Journal of Experimental Psychology: Learning, Memory, and Cognition

When Does Fading Enhance Perceptual Category Learning?

Harold Pashler and Michael C. Mozer

Online First Publication, February 18, 2013. doi: 10.1037/a0031679

Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/13/$12.00 DOI: 10.1037/a0031679

When Does Fading Enhance Perceptual Category Learning?

Harold Pashler
University of California, San Diego

Michael C. Mozer
University of Colorado

Training that uses exaggerated versions of a stimulus discrimination (*fading*) has sometimes been found to enhance category learning, mostly in studies involving animals and impaired populations. However, little is known about whether and when fading facilitates learning for typical individuals. This issue was explored in 7 experiments. In Experiments 1 and 2, observers discriminated stimuli based on a single sensory continuum (time duration and line length, respectively). Adaptive fading dramatically improved performance in training (unsurprisingly) but did not enhance learning as assessed in a final test. The same was true for nonadaptive linear fading (Experiment 3). However, when variation in length (predicting category membership) was embedded among other (category-irrelevant) variation, fading dramatically enhanced not only performance in training but also learning as assessed in a final test (Experiments 4 and 5). Fading also helped learners to acquire a color saturation discrimination amid category-irrelevant variation in hue and brightness, although this learning proved transitory after feedback was withdrawn (Experiment 7). Theoretical implications are discussed, and we argue that fading should have practical utility in naturalistic category learning tasks, which involve extremely high dimensional stimuli and many irrelevant dimensions.

Keywords: learning, perceptual learning, fading, category learning

While there is a large body of literature on the learning of perceptual categories by human beings, a majority of articles in this area have focused on the structure of categories and their mental representations, rather than the procedures used in training. Perhaps for this reason, while much has been learned about cognitive representations in category learning (e.g., Ashby & Maddox, 2005; Nosofsky, Palmeri, & McKinley, 1994; Shepard, Hovland, & Jenkins, 1961; Smith & Minda, 2002), less is known about the processes that establish these representations. Another consequence, perhaps, is that categorization research seems to have had relatively little translational impact on real-world training procedures despite the fact that perceptual category learning is a vital aspect of many important real-world skills in fields ranging from dermatology to mineralogy, from bird watching to naval aviation.

Harold Pashler, Department of Psychology, University of California, San Diego; Michael C. Mozer, Department of Computer Science, University of Colorado.

This work was supported by a Multidisciplinary University Research Initiative (MURI) award from the Office of Naval Research (25684A), by grants from the Institute of Education Sciences (Grant R305B070537 to Harold Pashler) and the National Science Foundation (Grant SBE-0542013; G. W. Cottrell, principal investigator), and a collaborative activity award to Harold Pashler from the J. S. McDonnell Foundation. We thank Joshua Martinez and Stephanie Ho for programming assistance, Michael Guo for assistance with data analysis, and Noriko Coburn for overseeing data collection and assisting with data analysis and with the preparation of this article.

Correspondence concerning this article should be addressed to Harold Pashler, Department of Psychology, University of California, San Diego 9500 Gilman Drive #0109, La Jolla, CA 92093-0109. E-mail: hpashler@ucsd.edu

The current article explores an instructional procedure that enjoys a long history within the field of experimental psychology and is often discussed under the rubric of *fading*. We used this term to refer to the deliberate exaggeration of a perceptual distinction in order to help the learner to acquire the distinction (as mentioned later, the term also has another common meaning that is quite different). As we will argue shortly, the literature provides little specific guidance about when fading is likely to help healthy individuals learn, and when it is not. The present research explores this issue by comparing fading with a more straightforward training regimen using a variety of categorization tasks.

Two Types of Fading

Fading is a term that has been used within the learning-theory and behavior-analysis literatures to refer to two quite different procedures for teaching discrimination tasks (encompassing what cognitive psychologists would call *category learning tasks*). In both types of fading procedures, a highly salient cue is introduced at the outset of training, making correct responding relatively easy. This cue is then gradually removed (*faded*), allowing the relevant stimulus property to acquire "control" over the response (i.e., to reliably elicit it). In one form of fading (*cross-dimensional fading*), the initial cue is a feature orthogonal to the discrimination the instructor seeks to teach. Thus, to teach someone to discriminate between two Turkish letters using cross-dimensional fading, one could start by making one of the letters red and the other green. After time, one could fade out these colors until the form discrimination came to control the response.

