

# Categorical Perception Effects Induced by Category Learning

Kenneth R. Livingston and Janet K. Andrews  
Vassar College

Stevan Harnad  
University of Southampton

The authors report a series of studies designed to determine whether effects similar to those observed in the innate categorical perception of color and phonemes are induced during the learning of simple unidimensional categories and more complex multidimensional ones. In Experiment 1 no evidence was found for such effects when stimuli varied on 1 dimension. Experiments 2 and 3 demonstrated a within-category compression effect but no between-category expansion effect for stimuli varying in 2 dimensions. Compression only was also shown in Experiment 4, which used pictures of actual objects. Multidimensional scaling analyses illustrate how within-category compression without expansion was sufficient to produce categorical clustering of items in the similarity space. These analyses also show that learning changed the dimensional structure of similarity space. Results are compared with those from other studies exploring similar phenomena and with neural network simulations.

Most contemporary models of category learning rely to some degree on relative similarities between category members and nonmembers (Bruner, Goodnow, & Austin, 1956; Carey, 1985; Estes, 1986; Keil, 1990; Komatsu, 1992; Medin, 1989; Medin & Schaffer, 1978; Murphy & Medin, 1985; Nosofsky, 1988; Pazzani, 1991; Posner & Keele, 1968; Rosch & Lloyd, 1978). A psychological similarity space is generally assumed, its axes defined by values on relevant dimensions (e.g., size, color, shape), along which objects or events might vary. In principle, every object or event can be located in this  $n$ -dimensional space, and its distance from other objects can be quantified (Hutchinson & Lockhead, 1977; Nosofsky, 1989; see also Tversky, 1988). In practice, it is often difficult to establish the metric of such spaces a priori, especially for natural objects. Nevertheless, some variation on exactly this procedure is used in the testing of most major models of categorization. Explanation-based theory first seemed to be an exception in this regard, having originally been formulated as a critique of similarity-based approaches to categorization (Murphy & Medin, 1985), but more recently it has been acknowledged that even an explanation-based approach is anchored in similarity comparisons at some point (Goldstone, 1994b; Medin, 1989; Medin, Goldstone, & Gentner, 1993).

Despite the importance of the idea of a psychological similarity space in almost all approaches to categorization, little direct work has been done to explore the properties of such spaces in the context of categorization. The default assumption is that these spaces have a fixed dimensional structure and that each dimension has a fixed and generally linear metric. Implicit in this view is the idea that any given object has a determinate location in this space and that its proximity to other items, including prototypes or other summary central tendencies, can be calculated in a straightforward fashion. The salience of various dimensions may shift as a function of category training, but the space itself is assumed to be an unchanging constant.

The problem with this view is that it does not explain how or why the set of items that comprise a category seem to be so much more coherent after one has learned that they are members of the same category (Komatsu, 1992). A possible explanation of this effect would be that the space undergoes some form of transformation or "warping" as a result of category training, with members of the same category moving closer together than they were before the training. Such "chunking" effects may be an important component of concept formation itself. For example, if category learning results in a warping of the space that compresses one region of a dimension, thereby decreasing the psychological distance between items in that region, this is tantamount to abstracting away from the stimulus values of the items on that dimension (within that region of the dimension). This is exactly the sort of effect that is required in order to understand how concepts manage to be abstract (i.e., treat distinct exemplars as equivalent, or relatively so compared with items lying outside the compressed region) while nevertheless referring to concrete particulars. There is a line of research that suggests that such a warping process is at least plausible. This framework for thinking about representational changes induced by categorization comes from research on categorical perception (Harnad, 1987).

Categorical perception is a psychophysical effect in which a continuous physical dimension such as intensity (of

---

Kenneth R. Livingston and Janet K. Andrews, Department of Psychology, Vassar College; Stevan Harnad, Department of Psychology, University of Southampton, Southampton, England.

This research was supported in part by grants from the Pew Charitable Trusts (T86-06597006) and by the Vassar College Undergraduate Research Summer Institute. We thank Ute Fischer and numerous Vassar College undergraduates for their assistance with these studies, with special thanks to Curt Degenhart and Gary Lubow. We also thank Barbara Malt and James Beale, and two anonymous reviewers for their helpful comments on an earlier version of this article. Portions of this work were presented at the Annual Meeting of the Society for Philosophy and Psychology, June 1994.

Correspondence concerning this article should be addressed to Kenneth R. Livingston, Department of Psychology, Vassar College, Poughkeepsie, New York 12604. Electronic mail may be sent to [livingst@vassar.edu](mailto:livingst@vassar.edu).

mechanical or electromagnetic energy) or frequency (of electromagnetic radiation or of acoustic vibration) is partitioned into discontinuous or bounded regions by some kind of perceptual mechanism. Color categories (Bornstein, 1987) and phoneme categories (e.g., Pastore, 1987) have been the two most fully studied cases of categorical perception.

In color perception, the gradations in the wavelength of light are continuous, but they are not perceived as continuous. Rather, the chromatic spectrum is segmented by the visual system into qualitatively distinctive categories of hue (Bornstein, 1987). Instead of being homogeneous (i.e., log-linear, like most psychophysical continua that obey Weber's Law), the spectral continuum is inhomogeneous, or warped, in the sense that it requires a larger change in wavelength to produce a just-noticeable difference (JND) in some regions than in others. Multiple values on the dimension are effectively chunked together in the interior of a color category. In other regions of the spectrum, at the category boundaries, a smaller relative change in wavelength is sufficient to produce a JND.

These kinds of effects also appear to occur in the perception of certain speech sounds. For example, voice-onset time (VOT) is the interval between the plosive release of air that marks the beginning of certain phonemes and the buzzing of the larynx that gives voice to these basic sounds. The phonemes /d/ and /t/, for example, differ in VOT when followed by a vowel: /ta/ sounds have longer VOTs than /da/ sounds. This difference in voicing is perceived categorically as either nonvoiced (/ta/) or voiced (/da/). Artificially generated versions of these phonemes can provide a continuous series of sounds differing only in VOT, but people hear the continuous series as dividing into just two categories: voiced /da/ sounds and nonvoiced /ta/ sounds (Eimas & Corbit, 1973). As in the color example, the effect consists of relative compressions of large regions of the VOT continuum and expansions at what is perceived to be the category boundary.

The mechanism responsible for the categorical perception of color, and for fixing the location of color boundaries, seems to be innate (see Bornstein, 1987). Some speech researchers claim that the same is true for phoneme categories (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971), but there is also evidence for some plasticity in both the location (e.g., Williams, 1977) and the sharpness (see Pastore, 1987) of speech category boundaries as a function of training and practice. Even in the case of color categories there are data to indicate that experience can induce some tuning of hue boundaries (Raskin, Maital, & Bornstein, 1983). It thus appears that even in the two classical cases, category boundaries are not always rigidly specified.

The fact that category boundaries remain plastic to some degree even in these two biologically prepared domains suggests that in less constrained domains categorical perception effects might be inducible by learning alone, without benefit of inborn boundaries. Given that brains can generate categorically perceived boundaries with dedicated neural structures, it is plausible that learning can reorganize less dedicated structures to generate similar categorical boundaries de novo during category learning. To make new,

biologically unprepared absolute judgments (Miller, 1956) among sets of highly interconfusable objects requires generating, on the basis of the objects' analog sensory projections, a categorical (effectively digital) output indicating the category to which each object belongs. This is precisely what the categorical perception mechanism does in the biologically prepared case of hue and voicing. The question is whether learning new categories in a domain is associated with a similar warping of the similarity space. If it is, then one could speak more generally of compression effects whenever a dimension (or portion of a dimension) of a psychological similarity space is warped so as to require larger changes to produce a JND (compared with a prewarped baseline). One could also speak generally of expansion effects whenever a dimension is warped so as to require smaller changes to produce a JND (again, compared with a prewarped baseline).

Our hypothesis is, accordingly, that categorical perception effects can be induced by learning alone, without benefit of an innate mechanism. More specifically, we suggest that acquiring the ability to categorize objects and events is associated with a warping of the psychological similarity space such that during category learning some regions become compressed and others expanded, relative to their precategorization configuration.

Two different lines of research, one from neural network simulations and the other from human experimental work, offer some support for the learned categorical perception hypothesis. First, Harnad and colleagues (Harnad, Hanson, & Lubin, 1991, 1995; see also Goldstone, Steyvers, & Larimer, 1996) used the back-propagation algorithm to train simple, three-layer nets to categorize inputs coded to correspond to lines of different lengths. The nets were first required to match their inputs exactly by producing as output a line identical in length to the input (a process called *auto-association*). Each input produces a pattern of activation in the hidden units that can then be thought of as representing that input uniquely as a point in  $n$ -dimensional hidden-unit space. Auto-association training has the effect of maximizing the pairwise interstimulus distance for all of the inputs in this hidden-unit space. This learning is not the focus of interest here, however: auto-association merely gets the nets to a state that is comparable with that of the normal human who can already do reliable pairwise sensory matching before learning any categories. (You may not yet know which lines are short, medium, or long, but you can already draw a line of the same length as any line presented to you: That is auto-associative matching, and the net, unlike the human, must first learn to do it.)

Once trained to auto-associate, the net is trained to sort the input lines into either two categories (short vs. long; Harnad et al., 1991) or three categories (short, medium, or long; Harnad et al., 1995), while still maintaining correct auto-association performance. Additional output units serve as the markers for category membership. Categorization training produces strong compression-expansion effects in hidden-unit space, as measured by comparing the interstimulus distances for auto-association alone with those for auto-association together with categorization. It thus appears to

be a natural feature of back-propagation or gradient descent networks to produce categorical perception-like compression–expansion effects in the course of category learning.

Data from neural network simulations support the plausibility of learned categorical perception, and the ability to examine the inner workings of the nets to understand how the effect is produced has considerable heuristic value, but evidence from a second line of research—research with human category learners—is even more directly relevant to the hypothesis that category learning is based on compression and expansion effects in similarity space. For example, Beale and Keil (1995, 1996) have found strong categorical perception effects along a continuum of faces produced by morphing John Kennedy’s face into Bill Clinton’s. These effects are the result of learning because they are reduced when the morphed faces are less familiar and absent when they are completely unfamiliar (see also Calder, Young, Perrett, Etcoff, & Rowland, 1996, and Etcoff & Magee, 1992, for analogous morphing boundaries for facial expression). A shortcoming that the morphing data share with the color and phoneme data is that, unlike the case of the network simulations, there is no precategorical baseline condition for comparison with the groups showing categorical perception effects. All that can be said is that instead of a log-linear Weberian continuum, one observes regions that, relative to each other, display compression and expansion; it is not known how those regions looked precategorically. But if categorical perception can be induced by learning in human as it can in neural nets, then with appropriate baseline control groups it should be possible to test whether learning-induced warping is associated with compression, expansion, or both.

