiana University introductory psychology students participated in the mock subliminal condition.

*Stimuli.* The 26 stimuli used in this experiment consisted of all 16 training items from the four regions used in Experiment 3, the four prototypes, and 6 of the 12 foils of Experiment 3. Appendix A indicates which of the foils in Experiment 3 were included in Experiment 4. All foils were from hue regions that were distinct from those of the training items.

*Procedure.* The procedure in the subliminal condition was similar to Experiment 1. On each of eight trials in the subliminal phase, participants saw a brief flash of a black rectangle. The instructions were nearly identical to those used in Experiment 1. Participants were told that various colors were being flashed with short durations on the screen, followed by a black rectangle, and that their task was to pay attention as best they could. On each trial of the familiarity rating phase that followed, participants judged how likely it was that the color was shown in the previous phase on a scale with responses ranging from 1 (*extremely unlikely*) to 9 (*extremely likely*). The test set included the prototype and one of the surrounding items for each region. The particular surrounding item was chosen randomly for each participant.

The remaining participants were randomly assigned to either a dense recognition condition or a sparse recognition condition. Participants in both conditions were trained on a total of eight items. In the dense conditions, participants trained on four items from each of two hue regions. In the sparse condition, participants trained on two items from each of the four hue regions. The hue regions in the dense condition were always contiguous regions so that interregion similarity was constant across the sparse and dense conditions. In the sparse condition, the (diagonal) pair of items that was shown within each region was randomly selected for each partic-

Table 3
*Probability of "Old" Responses for Each Type of Stimulus in Experiment 3*

| Item type | Old |
|---|---|
| Old | .83 |
| Prototype | .85 |
| Foil | .13 |

ipant, as was the subset of foils that was shown. As before, training consisted of the presentation of the stimuli on the computer screen, one at a time for 5 s each. Participants were asked to simply attend to the stimuli and were told that their memory for the colors would be tested in a subsequent phase. Because there were fewer training items in Experiment 4 than there were in the long-list condition of Experiments 2 and 3, we decided to use fewer training blocks so that memory sensitivity would not be too high. Therefore, participants trained on the stimuli for two blocks with the order of items in each block randomly determined.

At test, participants made recognition judgments to the old colors, the prototypes of the training regions, and the foils. In the dense condition there were eight old items, two prototypes, and six foils. In the sparse condition, there were eight old items, four prototypes, and four foils. Therefore, in both conditions there were eight new and eight old items in the test set. Participants were told that half of the items in the test phase were old and that half were new.

## Results

In the mock subliminal condition, the ratings of how likely a color was to have been shown in the first phase did not differ for the prototype (4.85) and the surrounding colors (4.80), $t(55) = 0.30$, $p > .1$, supporting our assumption that any difference in recognition probabilities in the standard recognition experiment is not due to stimulus specific effects.

Table 4 shows the observed probabilities of responding "old" to the various stimulus types in the standard recognition conditions of Experiment 4. The false-alarm levels of the prototype were significantly lower than the hit rates of the old items, $F(1, 98) = 15.79$, $p < .001$. In addition, participants in the dense condition were more likely to call the items in the region old, $F(1, 98) = 8.49$, $p < .01$. Most important, as predicted by the model, a significant interaction between region density condition and item type was observed, $F(1, 118) = 4.36$, $p < .05$, indicating a smaller difference between the hit rates to the old items and the false-alarm rates to the prototype in the dense condition. Although this interaction is in accord with the qualitative predictions from the model, we should acknowledge that it could be reflecting, at least in part, a scaling effect in which old probabilities are approaching a ceiling in the dense condition. To gain more incisive evidence bearing on the pattern of results, therefore, we turn to formal quantitative modeling of the data.

## Quantitative Modeling of the Data

It is straightforward to apply the GCM to predict the recognition data from Experiment 4. We assumed that the configuration within each hue region is as depicted in Figure 1. (In the sparse condition, the configuration is as depicted in Figure 1A, whereas in the dense condition the configuration is as depicted in Figure 1B.) We assume that the summed similarities for each type of item are determined primarily by the distance relations among the objects within the same hue region. For simplicity in the modeling, we assume that colors from differing hue regions have some small residual similarity between one another, which we represent with a free parameter $r$. Thus, consider for example the predictions for the old items from the dense condition, which occupy the four vertices of the square configuration in Figure 1B. A given old item has similarity 1 to itself, similarity $e^{-2c}$ to each of its adjacent neighbors on the square, and similarity $e^{-\sqrt{8}c}$ to the old item on the opposite diagonal of the square. Finally, this old item has residual similarity $r$ to each of the four old items from the other hue region presented on the study list. Thus, the summed similarity (familiarity value) for any given old item in the dense condition is given by

$$F(\text{old, dense}) = 1 + 2e^{-2c} + e^{-\sqrt{8}c} + 4r,$$

and the probability that an old item from the dense region is therefore judged as old is given by

$$P(\text{old, dense}) = \frac{F(\text{old, dense})}{F(\text{old, dense}) + k}.$$

Analogous prediction equations are easily derived for each of the other item types in each condition. Application of the model requires estimation of three free parameters: the overall sensitivity parameter $c$, the residual-similarity parameter $r$, and the response-criterion parameter $k$. We searched for the values of these free parameters that minimized the sum of squared deviations (*SSD*) between predicted and observed recognition probabilities. The best fitting parameters were $c = 1.00$, $r = .002$, and $k = .214$, with a resulting $SSD = .0003$. The predicted recognition probabilities are shown along with the observed probabilities in Table 4. The table shows that the GCM provides an excellent quantitative fit to the data. Thus, not only are the data consistent with the general qualitative predictions made by the model, but the GCM is capable of providing an excellent quantitative account of the blending effects observed in this study.

Table 4
*Observed and Predicted Recognition Probabilities in Experiment 4*

| | Dense condition | | Sparse condition | |
|---|---|---|---|---|
| Item type | Observed | Predicted | Observed | Predicted |
| Old | .87 | .86 | .83 | .83 |
| Prototype | .83 | .82 | .69 | .70 |
| Foil | .06 | .06 | .05 | .06 |

## Discussion

In summary, the goal of Experiments 2–4 was to test for blending effects in highly controlled situations in which similarity relations among items were precisely represented and in which artifactual stimulus-specific properties associated with the prototypes did not arise. Under these highly controlled conditions, the overall pattern of blending effects was in accord with the predictions of the summed-similarity exemplar model. In situations in which a new prototype item was located centrally between two old parent objects, the false alarms to the prototype were lower than were the hit rates to the old objects. However, in situations in which the new prototype item was centrally located in a configuration of four old parent items, the false-alarm rates to the prototype were virtually identical to the hit rates of the old items. This qualitative pattern of results is as predicted by the GCM under a wide range of settings of its memory sensitivity parameter. Furthermore, beyond effectively predicting the overall qualitative pattern of results, the model was capable of achieving precise quantitative fits to the old–new recognition data.

## Distinctiveness Effects

We now turn to the second main issue to be addressed in this research, namely the effects of typicality and distinctiveness of objects on old–new recognition judgments. We reemphasize that the terms *typicality* and *distinctiveness* are used in different ways by different researchers. As was the case in Busey and Tunnicliff's (1999) study, in the present research, typicality and distinctiveness are defined as the extent to which objects in a given experimental context are similar to other objects experienced in that context. A typical object is one that is located near other presented objects in the multidimensional similarity space and therefore has high summed similarity (value of $F_i$ in Equation 3); whereas a distinctive object is located in an isolated region of the similarity space and therefore has low summed similarity.