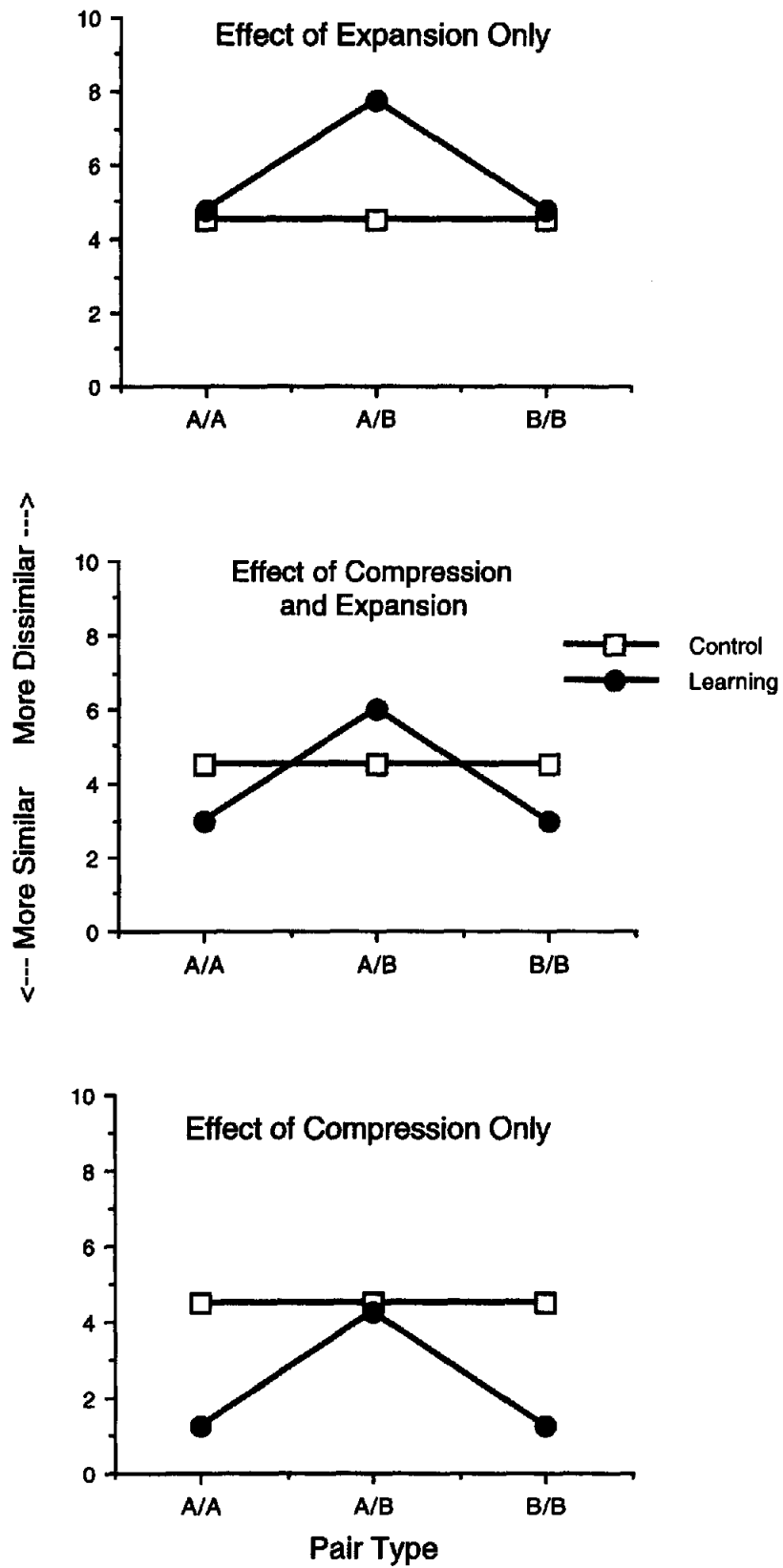
Learning-induced changes in the dimensional structure of psychological similarity space cannot be observed directly in humans as they can in neural nets; they can only be inferred. If category learners judge items in the same category as more similar than do control participants, we will interpret this as compression; if category learners judge items in different categories to be less similar than do control participants, this will be interpreted as expansion. Figure 1 illustrates the three possible forms of warping (compression only, expansion only, and both) using hypothetical plots of a similarity metric for pairs of items. Note that in these graphs higher numbers on the y-axis indicate lower similarity (greater dissimilarity). For each plot, three types of pairs are depicted: (a) those in which both members of the pair are from Category A, (b) those in which each member of the pair is from a different category, and (c) those in which both are from Category B. As the figure shows, compression alone, expansion alone, and both together have distinctive profiles. Figures 2a–2c show one way in which compression and expansion could be dissociated in a two-dimensional similarity space.

Using same–different judgments as the measure of similarity, Goldstone (1994a) did compare pre- and postlearning groups and reported finding “acquired distinctiveness” (what we are calling expansion effects) at learned category boundaries (see also Lawrence’s, 1950, “acquired distinctiveness of cues”). In Goldstone’s first experiment, the stimuli

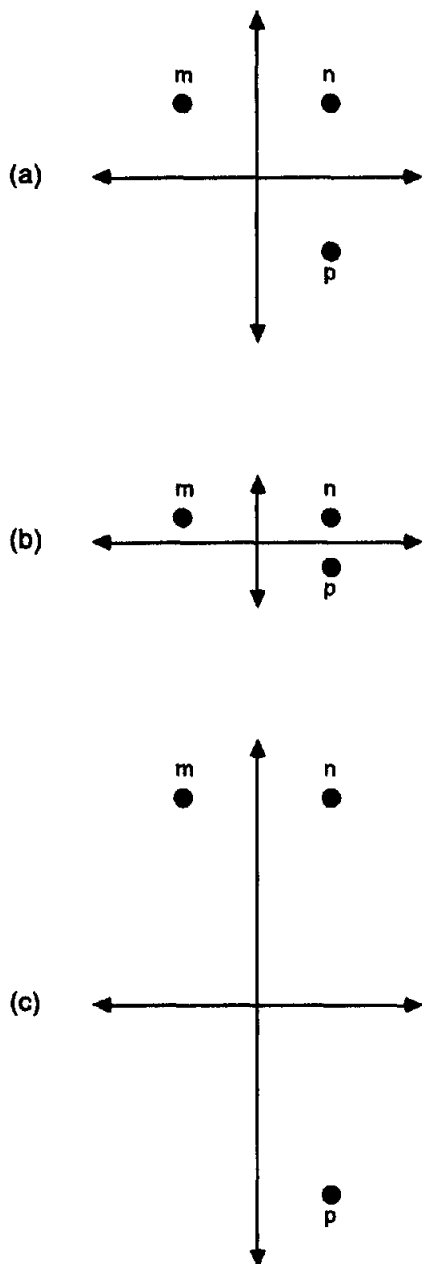
were squares varying in size and brightness, and in a later experiment he found similar effects using squares varying in saturation and brightness. He found no evidence of compression within category-relevant dimensions, though it did sometimes occur on dimensions not relevant to categorization. More recently, Kurtz (1996) compared the similarity judgments of a no-training control group with those of a group trained to classify simple drawings and found evidence for compression but not expansion. Stevenage (1998) compared people’s similarity judgments of photographs of identical twins before and after learning to distinguish them and found evidence for both compression and expansion.

Similar effects can also be found in the social categorization literature, particularly the work based on what Tajfel (1957, 1959, 1978) dubbed “accentuation theory.” According to this theory, when people are categorized into groups, between-group differences are accentuated (what we are calling an expansion effect, to emphasize the change in the similarity space), whereas within-group differences are minimized (what we are calling compression). The theory has been extended to a wide range of social judgment phenomena in recent years (see Eiser, 1996; McGarty & Penny, 1988), where it is generally assumed, without direct test, that compression–expansion effects are operative. Accentuation theorists believe that these effects reflect more general cognitive processes, largely on the basis of findings by Tajfel and Wilkes (1963) on a line-length estimation task. People in their experimental group were asked to view the lines labeled as either *As* or *Bs* to distinguish the short and long categories. In two control groups, people viewed the same lines either without labels or with labels assigned to them randomly (no relationship to line length). Tajfel and Wilkes found evidence for what they called contrast effects (expansion effects) at the boundary between short and long lines among people in the experimental group.

In summary, several lines of research bear on the proposal that categorization is associated with a warping of the psychological similarity space into which category instances are mapped. Neural network simulations show compression and expansion effects as a function of training and allow the direct examination, for heuristic purposes, of how item representations change during training. Related effects have been found in studies using human category learners, although the pattern of such effects is in some ways difficult to interpret. In many cases it is not possible to tell, relative to a baseline, whether the warping effects are the result of compression, expansion, or both, because prelearning control groups are not available. In the few cases where control groups are used, the effects are inconsistent. Sometimes there are compression effects alone, sometimes expansion alone, and sometimes both. Furthermore, some doubt is cast on at least one of these findings (Tajfel & Wilkes, 1963) by Richardson (1991), who failed to find either compression or expansion among people trained to categorize lines as long or short. Indeed, he found that people had great difficulty learning to make the long–short distinction. The scope and character of learned categorical perception effects thus remains unclear in many important respects, and until there is greater clarity about these matters, it is not possible to



*Figure 1.* Hypothetical plots of the learned categorical perception effect. In all cases the similarity ratings of control (nonlearning) participants are compared with those of people who learn a category distinction. The three possible patterns are (a) expansion at the category boundary without compression of items within the category, (b) both expansion and compression, and (c) compression within categories without expansion at the boundary.



*Figure 2.* In (a), item *n* is equidistant from items *m* and *p*. In (b), the compression of one dimension moves items *n* and *p* closer together (increased similarity) without increasing their distance from *m*. In (c), on the other hand, the result of learning is to expand one dimension, resulting in a greater dissimilarity between *p* on the one hand and *m* and *n* on the other, the latter two items maintaining their prelearning distance from one another.

explore the potential of this approach for understanding the relationship between similarity and categorization.

We begin the report of our investigations with a focus on whether these effects can be demonstrated reliably for stimuli varying in only one dimension. The one-dimensional case is particularly important because classical categorical

perception (color, phonemes) occurs for stimuli varying in one dimension. In addition, the learned version of the effect for line length, the foundation for accentuation theory in social psychology, is controversial, given Richardson's (1991) recent failure to find evidence for it. Finally, given the finding of learned categorical perception effects using stimuli of higher dimensionality (e.g., Goldstone, 1994a), it is important to know whether these effects are inherently multidimensional, or whether they might be understood as the result of the combination of many one-dimensional effects. Following a report on the one-dimensional case, we report results from experiments using stimuli of higher dimensionality, including one study using very complex real-world categories that allow multidimensional scaling (MDS) analyses as a way of analyzing directly the effects of category learning on dimensional structure. These studies extend the phenomenon to more ecologically valid categories and situations, and do so using control groups of people who have not learned the categories. Such controls are generally missing from the few studies that have been done in this area (see Goldstone, 1994a, and Kurtz, 1996, for exceptions).

### Experiment 1

In an effort to explore the expansion phenomenon observed by Tajfel for the one-dimensional case, and mindful that Richardson's (1991) work on line length was already ongoing, we worked with a number of other stimulus dimensions in our laboratory, including circle diameter, texture density, gray-scale saturation, and the size of notches cut from a square figure. Several efforts to calibrate the magnitude of the differences along these dimensions revealed that every difference we constructed resulted in one of two learning patterns: (a) In the first pattern it proved impossible to learn the distinction at or near the boundary within the 240 to 480 (depending on the study) learning trials offered, making it impossible to test for learning-induced expansion or compression effects. (b) In the second pattern the distinction was trivially easy to learn, so that the learning curve jumped to an asymptote immediately after one sampling of the range of variation. In the latter case, the fact that no real learning was required made it impossible to test the hypothesis that learning could produce the effects. These patterns themselves seemed categorical, with no discernible transition from one pattern to the other, and are very similar to those reported by Richardson.

In the course of these investigations, however, we came across another report in the literature that could be seen as relevant to the question of whether there is learned expansion or compression of similarity space on a single dimension. This study was reported as Experiment 2 in Cross, Lane, and Sheppard (1965), and was done as part of an effort to test the motor theory of speech perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). According to the motor theory, the perception of speech sounds is mediated by the movements that produce them. For example, stop consonants (/ba/, /da/, /ga/) are perceived categorically because the articulations that generate them are

categorical. To challenge this hypothesis, Cross et al. (1965) trained participants to sort visual stimuli into two categories, using two different consonant–vowel–consonant triads as labels. The labels (/b^b/ and /g^g/, rendered as *bub* and *gug*, respectively) were used as a kind of a caricature of the motor theory, because of course the stimuli themselves were unpronounceable, being visual and not acoustic. The visual stimuli consisted of circles with one of four pie-like wedges of different sized angles cut from the center out to the circumference symmetrically around the 12 o'clock position. The two narrowest wedges were assigned to one category and the two widest to the other. The participant had to learn to say "bub" when either of the narrowest wedges appeared and "gug" in response to the two widest. If the two stimuli labeled *bub* and the two labeled *gug* were perceived categorically after the training, this would refute the motor theory of speech perception because it would show that discrete motor production of the stimuli themselves was not the cause of the categorical perception; rather, the learning of discrete responses was. (Lane, 1965, explicitly related this to Lawrence's, 1950, behaviorist analysis of the acquired distinctiveness and similarity of cues, which was later elaborated in Gibson's, 1969, work on perceptual learning.)

Cross et al. (1965) trained 4 participants to a criterion of 50 correct successive categorizations of the four stimuli. Following training, all 4 participants were able to classify them with near perfect accuracy, so the authors were confident that category learning had taken place. Participants were then tested for their accuracy in discriminating them in an ABX paradigm, in which the participant sees two different stimuli, A and B (in this case successively) that might be drawn from the same or different categories (i.e., two different bubs, two different gugs, or a bub and a gug). Then, a third stimulus, X, is presented, which is identical either to A or B; the participant's task is to indicate which. A categorical perception effect occurs if experimental participants perform better when A and B are members of different categories (i.e., had different names) than when they are members of the same category (i.e., *bub–bub* or *gug–gug*).

Cross et al. (1965) reported finding learned categorical perception. Because they tested only 4 participants, no statistical analyses were reported, but all 4 showed the predicted pattern, with better discrimination accuracy when A and B were drawn from different sides of the category boundary than when they were drawn from the same side (the expansion effect).

Given the difficulties we encountered in teaching people categorical distinctions for several different unidimensional stimuli, and Richardson's (1991) similar failures for line length, it seemed to us important to replicate this one clear finding of something like the unidimensional learned categorical perception effect. Several efforts to replicate using computer-generated versions of the Cross et al. (1965) stimuli were unsuccessful. Variations in interstimulus intervals (ISIs), feedback for correct performance, and sizes of the wedges had no effect. We considered the possibility that having participants view the stimuli on a computer monitor might somehow account for the difference, so we decided to

try to replicate their results using the same kinds of equipment used in the original study. The methods described below are identical to those reported in Cross et al. (1965), except as noted.

## Method

**Participants.** Participants were 6 undergraduates who were paid for their participation.

**Stimuli and apparatus.** Wedges of 42°, 46°, 50°, and 54° were cut from 3-in. diameter black circles; these were oriented symmetrically relative to a vertical radius from the center of the circle. The circles were transferred to slides. Extreme care was taken by the slide maker to ensure that all images were precisely centered in their frames and that all visible blemishes were removed from the positive image from which the copies were made. In the only significant deviation from the original Cross et al. (1965) procedure, we made multiple copies of each wedge/angle rather than presenting the same four slides repeatedly in random sequence. Twenty copies of each wedge/angle were made for the training session and an additional 27 for the triads in the ABX procedure. The images were back-projected onto a screen 24 in. from the plane of the participant's eyes. The projected circle was 10 in. in diameter with the center adjusted to a height of 5 in. above eye height.

Slides were projected by a Model 8-2 Kodak carousel slide projector controlled by a cassette-driven electronic control system (Audiotronics Model No. 162 S-2, Audiotronics Corp., North Hollywood, California) interfaced to the projector.

**Procedure.** After being positioned in front of the screen, the participant was told that various shapes, differing only in the size of the wedge cut out of them, would appear on the screen. The participant's task was to identify each shape as it appeared by saying aloud either /b^b/ or /g^g/. When the participant correctly identified the shape, the sound of a nickel dropping into a glass jar could be heard (Cross et al., 1965, had used a penny dispenser, but clearly some correction for inflation was in order). The sound of the nickel dropping was a digital recording generated by a Macintosh IIfx computer rather than being produced by actually dropping a coin into a jar. However, the participant was paid according to the agreement. The more reliably the stimuli were classified, the more money the learner could earn.

The two smaller wedges were classified as /b^b/, and the two larger wedges were classified as /g^g/. Stimulus presentation was one third of a second (Cross et al., 1965, used 300 ms) with a stimulus-onset asynchrony (SOA) of 5 s. The four stimuli were presented in random order until the participant gave 50 consecutive correct responses, or until 160 training trials were complete. (Cross et al.'s participants all learned the distinction to criterion within 120 trials.)

Following training, the participant was tested 96 times using all permutations of the four stimuli appearing in random order. The participant's task was to identify each stimulus as either a /b^b/ or a /g^g/, but no feedback was given. Stimulus presentation time and SOA were the same as in the training phase.

Following identification testing, the participant performed a series of standard ABX discrimination trials. The second stimulus always differed from the first, and the third or "X" stimulus was identical to either the first or the second. When the third stimulus was presented, the participant said "first" if it matched the first one, or "second" if it matched the second one. All pairs differed by only one step (i.e., the 42° wedge followed by the 46° wedge, or the 54° wedge followed by the 50° wedge), and each adjacent pair was presented a total of 36 times, 9 times in each of the four possible triad sequences (ABA, ABB, BAA, and BAB). Stimulus duration

and SOA were as in the previous two sessions. The interval between triads was 15 s.

### Results

Not one of our 6 participants met the criterion of 50 consecutive correct responses during the 160 trial training period. Although the proportion of correct responses to the two extreme wedges ( $42^\circ$  and  $54^\circ$ ) reached 98.6% during the identification test phase of the study, the proportion was considerably lower for the two wedges that straddled the category boundary. Performance did reach 79.8% correct for these two items during the identification test phase, but the error rate remained high and no participant even approached the criterion of 50 correct consecutive responses met by all 4 of Cross et al.'s (1965) participants. Furthermore, all of Cross et al.'s participants showed the distinctive pattern of enhanced accuracy on the ABX trials across the category boundary as compared to within category trials. None of our participants showed this pattern.

### Discussion

What might explain this failure to replicate a prior demonstration of the compression–expansion effect for a unidimensional stimulus set? We believe that the explanation probably lies in the one significant divergence between our procedure and that of Cross et al. (1965). They used a programmable slide projector to show the same four slides repeatedly in a random sequence. We also presented the four stimuli in a random sequence, but instead of repeating the same four slides we presented a total of 80 different slides, 20 for each of the four stimuli. Even though we and the slide maker were obsessively careful about keeping our slides free of idiosyncratic dust specks and scratches, such damage is almost impossible to avoid entirely. Our suspicion is that the four slides used by Cross et al. contained unique identifying features in the form of such blemishes, with the result that it was possible to learn to tell the slides apart by paying attention to these unintended markers. This could certainly explain their excellent identification data. Because the hypothesis with which we began is that the categorical perception effect can be produced as a result of learning, procedures or stimuli that result in no learning should not be able to produce the effect. Hence, the fact that we were unable to produce the effect in a situation in which the categories could not be learned to a high standard does not fail to support the hypothesis, it simply fails to test it.

This negative finding, together with the lack of evidence of learning-produced categorical perception in the one-dimensional stimulus domains tested by Richardson (1991) and in our laboratory, suggests that unidimensional learned categorical perception is at best difficult to demonstrate. The major problem has turned out to be the difficulty of generating a set of stimuli that is hard enough to categorize to require extended training yet not so hard as to be impossible to learn. As is always true of negative findings, this does not show that it is impossible to produce one-dimensional learned categorical perception, but the range of

negative findings suggests that if it is possible, it is a fragile phenomenon, unlikely to be robust enough to explain category learning in general.

One possible explanation for difficulty in producing one-dimensional learned categorical perception is the following. Suppose that the effect actually does alter the way things appear. Then there is reason to expect early perceptual processes to be relatively immune to modification by categorization training, at least in adults. As noted earlier, there is some plasticity in the location of phoneme and color category boundaries, each of which varies along a single dimension. Boundary movement is difficult to generate, however, requiring prolonged practice (Pastore, 1987), and it appears to take the form of short-term adaptation effects (Wilson, 1987) rather than a long-term change. The only long-term learned categorical perception effects, those in young infants, take the form of a weakening or loss of inborn category boundaries, for example, the loss of the *r/l* boundary in speakers of Japanese rather than the *de novo* creation of a categorical perception boundary (Kuhl, 1987, p. 358; but see Logan, Lively, & Pisoni, 1991, for evidence of some recoverability with intensive training). The relative rigidity of isolated sensory dimensions is likely to be due to the perceiving organisms' need for a stable link between the world and the elemental dimensions of (at least) early perceptual processing. It would be easier to maintain stability if early perceptual processes were relatively resistant to categorical reorganization, and the modularity (Fodor, 1983) or cognitive impenetrability (Pylyshyn, 1984) of early perception has been stressed by other theorists.

Nevertheless, it is clear from the work of Goldstone (1994a) and that of Beale and Keil (1995, 1996) discussed above that people do show learned categorical perception effects when learning to categorize multidimensional stimuli. This would make sense if categorization is the result of a higher order attentional process that selects from among the many dimensions available in a multidimensional perceptual array. In this case, expansion effects might occur in category learners, relative to noncategorizing controls, when category-relevant dimensions are discovered and previously unattended distinctions at the boundary are made salient (see Figure 2 for an abstract example of how this might work). Alternatively, perceived differences among stimuli within a category might be ignored as irrelevant to a category distinction in the service of enhancing relative expansion at the category boundary, resulting in within-category compression. Either effect alone, or both together, might occur even when the dimensions along which stimuli vary are of the same kinds that fail to generate the compression–expansion effect when varied singly. The remaining experiments thus used multidimensional stimuli.

### Experiment 2

Before we can begin to develop good hypotheses about their role in categorization, learned categorical perception effects need to be demonstrated and analyzed in a wider range of stimuli than those explored thus far. In addition, there are too few studies of the phenomenon that include

proper baseline control conditions. Such controls are necessary to determine whether compression, expansion, or both has occurred for the learning group. In the following study, we explored multidimensional learned categorical perception effects using schematic stimuli that resemble single-celled organisms. These are more like actual categories in the world than squares of varying brightness or saturation, but because they are constructed, dimensions of variation can be carefully controlled. To determine whether compression or expansion effects occurred, we relied on similarity judgments as the dependent measure. Pilot work showed that, compared with measures of discriminability, similarity judgments are more sensitive indices of learning-induced changes. Particularly with complex, multidimensional stimuli, two objects that could still be discriminated following category learning might nonetheless have become more psychologically similar, just as two objects that were discriminable before category learning might become more dissimilar.

### Method

**Participants.** Participants were 16 Vassar College undergraduates and staff who either volunteered their participation as part of an introductory psychology course requirement or were paid for their time.

**Stimuli.** Stimuli were designed to resemble microorganisms (see Figure 3) and varied on two dimensions, including the length of a set of hairlike projections, or cilia, and the shape of a headlike portion of the figure. Head shape was varied as follows. A simple irregular polygon, open on one side, was constructed with four vertices or inflection points. These inflection points were moved incrementally prior to smoothing so as to produce variations in the depth of the lower (Gex) and upper (Zof) indentations observable in the right-hand portion of the sample stimuli shown in Figure 3. The pictured indentations represent extremes on this dimension; intermediate values gradually shift the indentation from upper to lower. Extensive pretesting of the psychophysical scaling properties of these features (see Livingston & Andrews, 1995, for details) ensured that differences between adjacent values on each of the dimensions were equally discriminable and well above the JND threshold. All stimulus presentations and data recording were automated by means of a SuperCard program running on a Macintosh IIcx or IIfx computer.

Eight discrete values of each dimension were chosen for the design of the stimuli used in this study. Two categories were created and identified by the nonsense labels *Gex* and *Zof*, names chosen for their low associative strength. *Gexes* had the four shorter cilia values, and the *Gex* head shape tended toward an indentation of varying depth in the lower half of the figure. *Zofs* had the four longer cilia and a head shape that tended toward an indentation in the upper half of the figure. The two categories were, thus, perfectly separable with no overlapping values on either dimension, and the intervals between values on both dimensions were equally discriminable between all pairs, whether they fell within a category or across the arbitrarily chosen category boundary. All other features of the stimuli were nonvarying in this study.

**Procedure.** Participants were assigned randomly to either a learning or a baseline (control) condition. The procedure for the learning participants was as follows. The participant was seated with eyes 70 cm from the screen, the distance at which psychophysical scaling of the stimuli was done. Stimuli were presented in blocks of 32, 16 from each of the two categories. Order of

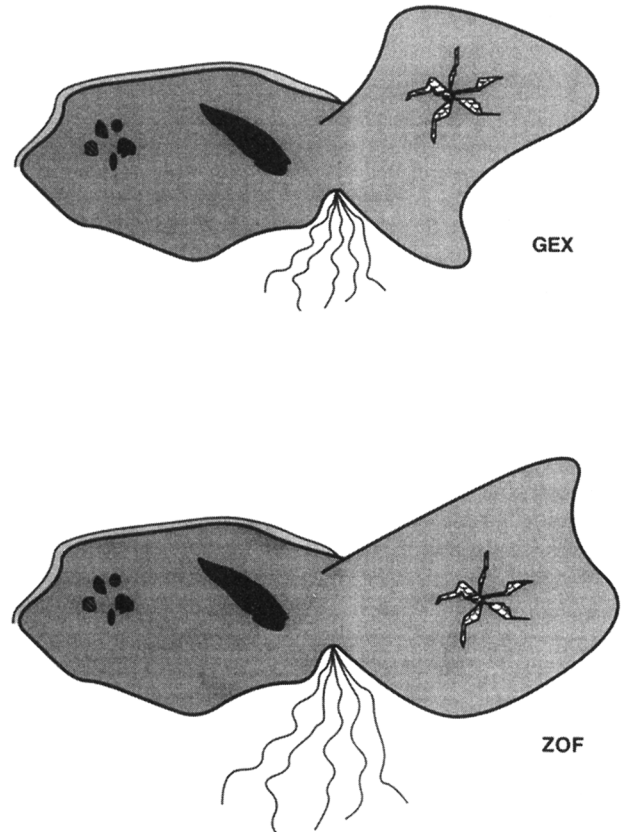


Figure 3. Examples of stimuli from the Gex and Zof categories from Experiments 2 and 3.

presentation within each block was determined randomly except that no more than four members of the same category could appear in sequence. Each stimulus appeared on screen for 2 s, and the participant indicated whether it was a *Gex* or a *Zof* by a keypress. Following the keypress, the participant heard either "Right, that was a (Gex/Zof)," or "Wrong, that was a (Gex/Zof)." These phrases were digitally recorded and played back simultaneously with the change in level of a bar chart at the bottom of the screen. The bar chart went up by one level for each correct response and went down by five levels for each error. The next stimulus in the series was presented 2 s after the feedback was delivered. All participants received 96 training trials, regardless of level of performance. Thereafter, the participant was judged to have reached criterion and training was stopped when all 32 stimuli in a block were identified correctly, or after 256 trials (eight trial blocks), whichever came first.