To review, with respect to new items, Busey and Tunnicliff (1999) observed that false alarms were higher for typical objects than for distinctive ones. This result is a robust one that is often observed in tasks of perceptual old–new recognition (e.g., Franks & Bransford, 1971; Metcalfe & Fisher, 1986; Nosofsky, 1991; Nosofsky et al., 1989; Omohundro, 1981; Shin & Nosofsky, 1992). This typicality effect on false alarms is also a central prediction made by global-familiarity models such as the GCM. The more challenging result for the GCM was Busey and Tunnicliff's observation that hit rates for distinctive old items exceeded hit rates for typical old items. Indeed, the GCM makes exactly the opposite prediction: Because the summed similarity for typical old items exceeds the summed similarity for distinctive old items, typical old–item hit rates should exceed distinctive old–item hit rates.

As noted in our introduction, however, a limitation associated with Busey and Tunnicliff's (1999) experimental design (as well as with numerous other studies involving perceptual recognition) was that the variables of typicality and distinctiveness were confounded with other properties associated with the individual objects. Conceivably, these individual-object properties could have exerted a powerful effect on old–new recognition judgments, above and beyond the roles of typicality and distinctiveness per se.

The central goal of the present studies, therefore, was to begin an investigation into the roles of typicality and distinctiveness on perceptual old–new recognition by explicitly manipulating these variables across conditions while holding fixed other properties of the individual objects.

## *Experiment 5*

Experiment 5 is illustrated schematically in Figure 3. There were three critical color regions (A, B, and C) in which the typicality and distinctiveness of objects were manipulated. Each region was defined by three potential old items that might be presented during study, as well as a new prototype item that was never presented during study. Across conditions, zero, one, or all three of the potential old items were presented from each critical region during study. For example, in Condition 1, zero old items were presented from Region A, one old item was presented from Region B, and all three old items were presented from Region C. Thus, in this condition, the old item from Region B was highly

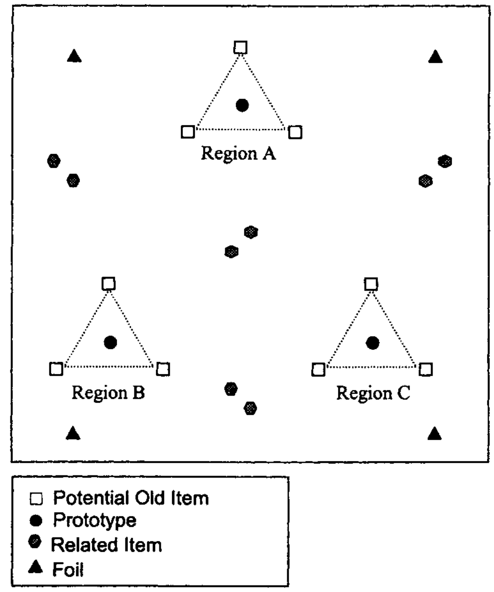


*Figure 3.* A schematic representation of the stimuli used in Experiment 5. Each triangular array of items represents one of the three critical color regions. The regions were arbitrarily labeled A, B, and C in the figure.

distinctive, whereas the old items from Region C were typical. (Subsequent modeling of our experimental results in terms of the GCM will verify that these manipulations successfully led to higher summed similarity values for the intended typical items and lower summed similarity values for the intended distinctive ones.) Following study, a test phase was conducted in which all items from all of the regions were presented; participants judged whether each item was old or new.

We expected, of course, that false alarms to the prototypes and other new items would increase with their typicality. Thus, in the condition described previously, false alarms to the prototype from Region C should exceed false alarms to the prototype from Region B; false alarms to the prototype from Region A should be lowest of all. The critical question concerned the hit rates for the old items: Would the hit rates for the distinctive old items (e.g., those from Region B in the condition described earlier) exceed the hit rates for the typical old items (i.e., those from Region C), as was observed in Busey and Tunnicliff's (1999) study? Or might the earlier results observed by Busey and Tunnicliff have been reflecting factors other than those of typicality and distinctiveness per se?

### *Method*

*Participants.* One hundred and twenty Indiana University undergraduates participated to complete an introductory psychology course requirement. A bonus of $15 was offered to the top two participants to ensure their motivation.

*Stimuli.* The stimuli were 24 colors that differed in hue, saturation, and brightness. A schematic diagram of the stimuli is shown in Figure 3. There were three critical hue regions. In each of these regions, one of these items (the prototype) was constructed to be approximately central to the other three. In addition to the items from these three regions, we also included four pairs of "related" colors. Each of these pairs consisted of two colors that were relatively close to each other but close to no other items in the space. These items were intended to serve as filler items in study and test lists. Finally, the stimulus set included four foils that were intended to be located relatively far from each other and other items. The actual stimuli were chosen by pilot testing.[4]

*Procedure.* In the training phase, participants saw eight different colors. Three of these colors were the colors that surrounded the prototype from one region. Participants also saw one nonprototypical item from another critical region. In addition, participants saw one item from each of the four related pairs. The participants were shown four blocks of these eight colors during the training phase. The colors appeared in a random order in each block. As in the previous experiments, participants were simply asked to carefully observe as the colors appeared on the screen and were told that their memory for the colors would be tested in a subsequent phase. Each of the colors remained on the screen for 5 s.

During training, we manipulated the density of the three regions. Participants were randomly assigned to one of six conditions. For each of these conditions, one of the three regions was designated the *zero-item region,* one of the regions was designated the *one-item region,* and the last region was designated as the *three-item region.* The item shown from the one-item region and the item shown from each of the related pairs were randomly determined for each participant.

In a test phase, participants were shown all 24 colors and were asked to indicate whether each item was new or old. The 24 colors included the old items, the new items from the critical regions (including the prototypes), the related pairs, and the foils. Note that in this experiment, participants saw all colors in all regions during test. This procedure was followed to maintain an identical testing situation in all conditions.

In the similarity-scaling part of the experiments, participants gave similarity ratings to displays of two colors. The ratings were given on a 9-point scale, with responses ranging from 1 (*least similar*) to 9 (*most similar*). The participants were instructed to use the full range of the scale. Participants made ratings to all unique pairs of 20 colors (all but the isolated foils) with one exception. To minimize the number of ratings that each participant made, especially ratings of items that were very dissimilar and therefore potentially more frustrating, we collected similarity judgments between only one of two items in a related pair and other critical region items. The participants made 142 similarity judgments. The similarity judgments were collected to derive an MDS solution for the colors. This MDS solution was used in conjunction with the GCM for purposes of formal modeling.

### Results

The probability with which each type of item was judged as old is reported in Table 5. The old target items were judged as old with higher probability than were the prototypes, and the prototypes were judged as old with higher probability than were the other new items. As expected, false-alarm rates to the new items that were similar to a single old item (New 1) exceeded false-alarm rates to new items that were similar to zero old items (New 0), $t(119) = 3.29, p < .01$. Likewise, as expected, false-alarm rates to the prototypes increased systematically with the number of old items to which they were similar: False-alarm rates to Prototype 3 were higher than to Prototype 1, $t(119) = 2.21$, $p < .05$, and false-alarm rates to Prototype 1 were higher than to Prototype 0, $t(119) = 6.08, p < .001$. Thus, the old–new recognition data show the usual pattern of results with regard to the effects of typicality on false-alarm rates to new items.

The critical data of interest are the hit rates to the old items. As can be seen in Table 5, in contrast to Busey and Tunnicliff's (1999) results, the distinctive old items (Old 1) were not recognized with higher probability than were the typical old items (Old 3). If anything, the results go slightly in the opposite direction, although the differences do not approach statistical significance.

Table 5
*Observed and Predicted Recognition Probabilities for Each Type of Item in Experiment 5*

| | Recognition probability | |
|---|---|---|
| Item type | Observed | GCM |
| Old 1 | .83 | .81 |
| Old 3 | .85 | .87 |
| Proto 0 | .27 | .28 |
| Proto 1 | .63 | .56 |
| Proto 3 | .77 | .77 |
| New 0 | .27 | .22 |
| New 1 | .42 | .50 |
| Related-old | .91 | .81 |
| Related-new | .56 | .67 |
| Isolated-new | .06 | — |

*Note.* Values (0, 1, or 3) following items denote the number of old training items to which a given test item is similar. The dash indicates predicted probabilities for isolated-new items are not available because these stimuli were not included as part of the MDS solution for the colors. Proto = prototype; GCM = generalized context model.