The similarity judgment task was the same for both learning participants and baseline or control-group participants; for the latter, this was the only task. The participant judged a total of 90 pairs of stimuli, including 30 *Gex*-*Gex* pairs, 30 *Zof*-*Zof* pairs, and 30 *Gex*-*Zof* pairs.<sup>1</sup> Using a city-block metric to define distances between items varying on two dimensions, we selected the within-category pairs randomly, with the constraint that the number

<sup>1</sup> Some pairs were presented for evaluation more than once in order to address issues not under discussion in this article. In those cases, only the first response to the pair was included in the computation of mean similarity ratings.



of pairs at each degree of separation be roughly equivalent. Between-category pairs were restricted to the same degrees of difference as the within category pairs so as not to artificially inflate the between-category differences, given that pairs selected across the category boundary could differ to a much greater degree than those selected from within. This had the effect of overrepresenting items nearer to the boundary in the between-category set, as compared with a truly random selection from all between-category pairs, but with the result that the degree of physical difference between pairs was equivalent for between- and within-category pairs. Order of presentation of the pairs for judging was random. The first member of the pair was presented for 2 s, followed by a blank screen for 1 s, and then the second member of the pair for 2 s. Similarity was rated on a 9-point scale from 1 (*most similar*) to 9 (*least similar*), using the computer keyboard to enter responses. A bar with the numbers 1 to 9 appeared on the screen representing the similarity scale and remained on until the participant made a response (the words *most similar* anchored the low end of the scale, and the words *least similar* anchored the high end of the scale).

### Results

All but one of the participants in the learning condition met criterion; the participant who failed to do so missed a single item in each of the last two blocks of 32 trials. This was judged to constitute sufficient evidence of having

learned the category distinction, so the participant's data were included in the similarity analysis. Mean number of items correct out of the first 16 items for all learning participants was 9.62, which is not significantly different from chance. The distinction between categories was therefore nontrivial and did require learning. The learning curve presented in Figure 4 is averaged across all participants. Note that this curve is representative of the learning curves for individual participants, which was gradual in every instance; it does not mask individual all-or-none learning effects.

A graph of mean similarity ratings by group (baseline vs. learning) is shown in Figure 5 (recall that higher numbers reflect lower similarity ratings). A 2 (group: baseline vs. learning)  $\times$  3 (pair type: Gex-Gex, Gex-Zof, Zof-Zof) analysis of variance (ANOVA) with repeated measures on the second variable revealed a marginal effect of condition,  $F(1, 14) = 4.217$ ,  $MSE = 2.183$ ,  $p < .06$ ; a significant effect of pair type,  $F(2, 28) = 3.384$ ,  $MSE = 0.652$ ,  $p < .05$ ; and a significant interaction effect,  $F(2, 28) = 7.889$ ,  $MSE = 0.652$ ,  $p < .002$ .

The compression-expansion model suggests, a priori, the need for tests of differences between three pairs of those means involved in the interaction effect. The critical compari-

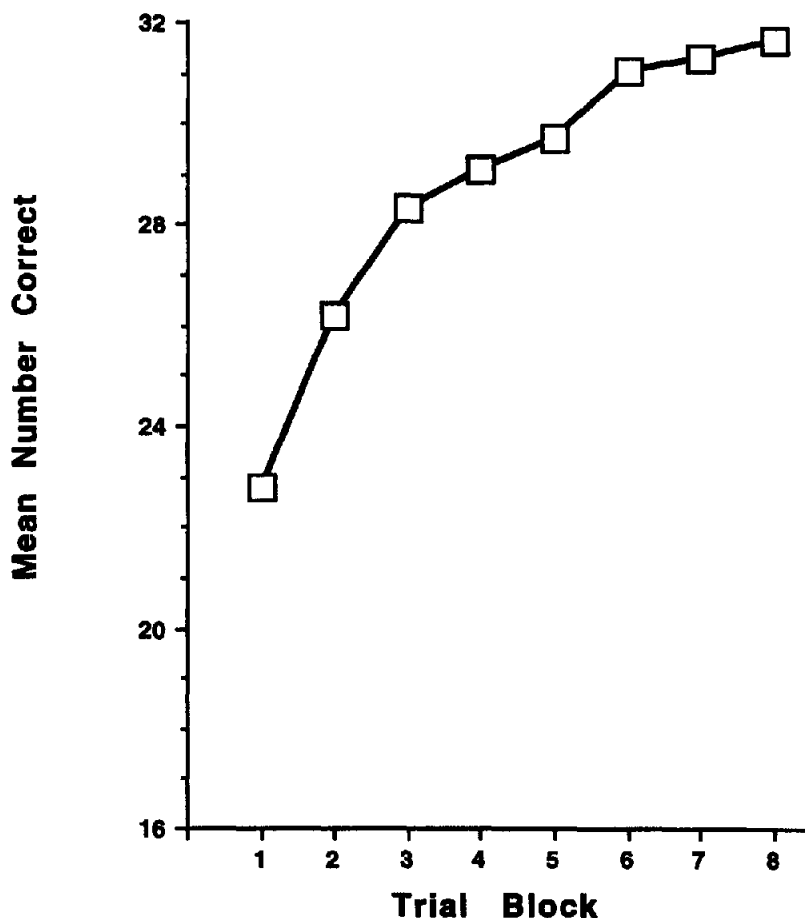


Figure 4. The learning curve for people in the learning condition in Experiment 2.

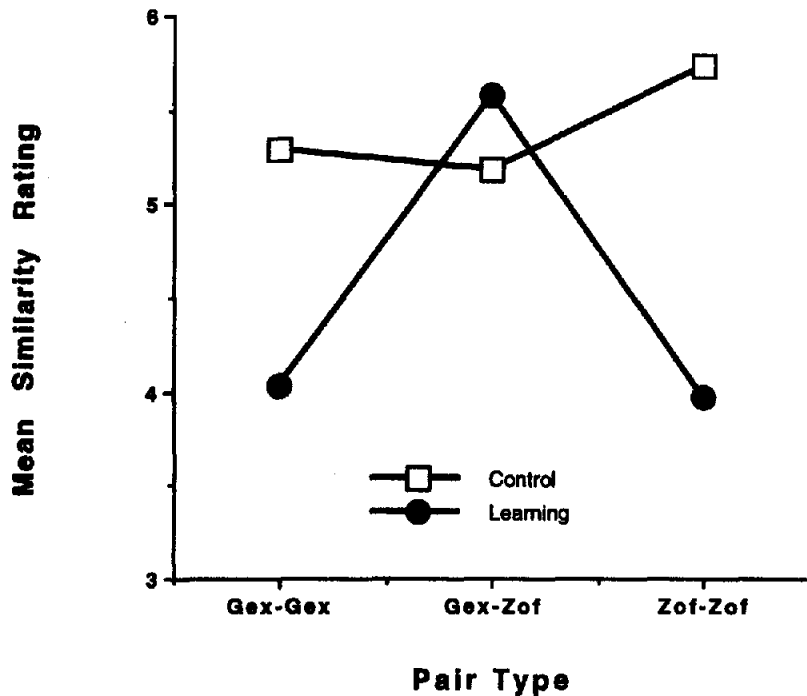


Figure 5. Plot of similarity ratings for both control ( $n = 8$ ) and learning ( $n = 8$ ) groups in Experiment 2. Post hoc analyses revealed compression effects without expansion at the category boundary.

sons are those that establish whether learning has an effect on judged similarity of stimulus pairs. We therefore performed  $t$  tests of the differences between the control and learning groups for three pair types: the two within-category comparisons (Gex-Gex [G-G], Zof-Zof [Z-Z]), and the between-category comparison (Gex-Zof [G-Z]).

Planned  $t$  tests on the differences between these means are presented in Table 1. The criterion for significance was Bonferroni adjusted to correct for multiple comparisons. Comparisons between the two groups of participants showed that the mean similarity rating for the G-G pairs in the learning condition differed from the corresponding mean rating in the baseline condition. The Z-Z ratings also reflected significantly reduced similarity ratings among participants in the learning group as compared with the baseline group. In contrast, the mean similarity ratings for the G-Z pairs did not differ for the two groups. Note that a post hoc contrast of within-category versus between-

category pairs for the baseline group showed no effect of pair type ( $F < 1$ ).

### Discussion

The significant interaction of Group  $\times$  Pair Type, combined with the finding of no significant difference in mean similarity ratings for different pair types in the control groups, is consistent with a warping of the psychological similarity space as a result of category learning. Planned  $t$  tests revealed that the effect was limited to a compression of similarity space within categories without increased expansion at the boundary between categories. This is in contrast to a Gibsonian theory of perceptual learning (Gibson, 1969), which predicts that one should see an increase in distinctiveness as a result of category learning. However, acquired distinctiveness and acquired similarity are obviously opposite sides of the same coin when one views the psychological similarity space nonlocally. If the members of a category acquire increased similarity (the compression effect), even if they are not in the process separated from nearby members of other categories, the resulting compression makes the categories more distinctive qua categories. When Gexes (for example) can be treated pretty much alike, and so can Zofs, then it is easier to distinguish Gexes as a category from Zofs.

Finding compression without expansion is also interesting because it is not consistent with the interpretation of these data as the result of a simple demand effect. The worry here is that category learners are merely trying to respond to the

Table 1  
Planned  $t$  Tests of Learning Versus Baseline Means  
for Gex-Zof Nonoverlap Study

Type of pair tested	Difference	$t(14)$	$p$
Gex-Gex	-1.265	-3.134	.0040*
Zof-Zof	-1.764	-4.370	.0002***
Gex-Zof	0.401	0.994	.3287

\*Significant at Bonferroni adjusted equivalent of  $p < .05$ . \*\*\*Significant at Bonferroni adjusted equivalent of  $p < .001$ .

experimenter's expectation that items that have been given the same label ought to be judged as more similar, whereas those with different labels should be judged as more dissimilar. The fact that the similarity of within-category pairs was greater for category learners than for nonlearners, whereas the between-category pairs were not judged to be less similar, casts some doubt on this interpretation. If people were complying with this demand, then one would have expected changes in both directions relative to the nonlearner baseline. Further evidence against the view that similarity judgments are driven by whether the same label is attached to a group of objects comes from cross-linguistic research by Sloman, Malt, and Shi (1997). They reported evidence that despite the fact that English and Chinese speakers have different linguistic category boundaries for the domain "containers," their perceptual similarity judgments of those objects are the same. They concluded that having labels for categories of objects does not in and of itself account for the shape of psychological similarity space, although it should be noted that the labels were not highlighted in the lab prior to giving similarity judgments. Nevertheless, given the absence of an effect of labeling per se, and the selective shift in the similarity space in our own data (compression but not expansion), we conclude that it is unlikely that demand effects account for the learned categorical perception effects we observed.

It is possible, however, that the compression without expansion in this study arose from the way we arranged the categories and category boundary. In the following experiment, we explored the effects of one possible change in this configuration.

### Experiment 3

One might hypothesize that a pattern of compression without expansion reflects the already clear separation between Gexes and Zofs. After all, they share no overlapping values on either dimension, and the dimensions were scaled so as to be well above JND levels. Under these conditions, learning to ignore or to minimize within-category differences is sufficient to partition the similarity space for efficient categorization. It is a reasonable hypothesis, however, that when the values on the category-relevant dimensions are not as distinctive, that is, when the boundary is particularly "noisy," a mechanism for enhancing separation may be more readily engaged. To explore this hypothesis, we created new versions of the Gex-Zof categories in which there were shared or overlapping feature values on both of the category-relevant dimensions.

### Method

**Participants.** Participants were 20 Vassar College undergraduates who participated in partial fulfillment of an introductory psychology course requirement.

**Stimuli.** Stimuli were constructed according to the same basic plan used to build the stimuli for the previous study, except that the values on the two varying features could overlap. Thus, Gexes had the five shortest cilia (out of eight), instead of the four in the previous study, and the five head shapes tending toward indentation

on the lower portion of the figure. Zofs had the five longest cilia and five heads tending toward indentation on the upper portion of the figure. Thus, the two middle values on both dimensions could occur in both Gexes and Zofs. Two of the four fully ambiguous stimuli were removed from the set to simplify the categorization rule, resulting in a total of 22 stimuli in each category. Despite the overlap in feature value range, all stimuli were unambiguously assigned to one of the two categories.

**Procedure.** Participants were again assigned randomly to either a baseline or a learning group. The learning procedure was the same as that used in the previous study, except that a block of trials consisted of presenting all 44 stimuli from the two categories in a random sequence. Training was continued for 352 trials unless the participant managed to respond correctly to all 44 stimuli in a trial block, in which case training was halted. The participant was allowed a break after Trials 132 and 264. The similarity judgment procedure was also the same as that used in the previous study. Again, only a subset of all possible pairs was rated, this time including 38 pairs of each of the comparison types (Gex-Gex, Gex-Zof, and Zof-Zof). The subset was selected according to criteria like those used in Experiment 2.

### Results

The boundary conditions in this study made the task of learning the category distinction much more difficult. Only 1 of the 10 participants in the learning condition actually managed a completely errorless trial block (on Block 5). All other participants continued training for the full 352 trials. The mean number correct out of the first 22 trials was 12.2, which is not significantly different from chance. An averaged learning curve is presented in Figure 6. With the exception of Participant 10, all participants gave clear evidence of having learned the distinction, despite the failure to achieve a perfect response record. Excluding Participant 10, the mean percentage correct for the last trial block was 89.4%. Participant 10 was still responding at chance (exactly 50%) on the last trial block, and this participant's data were excluded from the analyses that follow on the grounds that failure to acquire the category makes it impossible to test for predicted effects of category learning.

A graph of the mean similarity ratings by both baseline and learning groups is shown in Figure 7. A 2 (group: baseline vs. learning)  $\times$  3 (pair type: G-G, G-Z, Z-Z) ANOVA with repeated measures on the second variable revealed a marginally significant effect of condition,  $F(1, 17) = 3.829$ ,  $MSE = 1.225$ ,  $p < .10$ , and a significant interaction effect,  $F(2, 34) = 4.598$ ,  $MSE = 0.641$ ,  $p < .02$ . The same pattern of planned  $t$  tests explored in the previous study was also examined for these data; the results are shown in Table 2. The criterion for statistical significance was again Bonferroni adjusted. The comparisons between groups with respect to the same pair types revealed that the ratings of G-Z pairs did not differ for the two groups, suggesting no between-category expansion effect. The G-G pairs were judged to be significantly more similar by learning than by baseline participants, but the Z-Z difference only approached significance. As in Experiment 2, a post hoc contrast of within-category versus between-

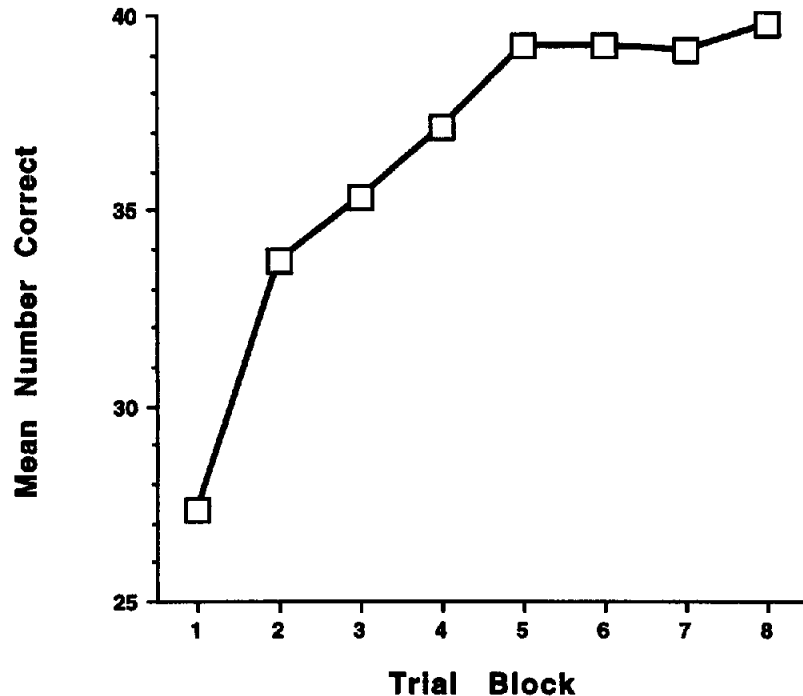


Figure 6. The learning curve for people in the learning condition in Experiment 3.

category pairs for the baseline group shows no effect of pair type,  $F(1, 17) = 2.231, p > .1$ .

### Discussion

As in the previous study, the significant interaction effect reflects the hypothesized change in the structure of the similarity space following category learning, but again the change involved within-category compression effects only, despite the fact that the boundary was much less well defined. Although the overall effect of compression was significant, the effect was weaker than in the previous experiment, as revealed by planned tests showing that the compression effect reached significance for only one of the categories. The failure to observe between-category expansion does not appear to be attributable to nonoverlap of dimensional values at the boundary. Even when the category boundary was hard to discriminate, compression of a subset of exemplars into a smaller region of the psychological similarity space was sufficient to establish a new, more tractable psychological unit (e.g., the Gex chunk), even while maintaining the same level of distinctiveness between items across the new category boundary. The effect is to increase relative distinctiveness by producing absolute compression without affecting distinctiveness across the boundary.

Thus far, the category distinctions we have manipulated depended on psychologically equal intervals along continuous dimensions. This raises questions about whether the learned categorical perception effect as thus far observed is peculiar to artificial stimulus domains. There are obvious

advantages to using artificial stimuli in studies of this kind. First, it is difficult to study concept formation in adults without using artificial categories because so much of what is knowable about the world is already organized into categorical form by adults. Second, artificial categories give the experimenter complete control of stimulus features and of category structure. Despite these advantages, it is important to investigate whether the findings from research with artificial stimuli generalize to real objects, if the theory is to be extended as an account of category learning outside the laboratory. It is not sufficient to assume that a phenomenon observed when stimuli are relatively simple and configured into artificial categories will apply when the stimuli are more richly multidimensional and have been classified by people other than the experimenters. One could address these concerns by building more elaborate stimuli in the laboratory and having study participants impose classifications, but in the next study we took advantage of an unfamiliar but natural category distinction to address these concerns.

### Experiment 4

There are some naturally occurring stimulus domains that are not familiar to adults (the Gex-Zof categories trade on the unfamiliarity to most adults of single-celled organisms), and one of these has been the subject of considerable speculation by cognitive scientists for some time. In fact, an entire mythology has grown up around the arcane process of learning to distinguish male from female day-old chicks by professional chicken sexers (e.g., Gibson, 1969; Lawrence, 1950). Contrary to legend, the distinction can be learned,

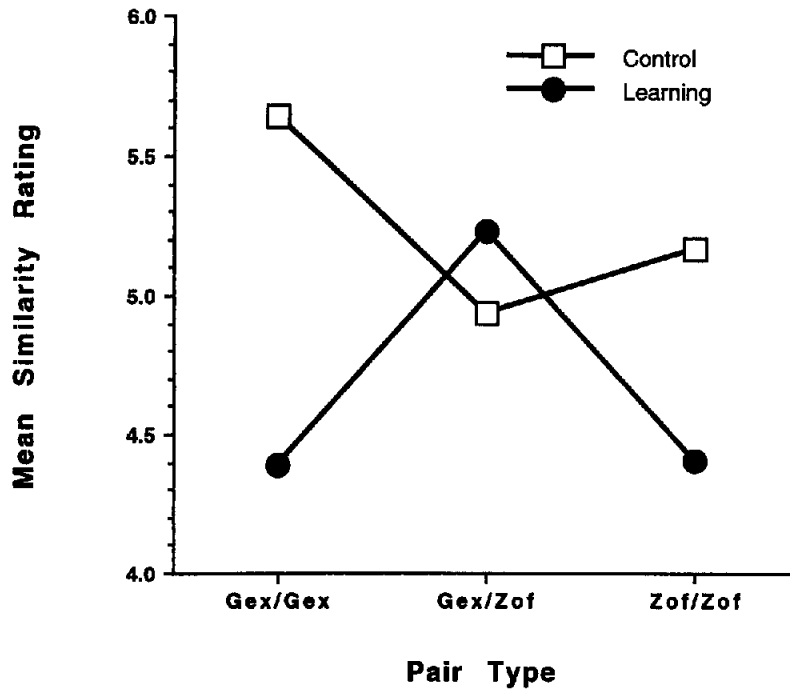


Figure 7. Plot of similarity ratings for both control ( $n = 10$ ) and learning ( $n = 9$ ) groups in Experiment 3. As in Experiment 2, post hoc analyses revealed compression effects without expansion at the category boundary.

and learned relatively quickly, given information about which features to look for (Biederman & Shiffrar, 1987).

In this experiment, we set out to study whether the learned categorical perception effect occurs when people learn to categorize drawings of chick cloaca. To minimize potential influences of any background knowledge about the morphology of sex differences (see Livingston & Andrews, 1995; Wisniewski & Medin, 1994, for findings related to the interaction of prior knowledge and stimulus structure), we told our participants that they would be learning to distinguish pictures of the larynxes of two different species of monkey. We thus had an opportunity to study whether the learned categorical perception effect occurs when participants encounter a stimulus set (a) that is more complex and multidimensional than those used in previous studies, and (b) in which the category membership of an instance is not an arbitrary stipulation by the experimenter but a fact of nature.

Table 2  
Planned *t* Tests of Learning Versus Baseline Means  
for Gex–Zof Overlap Study

Type of pair tested	Difference	<i>t</i> (17)	<i>p</i>
Gex–Gex	-1.252	-3.404	.0017**
Zof–Zof	-0.763	-2.074	.0457
Gex–Zof	0.291	0.792	.4341

\*\*Significant at Bonferroni adjusted equivalent of  $p < .01$ .

### Method

**Participants.** Participants were 20 volunteers recruited by public advertisements. They ranged in age from 18 to 44 years, and all participants were paid for their participation.

**Stimuli.** Stimuli were 14 black-and-white drawings, 2.4 cm wide  $\times$  1.9 cm high, of the genitalia of day-old chicks, seven of them male and the other seven female, taken from Biederman and Shiffrar (1987). Figure 8 shows a subset of these stimuli. Selection of the stimuli from the larger set was random, subject to the constraint that there be equal numbers of stimuli drawn from the two categories. Stimuli were cropped from high-quality copies and pasted onto 10 cm  $\times$  15 cm blank index cards and then placed into two-ring binders. One stimulus was placed in the center of each card for the learning trials. Six copies were made of each stimulus and six different random sequences of the 14 stimuli were generated. Separate binders were prepared for the similarity-judgment task, with the stimuli for a given pair placed side by side in the center of the same card and approximately 2.7 cm apart. All possible pairs of the 14 stimuli were included, and two sets were made, reversing the left–right position of the members of each pair between sets.

**Procedure.** Participants were assigned to either a learning condition or a baseline (control) condition. Participants in the learning condition were told that they would be viewing pictures of the larynxes of two different species of monkeys and that they would be asked to learn to tell which was which. No prior information was given about the nature of the differences. Each participant viewed the pictures in one of two sequences (each was simply the reverse of the other). The participant was asked to label each picture as either Type A or Type B and was given immediate feedback by the experimenter, who recorded the response and then

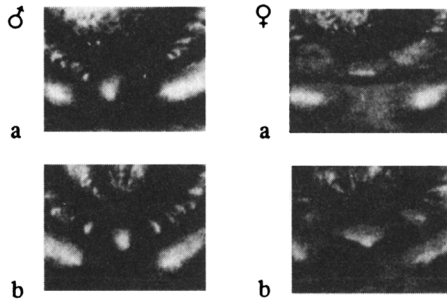


Figure 8. Examples of the pictures of male and female chicken cloaca used as stimuli in Experiment 4. From "Sex Determination of Day-Old Chicks. II. Type Variations," by T. H. Canfield, 1941, *Poultry Science*, 20, p. 328. Copyright 1941 by The Poultry Science Association. Reprinted with permission.

turned to the next picture in the sequence. Training continued until the participant met the criterion of two error-free trial blocks (28 stimuli), or until a total of 252 presentations (18 trial blocks), whichever came first.

Following training, the participant was asked to judge the degree of similarity between stimuli presented in pairs. A series of practice judgments was made to familiarize the participants with the scale. Judgments were then made using the same 9-point scale used in previous studies. Participants in the baseline condition judged the similarity of the same pairs without prior training.