Finally, before turning to the formal modeling analyses, we note that the hit rates to the *related-old* items were higher than the hit rates to those old items whose distinctiveness was manipulated across conditions. Because the distinctiveness of the related-old items was not experimentally manipulated across conditions, this result cannot easily be interpreted. Because the related-old items tended to be in even more distinctive regions of the similarity space than the manipulated old items (see ensuing MDS analyses), the result could indeed reflect an effect of distinctiveness on hit rates of old items. Unfortunately, the result could also reflect properties associated with these individual objects themselves. We attempted to disentangle these possibilities in a subsequent experiment.

### Theoretical Analyses

The matrix of averaged similarity ratings[5] between all pairs of colors was submitted to the alternating least squares scaling (ALSCAL) multidimensional scaling program using the standard Euclidean model. A three-dimensional solution gave a good fit to the data, yielding stress = .039 and accounting for 99.2% of the variance in the averaged similarity ratings. This three-dimensional solution for the colors is displayed in Figure 4. As can be seen, the colors from the three main manipulated groups each fall in relatively distinct regions of the similarity space, and the prototype of each group tends to lie more toward the center of each region than do the old target items. The members within each related pair of items lie close to one another, but each separate related pair is located in a distinct region of the space. This similarity structure satisfies the goals of the intended experimental design.

[4] The RGB values and the obtained MDS solutions for the stimuli in these experiments are available on request from Safa R. Zaki.

[5] We conducted an additional experiment that differed only slightly from Experiment 5, which we do not report in this article. Similarity ratings were combined from the 240 participants in that experiment and Experiment 5 to produce the averaged similarity matrix.

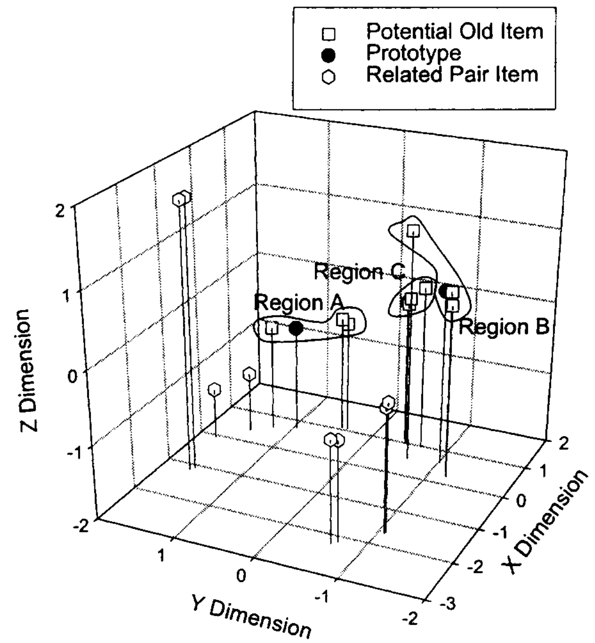


*Figure 4.* The obtained three-dimensional scaling solution for the stimuli used in Experiment 5.

We now use the GCM in combination with the derived three-dimensional MDS solution for the colors to quantitatively fit the old–new recognition data. In the baseline version of the model, we allowed only two free parameters to vary: the overall sensitivity parameter $c$ in Equation 2 and the response-criterion parameter $k$ in Equation 4. We assumed that participants weighted equally each of the three psychological dimensions (Equation 1), so the $w_m$ attention weight parameters did not enter into the model fits. All of the memory-strength parameters $M_j$ (Equation 3) were held fixed at 1.0. We conducted a computer search for the values of the free parameters ($c$ and $k$) that minimized the *SSD* between the predicted and observed recognition probabilities. The best fitting parameters were $c = 6.038$ and $k = .247$, yielding $SSD = .037$.[6]

The predicted recognition probabilities are reported along with the observed probabilities in Table 5. The two-parameter model accounts reasonably well for the pattern of data, particularly for the recognition probabilities associated with those objects in the manipulated regions of the similarity space. First, the model correctly predicts that the old items are recognized with higher probability than the prototypes and that the prototypes are recognized with higher probability than the other new items. The reason the old items are predicted to have the highest recognition probabilities is that they perfectly match their traces in memory, which provides a large boost to the summed similarity computation performed by the model. The prototypes are predicted to have slightly higher recognition probabilities than the other new items because they tend to be more centrally located in each subregion of the similarity space, so they have higher summed similarities.

Second, the model predicts correctly the steadily increasing false-alarm rates for both the new items and the prototypes as a function of their typicality. The model predicts this result because the greater the number of old items to which the new items and prototypes are similar, the greater is their overall summed similarity.

Finally, the baseline model predicts higher hit rates for the typical old items (Old 3) than for the distinctive old items (Old 1), as was observed in the present data set. The model appears to slightly overestimate this increase in hit rates, but the departures between the predicted and observed data values are not dramatic. Note that the model does predict correctly that increases in typicality lead to much smaller increases in old-item hit rates than in new-item false-alarm rates.

Perhaps the major shortcoming of the model is that it does not fit well the hit and false-alarm rates associated with the related-old and related-new items. These items were intended to serve as filler items on the study and test lists. Because the typicality and distinctiveness of the objects in these regions were not experimentally manipulated across conditions, we did not consider the results for these items to be of much theoretical importance. Indeed, the similarity-scaling results for these items are almost certainly not as precise as those for the remaining items. The reason is that each related pair lies in an extremely distinctive region of the similarity space, so the similarity ratings between these pairs and the other objects in the space were likely approaching some floor. We suspect that part of the problem in modeling the data for the related items is simply that the members of each pair were more discriminable from one another than was revealed by the derived MDS solution.

Nevertheless, the lack of fit for the related items does raise an important concern. In particular, the hit rate associated with the related-old items was greater than the hit rates associated with the old objects in the manipulated regions of the similarity space (Old 1 and Old 3). Even if the MDS results for the related items are not sufficiently precise, there is still no mechanism in the baseline version of the GCM that can predict this qualitative pattern of results: The related-old items lie in the most distinctive regions of the similarity space, so their summed similarity must be lower than that of the Old 1 and Old 3 items.

We have verified that extended versions of the GCM can fit these data by assuming that the memory strengths ($M_j$ in Equation 3) for the related-old items exceed the memory strengths for the old items in the manipulated regions. Such model-fitting exercises leave unanswered, however, the question of why the related-old items may have had higher memory strengths. Does it reflect a distinctiveness effect or an effect of these individual objects themselves? The purpose of Experiment 6 was to disentangle these possibilities and to gain further evidence bearing on the roles of typicality and distinctiveness on perceptual old–new recognition.

## *Experiment 6*

Detailed analysis of the Experiment 5 data revealed that related-old pair 3, 4 (the left-most and most elevated pair in Figure 4) had the highest recognition probabilities among all of the objects. Therefore, in Experiment 6, we tested these items again, except that now we manipulated the typicality and distinctiveness of items

[6] The magnitude of the sensitivity parameter estimates in this section cannot meaningfully be compared with those in the previous section because the scaling units of the MDS solution are arbitrary.

from this region across conditions. Furthermore, we were concerned that the colors from Regions B and C in Experiment 5 may not have been sufficiently distant from one another, which could dilute any distinctiveness effects. Therefore, in Experiment 6, we removed from the study list the colors from the previous Region C.