### Results

The mean proportion of correct responses in the first trial block was .529, a proportion not significantly different from

chance. Figure 9 shows the mean learning curve for all learning participants. The mean proportion of correct responses in the last trial block was .936, so there is clear evidence that the learning participants mastered the distinction. However, 4 of the 10 learning participants failed to reach the criterion for termination of training. The mean proportion of correct responses in the last block of 18 trials for these participants did reach .84, which is greater than chance and suggests that they had succeeded in learning something about the distinction. A 2 (criterion reached vs. not reached)  $\times$  3 (pair type: Type A–Type A, Type A–Type B, and Type B–Type B) ANOVA with repeated measures on the second variable confirmed that there was no difference in the similarity ratings of those participants who did reach criterion and those who did not,  $F(1, 18) = 1.402$ ,  $MSE = 2.784$ ,  $p > .2$ , and no interaction with pair type,  $F(2, 36) < 1$ ,  $MSE = 1.068$ . The data from all learning participants were therefore entered into the subsequent comparison of data from the learning and baseline conditions.

Figure 10 shows the mean similarity ratings by the baseline and learning groups for the three pair types (A–A, A–B, and B–B). A 2 (group)  $\times$  3 (pair type) ANOVA on similarity judgments with repeated measures on the second variable revealed significant main effects of group,  $F(1, 18) = 4.551$ ,  $MSE = 1.652$ ,  $p < .05$ , and pair type,  $F(2, 36) = 10.290$ ,  $MSE = 0.510$ ,  $p < .001$ . The hypothesized interaction of group with pair type was marginally significant,  $F(2, 36) = 3.183$ ,  $MSE = 0.510$ ,  $p = .053$ .

Table 3 contains the same planned  $t$  tests of the differences between means examined in the previous two studies, and the pattern is a familiar one. The learning and baseline

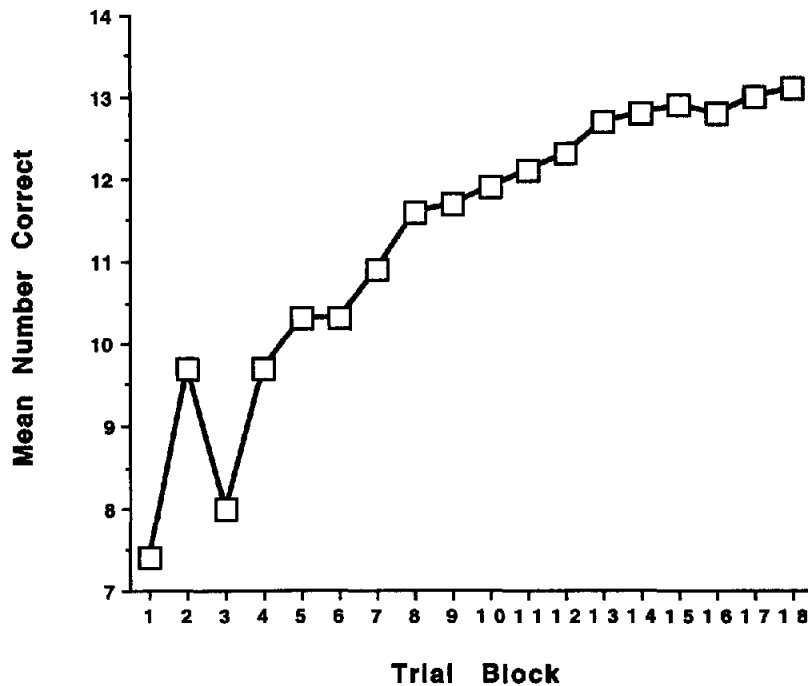


Figure 9. Learning curve for people in the learning condition of Experiment 4.

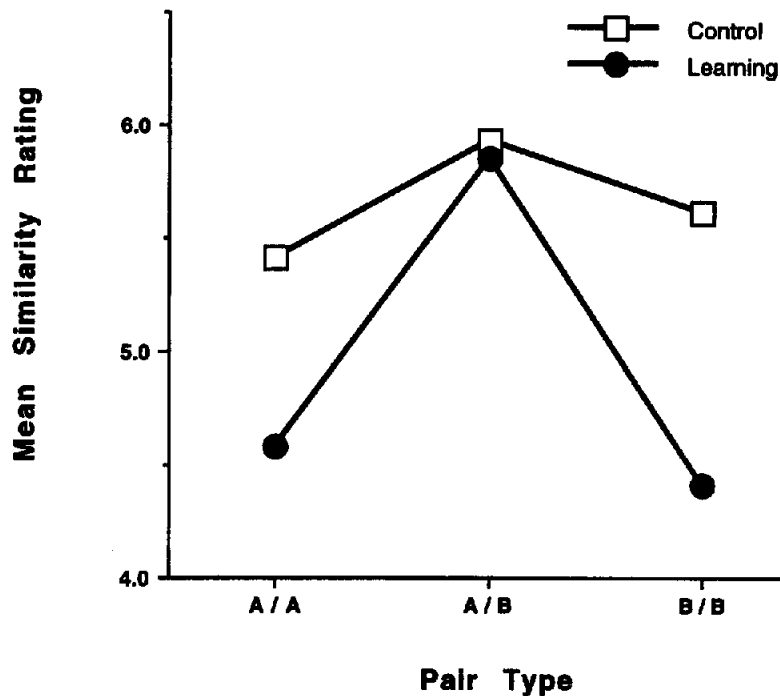


Figure 10. Plot of similarity ratings for both control ( $n = 10$ ) and learning ( $n = 10$ ) groups in Experiment 4. As in Experiments 2 and 3, post hoc analyses revealed compression effects without expansion at the category boundary.

groups did not differ in their judgments of the between-category pairs, but the learning participants judged the within-category pairs to be more similar than did the baseline participants.

Because we had similarity ratings from the complete set of possible pairs of category instances for this study, it was possible to use an MDS analysis to explore the changes in relationships among items that result from category learning. However, we did not initially force the data from the baseline and learning groups into a common solution for the simple reason that we do not know whether the dimensional structure of the stimuli in psychological similarity space remains the same following category training. Indeed, the available evidence suggests that category learning can change the relative salience of the dimensions along which a set of objects differs, even to the point of drawing attention to previously unattended features (Livingston & Andrews, 1995; Wisniewski & Medin, 1994). In recognition of this

possibility, the baseline and learning data were submitted to separate ALSCAL (Euclidean distance scaling) models (see Homa, Rhoads, & Chambliss, 1979, for similar arguments). For both models, a three-dimensional solution provides the best fit to the data, with stress = .128 and  $R^2 = .818$  for the baseline model, and stress = .129,  $R^2 = .850$  for the learning model.

In an effort to identify the dimensions extracted by the MDS analysis, we quantified seven dimensions along which the stimuli varied. These seven dimensions included all of the single, nonrelational features that the experimenters were able to identify as both separable and quantifiable. The first two authors (Kenneth R. Livingston and Janet K. Andrews) made all measurements or ratings independently, and these were correlated to determine reliability. The shape of the major central structure was quantified by rating the degree of its concavity-convexity. This was the stimulus property of this structure identified by Biederman and Shiffrar (1987) as diagnostic of the category difference for their chicken sexes. A value of +10 was assigned to maximally convex (essentially spherical) structures, and a value of -10 was assigned to maximally concave structures. The largest negative value assigned by either of the raters was -5, because none of the structures was maximally concave. Interrater reliability of ratings for this dimension was  $r = .802$ .

The second dimension rated was the number of different lobes comprising the central structure. These ranged from one to three in number, and reliability for this coding was

Table 3  
Planned *t* Tests of Learning Versus Baseline Means  
for Chicken Study

Type of pair tested	Difference	$t(18)$	$p$
Male-male	-0.833	-2.608	.0132*
Female-female	-1.205	-3.771	.0006**
Male-female	-0.086	<1	

\*Significant at Bonferroni adjusted equivalent of  $p < .05$ . \*\*Significant at Bonferroni adjusted equivalent of  $p < .01$ .

Table 4  
Forward Stepwise Regression of Chicken Stimulus Values  
on the Dimensions Extracted by the ALSCAL MDS Analysis

MDS dimension and stimulus variable	$\Delta R^2$	$\beta$	$t(11)$
Baseline Dimension 2 <sup>a</sup>			
Step 1: Beads	.425	0.773	4.988***
Step 2: Size	.322	0.581	3.747**
Learning Dimension 1 <sup>b</sup>			
Step 1: Shape	.696	1.011	6.449****
Step 2: Lobes	.099	0.361	2.304*
Learning Dimension 3 <sup>c</sup>			
Step 1: Height	.401	-0.833	-4.408**
Step 2: Overlap	.208	0.498	2.419*

Note.  $\beta$  is the standardized regression coefficient for each variable in the final equation that incorporates both variables. ALSCAL = standard Euclidean distance scaling algorithm; MDS = multidimensional scaling.

<sup>a</sup>Final  $R^2 = .747$ ,  $F(2, 11) = 16.258$ ,  $p < .001$ . <sup>b</sup>Final  $R^2 = .795$ ,  $F(2, 11) = 21.276$ ,  $p < .001$ . <sup>c</sup>Final  $R^2 = .609$ ,  $F(2, 11) = 8.571$ ,  $p < .01$ .

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ . \*\*\*\* $p < .0001$ .

1.00. The third dimension, apparent size of the central structure, was calculated by measuring the area of the two-dimensional projection of that portion of the image. Reliability was .942 for this measure.

Below or to either side of the central structure, one finds a bilaterally symmetrical pair of roughly tubular folds that vary in how close together they are. We measured the minimal distance between each pair, with a reliability of .983. In addition, the central structures of the stimuli vary in their height above these protruding folds, as measured from the center of the main structure to the point of closest approach of the protuberances. This dimension was measured with reliability of .819. The positioning of the central structure with respect to the protuberance resulted in varying degrees of overlap of these two features. The degree of overlap was measured as the distance from the furthest outside edge of the central structure to the point of closest approach of the protuberance. This measure was taken on both sides and averaged; reliability for this measure was .947. Finally, we counted the number of beadlike structures to either side of the central structure. The reliability of this count was .925.

Because reliabilities were acceptably high in all cases, we took the average of the values assigned by the two raters to each stimulus for entry into all subsequent analyses. Our goal in these analyses was to discover the identities of the dimensions extracted by the MDS analysis. To that end, we regressed each of the six dimensions identified by the MDS analysis (three baseline and three learning) onto the seven measures described, using a forward stepwise procedure.<sup>2</sup> A summary of these analyses is presented in Table 4. For three of the analyses (Baseline Dimensions 1 and 3, and Learning Dimension 2), none of the seven variables was significantly related to the MDS dimension; hence, we cannot infer the nature of these dimensions, although it should be noted that shape almost reached significance for Baseline Dimension 1

( $r = .480$ ,  $p < .09$ ). For Baseline Dimension 2, the number of beads and the size of the central structure together accounted for approximately 75% of the variance. For Learning Dimension 1, shape emerged as the predominant feature, accounting for approximately 70% of the variance, with number of lobes, which is also a feature of the central structure, accounting for an additional 10%. This result is consistent with Biederman and Shiffrar's (1987) account of shape as the most important factor for the chicken sexers in their study. Height and degree of overlap together accounted for 61% of the variance in Learning Dimension 3, suggesting that attention was paid to the relationship of the central structure to other structures.

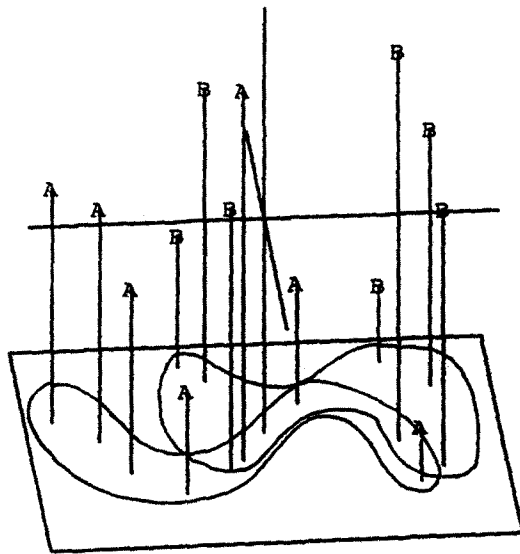
Thus, in so far as we are able to identify what the MDS dimensions reflect about the stimuli, they appear to be quite different for the baseline and learning conditions. Figure 11 shows separate plots for the baseline (control group) and learning solutions. Careful inspection does show that items become less dispersed and move toward linear separability in that a clear boundary between categories is easier to identify in the learning as compared with the baseline plots.