Experiment 6 is illustrated schematically in Figure 5. Again, there were three regions of the color space in which typicality and distinctiveness were manipulated. Regions A and B were the same ones used in Experiment 5. The new Region C consisted of colors from the region containing related-pair 3, 4 in Experiment 5. In each condition of Experiment 6, one, two, or three of the potential old items from Regions A, B, and C were presented during study. The prototypes from each region were never presented during study. At time of test, all items from all regions were presented, and participants judged whether each one was old or new. In addition, the test set included three foils from distant hue regions.

## Method

*Participants.* Ninety Indiana University introductory psychology students participated in the experiment to complete a course requirement. We offered a bonus of $15 to the top two participants.

*Stimuli.* The stimuli were 15 computer-displayed colors. As in the previous experiment, there were three critical regions of different hues. Each of these regions contained four colors of similar hues but with different levels of saturation and brightness. In each of these regions, one of these items was constructed to be approximately central to the other three. Regions A and B were the same regions as in Experiment 5. Region C was constructed around the related pair yielding the highest hit rates in the previous experiment. In addition to the items from these three regions, there were also three foils located relatively far from each other and from the other items in color space.

*Procedure.* In the training phase, participants saw six different colors. These colors were from three hue regions. In one of these regions, participants saw one color, in another the participants saw two, and in the third, the participants saw three. The assignment of number of items to regions was counterbalanced across all subjects. In each of four blocks of training, participants viewed these six colors in random order. Each color remained on the screen for 5 s. As in the previous experiments, the participants were simply instructed to carefully observe the colors on the screen and were told that their memory for the colors would be tested in the phase to follow.

In the test phase, participants were shown all 15 colors and were asked to indicate whether each item was new or old. The 15 colors included the old items, the new items from the critical regions, including the prototype, and three foils. Participants saw all colors in all regions during testing to maintain an identical testing situation across conditions.

The procedure in the similarity rating part of the experiment was almost identical to the procedure in the previous experiment. Participants were asked to give similarity ratings on a 9-point scale to pairwise combinations of the colors from the critical regions. Responses on the scale ranged from 1 (*least similar*) to 9 (*most similar*). Participants made ratings to all unique pairs of the 12 colors from the critical regions (the foils were not included in this set). To decrease the relative proportion of highly dissimilar pairs, and increase the precision of the scaling within each color region, participants saw combinations of colors from the same region twice, whereas combinations of colors from different regions were only seen once. All pairs of colors appeared in a random order. In total, each participant made 83 similarity judgments.

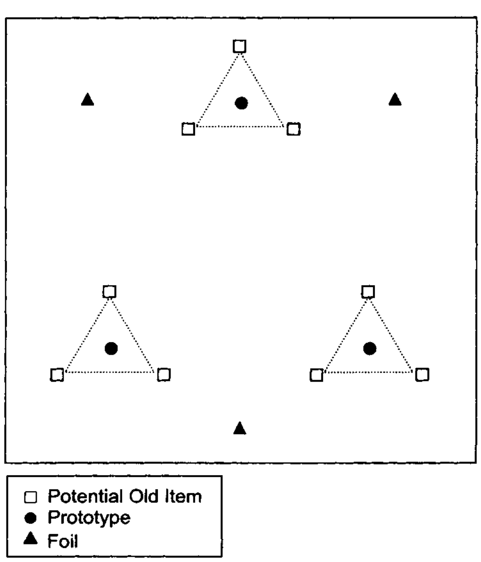


*Figure 5.* A schematic representation of the stimuli in Experiment 6.

## Results

The probability with which each type of item was judged as old is reported in Table 6. The pattern of results is similar to the one we observed previously in Experiment 5. Again, the old target items were judged as old with higher probability than were the prototypes, and the prototypes were judged as old with higher probability than were the other new items. In addition, false-alarm rates to the prototypes increased systematically with their typicality, $F(2, 178) = 3.41, p < .05$, although for some reason the other new items showed only a small nonsignificant increase in false-alarm rates as a function of typicality, $F(1, 89) = 0.21, p > .05$. Finally, we once again did not observe that distinctive old items had higher hit rates than typical old items, $F(2, 178) = 1.14, p > .05$. If anything, the results went in the opposite direction, in accord with the predictions from the GCM.

A detailed breakdown of the old–new recognition data from all three regions is provided in Table 7. The table reveals that the hit rates and false-alarm rates for old and new items tended to increase with the objects' typicality in all three manipulated regions of the similarity space, as predicted by the GCM. Interestingly, the mean hit rates associated with objects from Region C (i.e., the region containing related-pair 3, 4 from Experiment 5) were higher than the mean hit rates associated with objects from the other regions, $t(89) = 2.91, p < .01$. Thus, these colors do indeed appear to possess certain item-specific properties that give rise to their higher memory strengths. The higher hit rates for the related-old items in Experiment 5 were apparently not an effect of distinctiveness per se.

## Theoretical Analysis

A two-dimensional scaling solution gave a good fit to the matrix of averaged similarity ratings among the colors, yielding stress = .025 and accounting for 99.7% of the variance in the data. This two-dimensional scaling solution is illustrated graphically in Figure 6. Each color region gives rise to a highly compact and distinctive region in the similarity space. Unfortunately, because the regions are so compact and distinctive, it is unlikely that the between-regions similarity-ratings data provide sufficient informa-

Table 6
*Observed and Predicted Recognition Probabilities for Each Type of Item in Experiment 6*

| | Recognition probability | |
|---|---|---|
| Item type | Observed | GCM |
| Old 1 | .84 | .84 |
| Old 2 | .89 | .86 |
| Old 3 | .90 | .87 |
| Proto 1 | .58 | .42 |
| Proto 2 | .69 | .61 |
| Proto 3 | .77 | .71 |
| New 1 | .21 | .31 |
| New 2 | .23 | .53 |

*Note.* Values (0, 1, or 3) following items denote the number of old training items to which a given test item is similar. Proto = prototype; GCM = generalized context model.

tion to tightly constrain the within-region structures, so we should acknowledge that much of the precision of the MDS solution may be lost (see, e.g., Kruskal & Wish, 1978, pp. 29–30).

We used the GCM in combination with the derived two-dimensional scaling solution for the colors to fit the old–new recognition data. As was the case in modeling the results from Experiment 5, only two free parameters were estimated: the overall sensitivity parameter $c$ in Equation 2 and the response-criterion parameter $k$ in Equation 4. We conducted a computer search for the values of these free parameters that minimized the *SSD*s between the predicted and observed recognition probabilities. The best fitting parameter values were $c = 23.913$ and $k = 0.191$, yielding $SSD = .140$.

The predicted recognition probabilities are shown along with the observed probabilities in Table 6. The model predicts correctly that the old items were recognized with the highest probability, followed in turn by the prototypes and other new items. In addition, the model predicts correctly the increase in false-alarm rates for the prototypes as a function of their typicality, and the smaller increase in hit rates for the old items. The main shortcoming of the model is that it overpredicts the recognition probabilities for the other new items (New 1 and New 2), perhaps because the MDS representation for the colors within each region is too compact.

### *Discussion*

The results of this experiment are highly instructive. In the absence of an actual manipulation of typicality and distinctiveness across conditions, an investigator might have concluded that it was the distinctiveness of the old-related items from Experiment 5 that gave rise to their higher hit rates. The present experimental results suggest, however, that it was not a distinctiveness effect per se, but rather some other effect associated with these individual objects themselves. This result is analogous to the initial one from Experiment 1, in which we demonstrated that the extremely high false-alarm rates to morphed stimuli might not reflect a blending effect, but rather properties associated with the individual morphed stimuli themselves. The effects of typicality and distinctiveness on object hit and false-alarm rates were as predicted by the GCM, and the model gave a reasonably good quantitative account of the old–new recognition data.

## General Discussion

### *Summary*

In summary, the purpose of this research was to investigate effects of blending and distinctiveness on old–new perceptual recognition memory and to test a summed-similarity exemplar model (Nosofsky's, 1988, 1991, GCM) on its ability to account for these effects. The immediate motivation for the work involved some recently reported data sets from Busey and Tunnicliff (1999). These researchers used a large set of pictures of faces as stimuli and examined the probability with which old faces, new faces, and morphed faces were recognized. The morphed faces were blends of pairs of the old studied faces. The key results from that study were that observers false alarmed to the new blended faces with higher probability than they correctly recognized the old studied faces. In addition, observers displayed higher hit rates for distinctive old faces (ones located in isolated regions of the similarity space) than they did for typical old faces (ones located in dense regions of the similarity space). Both results pose challenges to the summed-similarity exemplar model of old–new recognition. Given the structure of Busey and Tunnicliff's face-space design, under no parameter settings can the GCM predict higher false recognition of the prototype faces (the morphs) than correct recognition of the old parent faces from which they were generated. In addition, because typical faces have greater summed similarity to old exemplars than do distinctive faces, the general prediction from the model is that hit rates for typical faces should exceed hit rates for distinctive ones.

Nevertheless, we pointed to some limitations of Busey and Tunnicliff's (1999) design that could have important bearing on the interpretation of the results. First, we noted that the similarity space for faces is likely to be highly complex, and, as acknowledged by Busey and Tunnicliff, the six-dimensional scaling solution they used for representing these similarities was likely to have

Table 7
*Recognition Probabilities for Old and New Items in the Three Regions of Experiment 6*

| Region | Old 1 | Old 2 | Old 3 | Proto/New 1 | Proto/New 2 | Proto/New 3 |
|---|---|---|---|---|---|---|
| Region A | .83 | .92 | .90 | .30 | .45 | .50 |
| Region B | .77 | .85 | .83 | .29 | .42 | .87 |
| Region C | .93 | .90 | .97 | .41 | .52 | .93 |

*Note.* Proto/New refers to the combined recognition probabilities for the prototypes and other new items in the three critical regions. Proto = prototype.

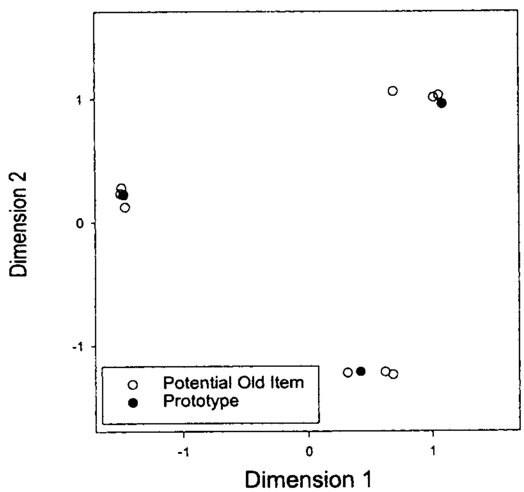


*Figure 6.* The obtained two-dimensional scaling solution for the stimuli in Experiment 5.

been inadequate. Second, we noted that in Busey and Tunnicliff's design, the fundamental variables of blending and distinctiveness were confounded with other properties of the individual objects themselves.

We addressed these concerns in the present research by using several approaches. First, we documented in a mock subliminal recognition study that the morphed faces did indeed have certain artifactual properties that apparently influence old–new recognition judgments above and beyond the role of blending per se. Second, we tested designs in which observers made old–new recognition judgments for color stimuli rather than faces. The colors have the advantage that the underlying properties of the similarity space are well documented and that blends created from local variations in color are unlikely to possess the same types of artifactual properties as the morphed faces. Finally, we tested designs in which distinctiveness was systematically manipulated across conditions while holding fixed properties of the individual objects themselves. Thus, in one condition, a given object might be a distinctive one in the similarity space of studied items, whereas in a comparison condition it might be a typical object.

Under these highly controlled conditions, we found effects of blending and distinctiveness that were compatible with the predictions from the summed-similarity exemplar model. With regard to blending, in all conditions in which two parent colors were used to form a blended prototype, the hit rates to the parents exceeded the false recognition of the new color blend. However, in conditions in which a configuration of four parent colors was used to form the blended prototype, the false recognition of the prototype was equal in magnitude to the hit rates of the old items. This overall qualitative pattern of results is consistent with the predictions from the GCM over a wide range of settings of its memory sensitivity parameter. Furthermore, we demonstrated that the model was capable of yielding precise quantitative fits to the complete sets of old–new recognition data obtained in these blending studies.

With regard to the effects of distinctiveness and typicality, we observed the following pattern of results. First, as the typicality of new objects increased, the probability with which observers false alarmed to the new objects increased. This finding is often observed in old–new perceptual recognition tasks and is a fundamental prediction made by the summed-similarity exemplar model. Second, in contrast to Busey and Tunnicliff's (1999) findings, we found that hit rates to old distinctive items were not greater than hit rates to old typical items. If anything, the results tended to go in the opposite direction, as predicted by the GCM. Furthermore, the magnitudes of the increasing false-alarm and hit rates (as a function of typicality) were in close quantitative accord with the predictions from the summed-similarity exemplar model.

### *Individual Object Properties*

Our interpretation is that at least some of the effects that Busey and Tunnicliff (1999) attributed to the variables of blending and distinctiveness may be attributable to other properties of the individual objects themselves. What are these properties and how can they be incorporated in a more fully specified exemplar model of recognition? We have already noted that the morphed faces differed from the standard parent faces in that they appeared younger, smoother, and more attractive. Conceivably, although summed similarity to stored exemplars may be the major force that drives old–new recognition judgments, other miscellaneous variables such as the ones listed earlier may contribute to such judgments as well. Thus, all other things being equal, an observer may have a bias to judge a more attractive face as more likely to have been a previously studied one.

Another important possibility to consider, however, is that summed similarity to stored exemplars is indeed the primary force that drives old–new recognition judgments. However, the exemplars that participate in the summed-similarity computations are not simply the ones that appeared in the experimental study lists. Rather, past exemplars that an observer has experienced in his or her lifetime, prior to entering the experimental setting, may enter into the summed-similarity computations as well. For example, Heit (1994) has demonstrated that effects of prior knowledge in categorization can often be well modeled by assuming that previously stored exemplars enter into the computations of the summed-similarity exemplar model. In a similar vein, Maddox and Estes (1997) studied the effects of prior familiarity of objects on old–new recognition performance. These authors tested designs in which prior to presenting participants with actual study lists, they were presented with familiarization lists in which the frequency of some of the to-be-studied items was manipulated. Maddox and Estes found that a summed-similarity exemplar model provided a good quantitative account of the old–new recognition judgments, as long as provision was made for the idea that similarities to exemplars presented on the familiarization lists also entered into the exemplar model's computations.

### *Distinctiveness Effects From the Face-Recognition Literature*

Besides Busey and Tunnicliff's (1999) study, a number of other studies from the face-recognition literature have also reported that hit rates for old distinctive faces are greater than hit rates for old typical faces (e.g., Bartlett, Hurry, & Thorley, 1984; Light et al., 1979; Vokey & Read, 1992). Indeed, by making reference to this literature, Valentine and Ferrara (1991) argued that the summed-similarity rule of exemplar models failed to account for such old–new recognition data: "By definition, there are likely to be more faces stored in memory which have a greater similarity to a typical face than to a distinctive face. Therefore the familiarity signaled by the summed-similarity rule will be greater for a typical face than a distinctive face" (p. 88).

We believe, however, that it is difficult to fully ascertain the implications of such results for the summed-similarity model of old–new recognition. First, the construct of distinctiveness is operationalized in a wide variety of ways in different research situations. For example, in the studies of Bartlett et al. (1984) and Vokey and Read (1992), the distinctiveness of objects was not experimentally manipulated across conditions. Rather, observers provided direct ratings of distinctiveness and typicality for individual faces, and the mean hit and false-alarm rates for the rated distinctive and typical faces were then compared. (In the Vokey and Read study, distinctiveness was operationalized in terms of participants' ratings of the memorability of individual faces.) Such studies are inherently correlational in nature and numerous factors besides "degree of isolation in a similarity space" may enter into participants' ratings of distinctiveness and memorability.