As a further check on these findings, we also performed an INDSCAL (individual differences or weighted Euclidean distance model) analysis on the data, forcing the two groups into the same solution space. This resulted in some loss of precision, as indicated by the lower  $R^2$  (.79 vs. .85 and .82 for the learning and control ALSCAL solutions, respectively), and it also indirectly suggests that the dimensional structure of the space was different for the two groups, although not necessarily in the same way suggested by the ALSCAL solutions. The INDSCAL solution confirmed that the most important dimension of similarity-dissimilarity was shape (see Table 5). Furthermore, the subject weights (which indicate the degree to which a particular dimension is important in explaining the distribution of similarities for each group) for the two groups differed in the expected direction. In particular, the shape dimension had a weight of .5888 for the control group, but a weight of .6927 for the learning group. The groups did not appear to differ on Dimension 2 (.4333 for the control group, .4584 for the learning group), but the control group weighted Dimension 3 more heavily (.4996) than the learning group (.3344).

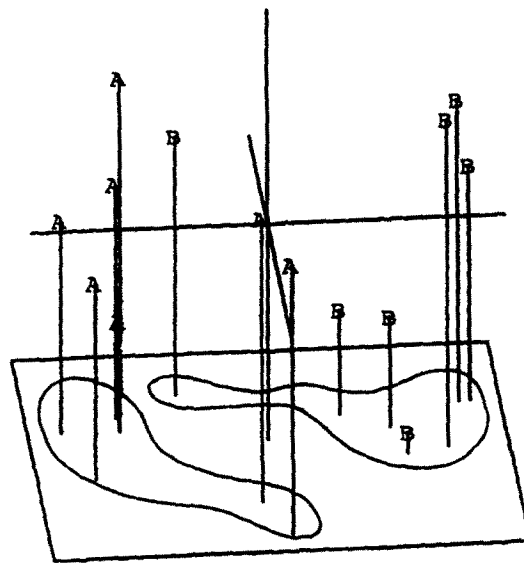
One remaining question that can be answered by these data is whether the compression effect was equivalent for all within-category pairs, or whether it was instead greater for some pairs than others. For each within-category pair, the mean similarity rating for the learning group was subtracted from the mean rating for the control group. The resulting differences were roughly trichotomized (small, medium, and large differences) and submitted to a one-way ANOVA, which was significant,  $F(2, 39) = 60.898$ ,  $MSE = 0.132$ ,

<sup>2</sup> We also explored the relationship of all possible ratios between pairs of variables (excluding shape) to the MDS dimensions and none emerged as superior to relationships with individual variables. For example, the ratio of number of beads to size does not better predict values on Baseline Dimension 2 than either variable taken alone. We therefore report only those analyses involving single variables.





CONTROL GROUP



LEARNING GROUP

Figure 11. Multidimensional scaling (ALSCAL [standard Euclidean distance scaling]) plots for similarity ratings of chicken cloaca stimuli for both the control and learning groups of Experiment 4. The scale is identical for the two plots. The two categories are represented by the letters A (the male stimuli) and B (the female stimuli). Curved lines drawn in the XY plane are intended to make identifying category groupings easier.

Table 5  
*Forward Stepwise Regression of Chicken Stimulus  
 Values on the Dimensions Extracted by the  
 INDSCAL MDS Analysis*

MDS dimension and stimulus variable	$\Delta R^2$	$\beta$	$t(11)$
Dimension 1 <sup>a</sup>			
Step 1: Shape	.487	0.941	4.764***
Step 2: Lobes	.187	0.496	2.510*
Dimension 2 <sup>b</sup>			
Step 1: Overlap	.365	-0.604	-2.629*
Dimension 3 <sup>c</sup>			
Step 1: Height	.344	-1.086	-4.919***
Step 2: Distance	.345	-0.770	-3.490**

Note.  $\beta$  is the standardized regression coefficient for each variable in the final equation that incorporates both variables. INDSCAL = individual differences or weighted Euclidean distance scaling; MDS = multidimensional scaling.

<sup>a</sup>Final  $R^2 = .674$ ,  $F(2, 11) = 11.366$ ,  $p < .01$ . <sup>b</sup>Final  $R^2 = .365$ ,  $F(2, 11) = 6.910$ ,  $p < .05$ . <sup>c</sup>Final  $R^2 = .689$ ,  $F(2, 11) = 12.177$ ,  $p < .01$ .

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

$p < .0001$ . Thus, not all pairs increase in similarity to the same degree following learning.

### Discussion

The finding that category learning results in compression of the psychological similarity space without expansion confirms and extends the findings from our studies using artificial stimuli and categories. The use of MDS analysis was new to this study, and although it is not easy to compare the solutions for the baseline and learning conditions, the analysis does help to visualize the kinds of changes in psychological similarity space that occur with multidimensional category learning. The results of the regression analyses reported above show that some dimensional changes did in fact occur in the course of category learning, as expected. Because the dimensions are different for the two solutions, the resulting MDS plots are of heuristic value only, but they do show concretely how it is that increased local compression could occur without an increase in global separation between categories. The plots show how increased similarity is by its nature accompanied by increased distinctiveness.

Complementary INDSCAL solutions confirm the shift in dimensional structure, although by forcing the two groups into the same dimensional space it may actually underestimate the extent of the shift. For example, the ALSICAL analysis makes clear that the control group appeared to give considerable attention to number of beads as a dimension, but this did not emerge as an important dimension when the groups were forced into the same similarity space. Additional work will be required to develop more suitable ways of representing the multidimensional changes that occur with learning.

Finally, the fact that there were significant differences in the amount by which pairs increased in similarity between the control and learning groups is inconsistent with a

demand-effect interpretation of the learned categorical perception effect.

### General Discussion

A good first test of whether an experimental psychological effect is a nontrivial one is whether it would surprise our grandmothers. By this standard, an experiment that demonstrated that objects in the same category resemble one another more than they resemble objects in other categories would fail the grandmother test, for Grandma would comment, "Of course they look more alike, that's why we put them in the same category in the first place!" Grandma would be much more surprised if we told her that we had evidence that things strike us as being more similar because we have learned to put them in the same category, not the other way around.

In the introduction, we discussed cases in which the structure of a domain is perceived as categorical a priori, the effect that has come to be called categorical perception, with things being seen as more similar or different than one would have predicted on the basis of a purely physical measure of their properties (Hamad, 1987). Perhaps the most prominent example of a priori categorical perception is the rainbow. If changes in appearances were directly proportional to changes in the wavelength of light, then we should see a smooth continuum of variation, like shades of gray, or of any other hue; we should not see colors themselves, as the qualitatively distinct bands they appear to be.

However, the physics of light alone is insufficient to tell us what colors ought to look like or even that the continuum of variation in wavelength ought to look like a continuum. In fact, there is nothing in the physics of light that dictates that the small portion of the spectrum from the invisible ultraviolet to the likewise invisible infrared range should look like anything at all. But it is in the nature of human perceptual systems to respond categorically to the visible spectrum of light, and colors look categorical from birth. Color names do follow from their appearances, not the other way around, just as Grandma would expect.

There is a similar lesson from the other prominent example of categorical perception that we reviewed: phoneme perception. As in the case of color, humans are born biologically prepared with "frets" along several speech-sound continua, sorting them into qualitatively distinct categories such as ba, da, and ga. Grandma's intuition—that people put things into the same categories when they appear similar—is again supported, and again the boundaries of these categories have been established by the "Blind Watchmaker" rather than by language, just as in the case of color.

The examples of color and phonemic boundaries are striking, but the great majority of the categorical distinctions that people make are not so fully prepared in advance by evolution. Instead people need to learn them. The aim of our research has accordingly been to identify a mechanism that could explain how feedback from the consequences of one's categorizations actually results in categorical responding. Our results suggest that learning to categorize a set of items

produces a change in the multidimensional similarity space for those items, and it is this change that constitutes the learned categorical perception effect. Note that, logically, the relevant change could consist in either (a) shifting of item locations within a stable dimensional structure or (b) a change in the dimensional structure itself. The evidence from the MDS analysis of our last experiment suggests that it is the latter, not the former, that occurs, although a detailed understanding of the nature of these changes will require further study. In particular, we do not have comparable evidence for the artificial stimuli of Experiments 2 and 3, where there were only two dimensions of variation, both of which were carefully scaled. However, these experiments also showed within-category compression without between-category expansion, just as in the chicken cloaca study.

The repeated pattern of compression effects without expansion effects in our studies occurred despite the differences across experiments. This is consistent with data from two other studies that can be interpreted as showing compression without expansion following category learning. First, Shin and Nosofsky (1992), who were testing the goodness of fit of exemplar versus prototype versus combined models of category structure to data on the classification of dot patterns, measured similarity ratings for pairs of patterns, and also found what look like compression effects following category learning. Their data show compression for both within- and between-category pairs, but because there was more compression for within-category pairs than for between category pairs, the overall effect was the same as if there had been only compression within categories. Second, Homa et al. (1979) used MDS analyses to explore prototypicality effects produced by learning to categorize random dot patterns. We interpret their results (see their Table 1, p. 18) as showing compression effects based on their MDS analysis of their data. However, they confronted the same dilemma we did in constructing their MDS analyses and also chose to submit the data from the before- and after-training conditions to separate analyses. Our analyses of the data from Experiment 4 suggest that it is questionable whether one can extract evidence for compression or expansion from MDS solutions alone when they are performed separately on pre- and posttraining data sets. Because there is no evidence of compression effects apart from the MDS data (unlike in our experiment, which provided independent evidence of compression from our similarity measures), the Homa et al. results should be interpreted with caution, because the dimensions themselves might be different for their learning and control groups.

These results are partially at odds with those reported by Goldstone (1994a). He also observed the learned categorical perception effect, but in his studies it occurred primarily in the form of expansion effects. At present, the reason for this discrepancy is unclear. Goldstone's dependent measure was same-different judgments, not similarity ratings, and it is possible that the dependent measure may be a crucial factor, although why or how remains to be determined. For example, the same-different task requires the presentation of many "same" pairs, but the similarity rating task does not; in fact, none of the studies discussed above (our own; Homa

et al., 1979; and Shin & Nosofsky, 1992) presented pairs of identical stimuli. Furthermore, the same-different task requires a binary response, whereas the rating scales require more graded responses (from 1 to 9 in our studies and in Shin & Nosofsky, and from 0 to 10 in Homa et al.). In addition to the difference in dependent measure, there are also differences in the stimuli themselves. For example, our stimuli (and, apparently, those of Shin & Nosofsky and Homa et al.) differ from one another along their various dimensions by amounts that seem to be well above the JND, whereas Goldstone's stimuli were intentionally scaled to make adjacent pairs less than a JND (50% accuracy criterion) apart. The only way to really improve performance at the boundary under these latter conditions is to enhance perceptual discriminability along the relevant dimension(s), an effect that may be difficult to limit to that portion of the dimension that falls at the boundary. This interpretation is consistent with the "acquired distinctiveness" that occurs even within categories (though not to the same extent as between) for many of Goldstone's learning groups. It is interesting that the only cases of compression in Goldstone's studies occurred for dimensions that were not relevant to categorization. Those were the conditions under which it makes sense to neglect the differences among stimuli (i.e., to compress them). In our studies, where the differences between stimuli were readily identified, the categorization task did not present the perceptual discrimination problem encountered in Goldstone's paradigm, and the task could be most efficiently performed by, in Goldstone's (1994a) words, "learning to neglect differences along a categorization-relevant dimension" (p. 197; see also Kelley, 1984, for a theoretical prediction of a similar effect). Further research will be required to establish whether these kinds of differences in stimuli and procedure reliably predict whether the categorical perception effect is one of compression, expansion, or both.

Behavioral experiments such as ours and Goldstone's (1994a) only show that categorical perception can indeed be induced by learning alone. They cannot reveal what functional role, if any, compression-expansion might be playing in perceptual learning and representation. One possibility is suggested by simple feedforward networks like those studied by Harnad et al. (1991, 1995) and discussed in the introduction to this article. In the net, the functional role of the categorical perception effect appears to be to move the stimulus representations into positions where a plane can separate them into categories (linear separability) and minimize the confusability of items at the boundaries. Our MDS analyses of the data from Experiment 4 show that the process of category learning (see Figure 11) resulted in within-category compression sufficient to produce clear linear separability even without between-category expansion.

Thus, both the human data and the network simulation data suggest that categorical perception effects may occur in the service of categorization. In fact, one possibility is that they are the locus of a system's capacity to sort correctly. If this were the case, then both nets and people manage to sort by manipulating interstimulus differences. At this point, this

remains a hypothesis only, but one that is at least consistent with the data presented here. Note that the hypothesis in its strong form could be disconfirmed by finding a case in which people (or nets) manage to categorize but do not show learned categorical perception effects. Thus far, there are no definitive counterexamples,<sup>3</sup> and the available evidence does make one thing clear: The idea that it is obvious how to categorize on the basis of how things look is simplistic, at least when they vary in more than one respect. Categorization is a much more active process than Grandma suspected.