Furthermore, suppose that the faces that observers rate as distinctive do indeed tend to be ones that lie in isolated regions of the similarity space. It is still possible that such faces possess certain individual-object properties that play a crucial role in old–new recognition (above and beyond the role of distinctiveness per se). For example, in Busey and Tunnicliff's (1999) study, the faces located in the isolated regions of the similarity space also tended to have beards, whereas the typical faces did not. Conceivably, this individual-object property may result in the object having a stronger and more salient representation in memory than do other more typical objects without such a property. To fully account for such effects, a theory is needed for how the memory-strength parameters in the exemplar model (Equation 3) may vary across the individual objects.

### *Other Issues for Future Research*

The central theme we emphasized in this research is that it is important to separate effects of blending and distinctiveness from effects of other properties of the individual objects themselves. These factors have tended to be confounded in numerous studies of face recognition. However, there are other important differences between the nature of our studies and those from the face-recognition literature.

First, there may well be stimulus domain differences in the effects that blending and distinctiveness have on old–new recognition. Although we chose to study color recognition because the similarity space of colors is well understood, it may be that distinctive colors are not as easily recognized as are distinctive faces or that our manipulations of color distinctiveness were not sufficiently extreme. However, it is important to point out that in stimulus domains involving verbal materials (Criss & Shiffrin, 2000; Shiffrin et al., 1995) and faces (Criss & Shiffrin, 2000), researchers have observed effects that were analogous to the ones reported here.

A second possibility concerns the manner in which we operationalized distinctiveness in the current work, namely, in terms of local neighborhood density. Perhaps alternative operationalizations may come closer to capturing the construct of distinctiveness. For example, distinctive items may simply be those furthest from the average item in a given domain, regardless of the number of neighboring items. Alternative operational definitions of distinctiveness may lead to different results, which may or may not be well accommodated by an exemplar model.

One final idea we are currently pursuing is that the contrasting results from the face-recognition literature and those from the present study and that of Shiffrin et al. (1995) may reflect differences in the nature of the testing situation. In the present research, we defined distinctive old objects as those lying in isolated regions of the similarity space of studied items. In the face-recognition literature, however, the nature of the testing situation tends to be confounded with the nature of the study situation. For example, in Busey and Tunnicliff's (1999) experiments, a sample of faces with differing properties was generated and some faces were randomly selected as old and other faces as new. Furthermore, some of the faces happened to lie in dense regions of the similarity space (i.e., typical faces), whereas others happened to lie in isolated regions (i.e., distinctive faces). At time of test, all of the faces were 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. Furthermore, another consequence of this procedure is that participants will be tested on a much larger number of faces from typical regions than from distinctive ones.

By contrast, in the present studies and in the experiments reported by Shiffrin et al. (1995), the nature of the testing situation was held fixed across conditions and was not confounded with the properties of the study phase. In the present studies, for example, participants were tested in a uniform manner with all objects from all regions, regardless of which particular objects were presented at time of study. Shiffrin et al. used a similar testing procedure.

Conceivably, in the face-recognition studies, participants may gain some meta-knowledge of the structure of the testing situation. This meta-knowledge could influence their recognition judgments. For example, a participant may come to realize that he or she is never being tested on new objects that are similar to old objects that lie in distinctive regions of the similarity space. If so, the participant could optimize performance by using a strategy of responding "old" to any tested object that is at all similar to distinctive ones studied. To account for such effects, current global-familiarity models of old–new recognition would need to be augmented by allowing for item-specific or region-specific adjustments in criterion settings. It is important for researchers to pursue these possibilities in the future.

### *False Recognition in More Complex Domains*

The applications of the GCM in the present study were limited to the domain of colors. Can the model be applied in richer stimulus domains and can it handle still more challenging phenomena involving old–new recognition memory? For example, in the present experiments, the GCM successfully modeled cases in which the false recognition rates of the prototypes were at the same level as the correct hit rates of the old items. These results were obtained in a situation in which the prototype was surrounded by four old items configured in a two-dimensional spatial region. In

Table 8
*Summed Similarities of Old Exemplars and New Prototype to Studied Items in 16-Dimensional Space as a Function of Overall Sensitivity (c) and Number of Exemplars on the Study List*

| | No. of items in study list | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 2 | | 4 | | 8 | | 16 | |
| $c$ | O | P | O | P | O | P | O | P |
| 1.0 | 1.50 | 1.21 | 2.49 | 2.43 | 4.49 | 4.85 | 8.47 | 9.70 |
| 2.0 | 1.25 | 0.74 | 1.75 | 1.47 | 2.75 | 2.94 | 4.76 | 5.89 |
| 3.0 | 1.13 | 0.45 | 1.38 | 0.89 | 1.88 | 1.79 | 2.90 | 3.57 |
| 4.0 | 1.06 | 0.27 | 1.19 | 0.54 | 1.46 | 1.08 | 1.97 | 2.17 |

*Note.* The summed similarity computations assumed equal attentional weighting of the 16 dimensions. O = the summed similarity of an old item to all of the studied items; P = the summed similarity of the prototype to the old studied items.

semantic memory experiments, however, there are reported cases in which false recognition of the prototype actually exceeds correct hit rates of the old items, especially as category size increases (e.g., Robinson & Roediger, 1997). Adequate modeling of similarity relations among the concepts from semantic domains undoubtedly would require MDS solutions of much higher dimensionality than the ones used here. Although any attempt to derive such representations goes well beyond the scope of the present research, we can nevertheless explore the general behavior of the model in this type of situation.

In the following simulations, we consider predictions from the GCM in a case in which the stimuli are assumed to reside in a 16-dimensional similarity space and the category size is 2, 4, 8, or 16 items. In each of 1,000 simulations, a random set of old items was generated. Each item was defined by a 16-dimensional vector consisting of a random sequence of zeros and ones. Thus, each old item occupied a random vertex of a 16-dimensional unit hypercube. (Analogously, in the blending studies reported in our experiments, the four old items occupied the vertices of a two-dimensional square.) A prototype was also defined, located at the centroid of the 16-dimensional hypercube. This item was a 16-element vector with each element set at value .5.

In Table 8, we report mean summed similarity values of the old items and the new prototype for each of the four category sizes as a function of values of the sensitivity parameter ($c$) ranging from 1.0 to 4.0. For the smaller category sizes (2 and 4 items), the summed similarity of the old items always exceeds the summed similarity of the prototype. Interestingly, however, at category size 8, the summed similarity of the prototype exceeds the summed similarity of the old items when memory sensitivity is low. For the larger category size (16 items), the summed similarity of the prototype exceeds the summed similarity of the old items, even at higher sensitivity levels. Thus, cases in which false recognition of the prototype exceeds correct recognition of the old items are compatible with the predictions from the GCM. Furthermore, the simulation results indicate that such cases are most likely to occur when the psychological structure of the stimuli is highly multidimensional and when category size is large.

To apply the GCM to quantitatively predict recognition probabilities for individual items in complex domains, however, techniques are needed to derive the underlying similarity representations. Note that although in past work, the GCM has been combined only with low-dimensional spatial solutions for modeling similarity, there is no reason in principle why the model cannot be combined with alternative methods of similarity representation. With regard to face recognition, for example, one possibility might involve combining the GCM with the principal-component analyses of face representation used successfully by Abdi, Valentin, O'Toole, and their colleagues (Abdi, Valentin, Edelman, & O'Toole, 1995; Abdi, Valentin, & O'Toole, 1997). Likewise, by making use of word association norms as a database, Steyvers (2000) has applied the singular value decomposition technique from Landauer and Dumais (1997) to derive high-dimensional vector representations for the similarity relations among words. Future research is needed to explore the extent to which the use of such similarity-scaling techniques may allow for successful extensions of global-memory models such as the GCM to domains such as face and semantic concept recognition.