Although these results go some way toward demonstrating the importance of learned categorical perception effects for category learning, they raise a great many new questions in the process. The following explicit discussion of some of the most important of these is intended to indicate directions for further research.

First, the reader will no doubt have noticed that Harnad et al.'s (1991, 1995) network simulations are capable of learning to categorize coded forms of one-dimensional stimuli that are difficult for human participants to categorize. As noted in our discussion of Experiment 1, if learned categorical perception actually alters the way things look, then it is likely to be resisted by the relative cognitive impenetrability of early perceptual processes. It should be pointed out, however, that the category training we gave our participants was relatively brief. Tasks that involve difficult perceptual discrimination are likely to require long-term training, with spaced rather than massed practice, before reliable categorization and warping of the similarity space can emerge.

Furthermore, the network simulations are not intended to capture the entire human process of category learning, only to model certain functional components of it. In this respect, network models are not as different from symbolic AI models as is often claimed (Aizawa, 1992). For example, the simulations obviously do not have their own transducers and early perceptual processes. At this level, as we have already suggested, it may be hard-wired sensory processes that account for the fact that people fail to show the categorical perception effect for one-dimensional stimuli. This hypothesis could itself be explored by building more elaborate network models that incorporate more of the features of early sensory-perceptual systems, a strategy that should help to refine the understanding of where the learned categorical perception effect occurs in the human case.

Second, there are questions about the relationship between the learned categorical perception effect and many other well-documented phenomena in the concepts literature. Changes in reaction time and in typicality ratings that occur during the course of category learning might both be explained by the pattern of changes in the relationships among items in psychological similarity space. For example, both compression and expansion effects should result in faster reaction times as category members become less confusable. These findings also raise interesting questions about the role of language in category learning. The idea that the lexicon and grammar of one's language determine or influence the structure of one's thought, including one's categories, is well known as the Whorf (1956) hypothesis; softened versions of the hypothesis have received some

empirical support (e.g., Lucy, 1992). The bulk of the evidence seems more negative (e.g., Au, 1983; Berlin & Kay, 1969; Heider, 1972; Sloman et al., 1997; Pullum, 1991), but the occurrence of learned categorical perception effects might be construed as support for a weak Whorf hypothesis. Unfortunately, the studies reported here do not compare cases in which category learning is mediated by words and those in which it is not. Our data do not, therefore, speak directly to the Whorf hypothesis, but questions about its relevance to category learning deserve further examination.

Third, effects of category learning and naming need to be investigated during infancy and childhood when perceptual, cognitive, and linguistic changes are particularly dramatic. We have begun pilot work in our labs to adapt our measures to studies with preschool children as a first step in exploring the pattern of learned categorical perception effects across ages. Similar questions could be asked about nonhuman species, where data on innate categorical perception exist (Ehret, 1987), and perceptual learning and categorization are being studied (Mackintosh, 1995; McLaren, Leavers, & Mackintosh, 1994), but learned categorical perception has yet to be investigated.

Finally, electrophysiological measures are beginning to be used to study speech categorical perception (Maiste, Wiens, Hunt, Scherg, & Picton, 1995; Sharma, Kraus, McGee, Carrell, & Nichol, 1993) and even perceptual learning (Skrandies & Fahle, 1994), but there are no data yet using these measures to study learned categorical perception. Current research in our laboratories is designed to investigate where and when in the processing of perceptual inputs the compression-expansion effects occur. Event related potentials (ERPs) measured before and after categorical perception-inducing category learning are compared with ERPs from people with equal exposure to the same stimuli without the feedback that produces learning. Analyses will attempt to identify features of the ERP that co-vary with the appearance of categorical perception effects. These data can then be used to guide later studies using more sophisticated mapping techniques to further refine the time-course and patterning of changes that occur during category learning.

In conclusion, we hope that these experiments point the way to a new theoretical and empirical strategy for understanding mechanisms of category learning. Instead of descriptive modeling of conceptual structure, this paradigm raises new kinds of questions about the relationships among category learning, the structure of psychological similarity space, and the nature of concepts. They should certainly give Grandma something new to ponder.

<sup>3</sup> It is possible that the pattern of results reported by Sloman et al. (1997) may reflect such a counterexample in that items that receive the same category label are not necessarily judged to be more similar than items not receiving the same label. However, because there is no prelearning baseline measure of similarities for these items, it cannot be known whether a process of compression does account for the observed categorization. Appropriate developmental studies would be required to establish the character of the prelearning similarity spaces for such natural categories.

## References

- Aizawa, K. (1992). Biology and sufficiency in connectionist theory. In J. Dinsmore (Ed.), *The symbolic and connectionist paradigms: Closing the gap* (pp. 69–88). Hillsdale, NJ: Erlbaum.
- Au, T. K.-F. (1983). Chinese and English counterfactuals: The Sapir-Whorf hypothesis revisited. *Cognition*, 15, 155–187.
- Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, 57, 217–239.
- Beale, J. M., & Keil, F. C. (1996). Categorical perception as an acquired phenomenon: What are the implications? In L. Smith & P. Hancock (Eds.), *Neural computation and psychology: workshops in computing series* (pp. 176–187). Berlin, Germany: Springer-Verlag.
- Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.
- Biederman, I., & Shiffrar, M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 640–645.
- Bornstein, M. H. (1987). Perceptual categories in vision and audition. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 535–565). New York: Cambridge University Press.
- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.
- Calder, A. J., Young, A. W., Perrett, D. I., Ectoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition*, 3, 81–117.
- Canfield, T. H. (1941). Sex determination of day-old chicks. II. Type variations. *Poultry Science*, 20, 327–328.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.
- Cross, D., Lane, H., & Sheppard, W. (1965). Identification and discrimination functions for a visual continuum and their relation to the motor theory of speech perception. *Journal of Experimental Psychology*, 70, 63–74.
- Ehret, G. (1987). Categorical perception of sound signals: Facts and hypotheses from animal studies. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 301–331). New York: Cambridge University Press.
- Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 4, 99–109.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 303–306.
- Eiser, J. R. (1996). Accentuation revisited. In W. P. Robinson (Ed.), *Social groups and identities: Developing the legacy of Henri Tajfel* (pp. 121–142). Oxford, England: Butterworth Heinemann.
- Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology*, 18, 500–549.
- Ectoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, 44, 227–240.
- Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. Englewood Cliffs, NJ: Prentice Hall.
- Goldstone, R. L. (1994a). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, 123, 178–200.
- Goldstone, R. L. (1994b). The role of similarity in categorization: Providing a groundwork. *Cognition*, 52, 125–157.
- Goldstone, R. L., Steyvers, M., & Larimer, K. (1996). Categorical perception of novel dimensions. *Proceedings of the 18th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- Harnad, S. (Ed.). (1987). *Categorical perception: The groundwork of cognition*. New York: Cambridge University Press.
- Harnad, S., Hanson, S. J., & Lubin, J. (1991). Categorical perception and the evolution of supervised learning in neural nets. In D. W. Powers & L. Reeker (Eds.), *Proceedings of the AAAI Spring Symposium on Machine Learning of Natural Language and Ontology* (pp. 65–74). Kaiserslautern, Germany: DFKI.
- Harnad, S., Hanson, S. J., & Lubin, J. (1995). Learned categorical perception in neural nets: Implications for symbol grounding. In V. Honavar & L. Uhr (Eds.), *Symbol processors and connectionist network models in artificial intelligence and cognitive modeling: Steps toward principled integration* (pp. 191–206). Boston: Academic Press.
- Heider, E. R. (1972). Universals in color naming and memory. *Cognitive Psychology*, 3, 337–354.
- Homa, D., Rhoads, D., & Chambliss, D. (1979). Evolution of conceptual structure. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 11–23.
- Hutchinson, J. W., & Lockhead, G. R. (1977). Similarity as distance: A structural principle for semantic memory. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 660–678.
- Keil, F. C. (1990). Constraints on constraints: Surveying the epigenetic landscape. *Cognitive Science*, 14, 135–168.
- Kelley, D. (1984). A theory of abstraction. *Cognition and Brain Theory*, 7, 329–357.
- Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin*, 112, 500–526.
- Kuhl, P. K. (1987). Categorization by animals and infants. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 355–386). New York: Cambridge University Press.
- Kurtz, K. J. (1996). Category-based similarity. *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Erlbaum.
- Lane, H. (1965). The motor theory of speech perception: A critical review. *Psychological Review*, 72, 275–309.
- Lawrence, D. H. (1950). Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *Journal of Experimental Psychology*, 40, 175–188.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–461.
- Livingston, K. R., & Andrews, J. K. (1995). On the interaction of prior knowledge and stimulus structure in category learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 48(A), 208–236.
- Logan, J. S., Lively, S. E., & Pisoni, D. P. (1991). Training Japanese listeners to identify /r/ and /l/. *Journal of the Acoustical Society of America*, 89, 874–886.
- Lucy, J. A. (1992). *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis*. Cambridge, England: Cambridge University Press.
- Mackintosh, N. J. (1995). Categorization by people and pigeons. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, 48(B), 193–214.
- Maiste, A. C., Wiens, A. S., Hunt, M. J., Scherg, M., & Picton, T. W. (1995). Event-related potentials and the categorical perception of speech sounds. *Ear and Hearing*, 16, 68–90.
- McGarty, C., & Penny, R. E. (1988). Categorization, accentuation, and social judgement. *British Journal of Social Psychology*, 27, 147–157.

- McLaren, I. P. L., Leevers, H. J., & Mackintosh, N. J. (1994). Recognition, categorization, and perceptual learning (or, how learning to classify things together helps one to tell them apart). In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 889–909). Cambridge, MA: MIT Press.
- Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist*, *44*, 1469–1482.
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, *100*, 254–279.
- Medin, D., & Schaffer, M. (1978). Context theory of classification. *Psychological Review*, *85*, 207–238.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.
- Murphy, G., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*, 289–316.
- 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. (1989). Further tests of an exemplar-similarity approach to relating identification and categorization. *Perception & Psychophysics*, *45*, 279–290.
- Pastore, R. E. (1987). Categorical perception: Some psychophysical models. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 29–52). New York: Cambridge University Press.
- Pazzani, M. J. (1991). Influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 416–433.
- Posner, M., & Keele, S. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*, 353–363.
- Pullum, G. K. (1991). *The great Eskimo vocabulary hoax, and other irreverent essays in the study of language*. Chicago: University of Chicago Press.
- Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.
- Raskin, L. A., Maital, S., & Bornstein, M. H. (1983). Perceptual categorization of color: A life-span study. *Psychological Research*, *45*, 135–145.
- Richardson, G. P. (1991). *Category names and category learning*. Unpublished doctoral dissertation, University of Stirling, Stirling, Scotland.
- Rosch, E., & Lloyd, B. B. (Eds.). (1978). *Cognition and categorization*. Hillsdale, NJ: Erlbaum.
- Sharma, A., Kraus, N., McGee, T., Carrell, T., & Nichol, T. (1993). Acoustic versus phonetic representation of speech as reflected by the mismatch negativity event-related potential. *Electroencephalography and Clinical Neurophysiology*, *88*, 64–71.
- Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General*, *121*, 278–304.
- Skrandies, W., & Fahle, M. (1994). Neurophysiological correlates of perceptual learning in the human brain. *Brain Topography*, *7*, 163–168.
- Sloman, S. A., Malt, B. C., & Shi, M. (1997). The relation of similarity to naming: Chinese versus American conceptions of bottles and jars. *Proceedings of the 19th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- Stevenage, S. V. (1998). Which twin are you? A demonstration of induced categorical perception of identical twin faces. *British Journal of Psychology*, *89*, 35–57.
- Tajfel, H. (1957). Value and the perceptual judgement of magnitude. *Psychological Review*, *64*, 192–204.
- Tajfel, H. (1959). Quantitative judgement in social perception. *British Journal of Psychology*, *50*, 16–29.
- Tajfel, H. (Ed.). (1978). *Differentiation between social groups: Studies in the social psychology of intergroup relations*. New York: Academic Press.
- Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgement. *British Journal of Psychology*, *54*, 101–114.
- Tversky, A. (1988). Features of similarity. In A. M. Collins & E. E. Smith (Eds.), *Readings in cognitive science: A perspective from psychology and artificial intelligence* (pp. 290–302). San Mateo, CA: Morgan Kaufmann.
- Whorf, B. L. (1956). *Language, thought, and reality*. Cambridge, MA: MIT Press.
- Williams, L. (1977). The perception of stop consonant voicing by Spanish–English bilinguals. *Perception & Psychophysics*, *21*, 289–297.
- Wilson, M. (1987). Brain mechanisms in categorical perception. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 387–417). New York: Cambridge University Press.
- Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science*, *18*, 221–290.

Received May 28, 1996

Revision received July 21, 1997

Accepted August 5, 1997 ■