## References

Abdi, H., Valentin, D., Edelman, B., & O'Toole, A. J. (1995). More about the difference between men and women: Evidence from linear neural networks and the principal-component approach. *Perception, 24,* 539–562.

Abdi, H., Valentin, D., & O'Toole, A. J. (1997). A generalized autoassociator model for face processing and sex categorization: From principal components to multivariate analysis. In D. S. Levine & W. R. Elsberry (Eds.), *Optimality in biological and artificial networks? The International Neural Networks Society series* (pp. 317–337). Hillsdale, NJ: Erlbaum.

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

Busey, T. A. (1998). Physical and psychological representations of faces: Evidence from morphing. *Psychological Science, 9,* 476–483.

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.

Clark, S. E. (1988). *A theory of classification and memory retrieval.* Unpublished doctoral dissertation, Indiana University.

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.

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

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

Estes, W. K., & Maddox, W. T. (1995). Interactions of stimulus attributes, base rates, and feedback in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1075–1095.

Franks, J. J., & Bransford, J. D. (1971). Abstraction of visual patterns. *Journal of Experimental Psychology, 90,* 65–74.

Galton, F. (1878). Composite portraits. *Nature, 18,* 87–100.

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

Heit, E. (1993). Modeling the effects of expectations on recognition memory. *Psychological Science, 4,* 244–252.

Heit, E. (1994). Models of the effects of prior knowledge on category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1264–1282.

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

Homa, D., Goldhardt, B., Burruel-Homa, L., & Smith, J. C. (1993). Influence of manipulated category knowledge on prototype classification and recognition. *Memory & Cognition, 21,* 529–538.

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

Kayser, A. (1985). *Heads.* New York: Abbeville Press.

Kroll, N. E. A., Knight, R. T., Metcalfe, J., Wolf, E. S., & Tulving, E. (1996). Cohesion failure as a source of memory illusions. *Journal of Memory and Language, 35,* 176–196.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling.* Beverly Hills, CA: Sage.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240.

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.

Loftus, E. F. (1977). Shifting human color memory. *Memory & Cognition, 5,* 696–699.

Maddox, W. T., & Estes, W. K. (1997). Direct and indirect stimulus-frequency effects in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 539–559.

Metcalfe, J. (1990). Composite holographic associative recall model (CHARM) and blended memories in eyewitness testimony. *Journal of Experimental Psychology: General, 119,* 145–160.

Metcalfe, J., & Fisher, R. P. (1986). The relation between recognition memory and classification learning. *Memory & Cognition, 14,* 164–173.

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

Neumann, P. G. (1977). Visual prototype formation with discontinuous representation of dimensions of variability. *Memory & Cognition, 5,* 187–197.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 104–114.

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., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 282–304.

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9,* 247–255.

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

Reed, S. K. (1978). Category vs. item learning: Implications for categorization models. *Memory & Cognition, 6,* 612–621.

Robinson, K. J., & Roediger, H. L., III. (1997). Associative processes in false recall and false recognition. *Psychological Science, 8,* 231–237.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803–814.

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

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.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin and Review, 4,* 145–166.

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

Solso, R. L., & McCarthy, J. E. (1981). Prototype formation of faces: A case of pseudo-memory. *British Journal of Psychology, 72,* 499–503.

Steyvers, M. (2000). *Modeling semantic and orthographic similarity effects on memory for individual words.* Unpublished doctoral dissertation, Indiana University.

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.

# Appendix A

## Munsell Specifications and Stimulus Red, Green, and Blue (RGB) Values for Experiments 2–4

Table A1
*The Munsell Hue, Saturation, and Brightness Specifications and RGB Values for Each Stimulus in Experiment 2*

| Stimulus | Stimulus type | Hue | Saturation | Brightness | R | G | B |
|---|---|---|---|---|---|---|---|
| 1 | Region 1 old | 2.5 BG | 6 | 10 | 0 | 198 | 161 |
| 2 | Region 1 proto | 2.5 BG | 7 | 8 | 91 | 233 | 193 |
| 3 | Region 1 old | 2.5 BG | 8 | 6 | 150 | 255 | 222 |
| 4 | Region 2 old | 2.5 B | 5 | 2 | 109 | 140 | 141 |
| 5 | Region 2 proto | 2.5 B | 6 | 4 | 109 | 178 | 184 |
| 6 | Region 2 old | 2.5 B | 7 | 6 | 115 | 217 | 229 |
| 7 | Region 3 old | 5 PB | 2 | 6 | 14 | 53 | 97 |
| 8 | Region 3 proto | 5 PB | 3 | 8 | 0 | 75 | 141 |
| 9 | Region 3 old | 5 PB | 4 | 10 | 0 | 105 | 190 |
| 10 | Region 4 old | 7.5 P | 6 | 10 | 208 | 134 | 215 |
| 11 | Region 4 proto | 7.5 P | 7 | 8 | 236 | 168 | 241 |
| 12 | Region 4 old | 7.5 P | 8 | 6 | 255 | 205 | 255 |
| 13 | Region 5 old | 10 RP | 3 | 8 | 154 | 0 | 43 |
| 14 | Region 5 proto | 10 RP | 4 | 6 | 165 | 62 | 79 |
| 15 | Region 5 old | 10 RP | 5 | 4 | 171 | 104 | 115 |
| 16 | Region 6 old | 2.5 YR | 6 | 10 | 255 | 108 | 0 |
| 17 | Region 6 proto | 2.5 YR | 7 | 8 | 255 | 133 | 80 |
| 18 | Region 6 old | 2.5 YR | 8 | 6 | 255 | 188 | 145 |
| 19 | Region 7 old | 5 Y | 3 | 4 | 97 | 72 | 0 |
| 20 | Region 7 proto | 5 Y | 4 | 6 | 127 | 99 | 0 |
| 21 | Region 7 old | 5 Y | 5 | 8 | 157 | 128 | 0 |
| 22 | Region 8 old | 7.5 GY | 6 | 4 | 144 | 177 | 104 |
| 23 | Region 8 proto | 7.5 GY | 7 | 6 | 170 | 212 | 98 |
| 24 | Region 8 old | 7.5 GY | 8 | 8 | 202 | 255 | 87 |
| 25 | Foil | 7.5 BG | 9 | 2 | 227 | 255 | 255 |
| 26 | Foil | 10 B | 2 | 2 | 51 | 55 | 54 |
| 27 | Foil | 10 PB | 2 | 6 | 62 | 34 | 89 |
| 28 | Foil | 5 RP | 6 | 12 | 255 | 81 | 162 |
| 29 | Foil | 7.5 R | 5 | 14 | 255 | 4 | 0 |
| 30 | Foil | 7.5 YR | 7 | 16 | 255 | 144 | 0 |
| 31 | Foil | 10 Y | 8.5 | 12 | 255 | 255 | 0 |
| 32 | Foil | 5 G | 5 | 10 | 0 | 165 | 86 |

*Note.* proto = prototype; B = blue; G = green; Y = yellow; R = red; P = purple.

(*Appendixes continue*)

Table A2
*The Munsell Specifications and RGB Values for Each Stimulus in Experiments 3 and 4*

| Stimulus | Stimulus type | Hue | Saturation | Brightness | R | G | B |
|---|---|---|---|---|---|---|---|
| 1 | Region 1 | 7.5 GY | 6 | 4 | 147 | 170 | 112 |
| 2 | Region 1 | 7.5 GY | 6 | 8 | 119 | 183 | 54 |
| 3 | Region 1 | 7.5 GY | 8 | 4 | 216 | 239 | 175 |
| 4 | Region 1 | 7.5 GY | 8 | 8 | 206 | 248 | 127 |
| 5 | Region 2 | 5 Y | 5 | 4 | 142 | 121 | 53 |
| 6 | Region 2 | 5 Y | 5 | 8 | 147 | 117 | 0 |
| 7 | Region 2 | 5 Y | 7 | 4 | 205 | 191 | 125 |
| 8 | Region 2 | 5 Y | 7 | 8 | 216 | 179 | 44 |
| 9 | Region 3 | 2.5 YR | 5 | 8 | 186 | 72 | 9 |
| 10 | Region 3 | 2.5 YR | 5 | 12 | 210 | 71 | 0 |
| 11 | Region 3 | 2.5 YR | 7 | 8 | 255 | 147 | 115 |
| 12 | Region 3 | 2.5 YR | 7 | 12 | 255 | 137 | 76 |
| 13 | Region 4 | 10 RP | 4 | 6 | 154 | 63 | 78 |
| 14 | Region 4 | 10 RP | 4 | 10 | 183 | 30 | 73 |
| 15 | Region 4 | 10 RP | 6 | 10 | 253 | 120 | 143 |
| 16 | Region 4 | 10 RP | 6 | 6 | 214 | 134 | 148 |
| 17 | Region 1 proto | 7.5 GY | 7 | 6 | 173 | 204 | 118 |
| 18 | Region 2 proto | 5 Y | 6 | 6 | 177 | 155 | 51 |
| 19 | Region 3 proto | 2.5 YR | 6 | 10 | 238 | 119 | 61 |
| 20 | Region 4 proto | 10 RP | 5 | 8 | 200 | 91 | 111 |
| 21 | Region foil[a] | 7.5 GY | 9 | 4 | 246 | 255 | 216 |
| 22 | Region foil[a] | 5 Y | 3 | 4 | 84 | 69 | 9 |
| 23 | Region foil[a] | 2.5 YR | 3 | 6 | 120 | 53 | 22 |
| 24 | Region foil[a] | 10 RP | 8 | 6 | 255 | 213 | 222 |
| 25 | Isolated foil | 7.5 BG | 9 | 2 | 227 | 255 | 255 |
| 26 | Isolated foil | 10 B | 2 | 2 | 51 | 55 | 54 |
| 27 | Isolated foil | 10 PB | 2 | 6 | 62 | 34 | 89 |
| 28 | Isolated foil | 6.25 PB | 5 | 12 | 96 | 142 | 228 |
| 29 | Isolated foil | 7.5 R | 5 | 14 | 255 | 4 | 0 |
| 30 | Isolated foil[a] | 10 YR | 8 | 14 | 255 | 144 | 0 |
| 31 | Isolated foil | 10 Y | 8.5 | 12 | 255 | 255 | 0 |
| 32 | Isolated foil[a] | 5 G | 5 | 10 | 0 | 165 | 86 |

*Note.* proto = prototype; B = blue; G = green; Y = yellow; R = red; P = purple.
[a] Items not used in Experiment 4.

# Appendix B

## Verification of Scanning Procedure

The purpose of this study was to establish a technique for creating on computer displays color stimuli that are at known psychological distances from each other. Munsell color chips provide an ideal stimulus set because extensive psychological scaling of these stimuli has already been conducted. However, use of these chips does not allow the advantages of computer-controlled experiments. The goal of Experiment 1 was to test whether Munsell color chips scanned into a computer would maintain their respective perceptual distances and psychological structure. Therefore we scanned these colors and then collected similarity ratings to ensure that the relationship between various color chips in Munsell space corresponded to the relationship in the obtained MDS solution. More specifically, the idea was to test whether certain types of stimulus triplets used in the experiments in the text would be represented such that the prototype would be centrally located between the other two items.

### Method

*Participants*

The participants in the experiment were 51 Indiana University undergraduate students. They participated to satisfy an introductory psychology course requirement.

*Stimuli*

The stimuli were 12 colors presented on a computer screen. To arrive at these computer displayed colors, Munsell color chips from the 5 Red hue were scanned into Adobe Photoshop using a HP ScanJet 6200C. The obtained red, green, blue (RGB) values were used as input to the experimental program. The Munsell configuration for the 12 colors is illustrated in Panel A of Figure B1.

*Procedure*

Prior to the similarity rating phase, participants were shown each of the 12 colors, one at a time, on a computer screen and were instructed to merely familiarize themselves with the range of colors in the experiment. Following this phase, participants were asked for similarity ratings to all pairwise colors. All 66 unique pairs of colors were displayed with the right/left placement randomly determined. Pairs of colors appeared in a different random order for each participant. The two colors appeared on the screen as 2-in. squares separated by 1 in. Participants made similarity ratings using a 9-point scale and were instructed to try to use the full range of ratings. Responses on the scale ranged from 1 (*least similar*) to 9 (*most similar*). Participants made ratings in two blocks of these colors.

### Results and Discussion

A multidimensional scaling solution was derived by fitting the standard Euclidean model to the mean similarity ratings. The two-dimensional scaling solution provided a reasonable fit to the data, accounting for 99.6% of the variance. The solution had a stress level of 0.023. A plot of the derived solution (rotated to achieve maximal correspondence with the Munsell configuration) is shown in Panel B of Figure B1. The correspondence between the solution and the Munsell values is reasonable. More important, however, certain relationships critical to the design of the following experiments appear to hold. That is, we were interested in verifying that three contiguous items on a diagonal of the brightness and saturation space would be represented such that the middle color would be approximately centrally located. This relationship can be checked in a number of triplets [(1,3,6); (2,4,7); (3,6,9); (5,8,11); (2,3,5); (4,6,8); (7,9,11); (6,8,10)] and generally holds quite well. Therefore, it appears that the scanning procedure used in these experiments is adequate for our research goals.

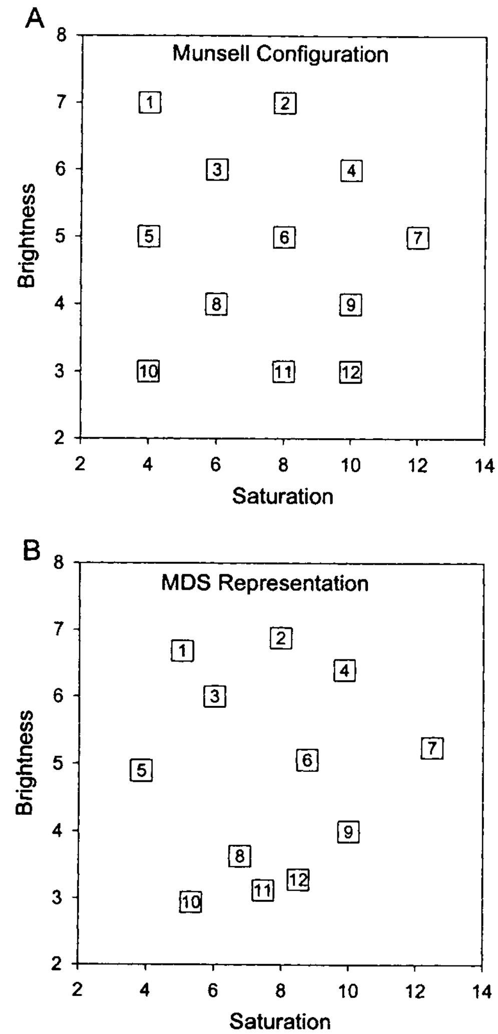


*Figure B1.* The schematic illustration of the Munsell configuration used in the scanning verification study reported in Appendix B (A) and the obtained two-dimensional solution (B). MDS = multidimensional scaling.

Received May 5, 2000
Revision received September 28, 2000
Accepted December 7, 2000 ■