Cross-dimensional fading has sometimes been found effective in training animals (Doran & Holland, 1979; Terrace, 1963). It has also been extensively studied as a means of training children who are developmentally delayed (Jones & Eayrs, 1992). However, Ploog and Williams (1995) found in a rigorous study with pigeons that it was actually less effective than trial-and-error learning. As Ploog and Williams pointed out, this should perhaps not be altogether surprising, because cross-dimensional fading is identical to the associative blocking procedure, which is famous for slowing, rather than enhancing, associative learning (Kamin, 1969).

The second form of fading—which is the focus of the current article—works quite differently. This procedure, termed *transfer-along-a-continuum fading* by Ploog and Williams, was apparently first explored with animals by Lawrence (1952). In this procedure (which will be referred to henceforth simply as *fading*), the learner is exposed to exaggerated versions of the stimulus discrimination that must be mastered. Ploog and Williams (1995) found this form of fading highly advantageous in teaching pigeons a flicker-rate discrimination. The birds were taught to peck in response to a flickering light (reinforced stimulus or $S+$) that flashed on and off, remaining on for 88 ms and off for 110 ms, and ignore a light ($S-$) that remained on for 88 ms and off for 88 ms. In the fading condition, the initial S+ involved a more extreme difference (88 ms and 1,770 ms, respectively); this difference was then reduced over the course of training. Fading produced much more rapid learning than a trial-and-error procedure in which flicker rate was not changed over the course of training. Working with the octopus, Sutherland, Mackintosh, and Mackintosh (1963) found similar results.

Turning to studies involving human beings, there are some hints in the literature that fading can be helpful. The most systematic body of favorable evidence comes from research by Jamieson and his colleagues in the field of audiology (Jamieson & Morosan, 1989; Jamieson & Rvachew, 1992). Their primary focus was on methods of teaching native speakers of French the voiced/voiceless "th" distinction (*theta* vs. *the*), which is critical in comprehending English but notoriously difficult for francophones to master. Jamieson and colleagues found that a fading procedure using synthetic speech sounds that exaggerated the underlying formant differences greatly facilitated francophones' acquisition of this distinction (see also Temple et al., 2003). However, more recent studies of fading in teaching speech contrasts did not find very large benefits of fading when feedback was provided, although in the presence of feedback, the fading group did generally show somewhat faster learning (McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002; McClelland, Fiez, & McCandliss, 2002). There is also a substantial literature demonstrating benefits of fading in training of developmentally delayed children (e.g., Strand & Morris, 1986). Further afield, Ahissar and Hochtein (1997) found that pre-exposing people to a display for a considerable period of time enhanced their subsequent learning from very brief masked presentations.

Current Research

The goal in the present research was to examine the relative effectiveness of fading in several different situations that have not, to our knowledge, been contrasted. All our experiments involved training with feedback using healthy adult learners. In the first three experiments, we compared a *fading* condition with a *difficult* control condition using category-critical dimensions that the subjects were familiar with (namely, temporal duration and line length) and informing the subjects about the dimension on which they were to focus. The tasks required learning relatively fine perceptual discriminations, presumably meaning that perceptual noise was a main aspect of what makes the tasks difficult. In the first two experiments, fading was accomplished using a staircase algorithm that adjusted difficulty to keep the task relatively easy for subjects, while the third experiment started with a very easy discrimination and increased the difficulty according to a fixed schedule that operated without regard to the learner's successes or failures. After that, we present experiments in which, in addition to a category-relevant dimension, the stimuli also contained variation on category-irrelevant dimensions.

As will be argued in the General Discussion, the efficacy of fading has substantial theoretical interest as well as a great deal of practical interest. Fading greatly reduces the number of errors made in training, and for this reason, examining its effects may help in assessing the validity of error-correction models that propose that representations of category structure are modified only after errors. The efficacy of fading may also shed light on computational models according to which category members lying close to the decision boundary constrain the learner's internal representation of the boundary. Finally, as we see in the General Discussion, it may also shed light on the role of attention in perceptual category learning.

Experiment 1: Auditory Duration Discrimination

In Experiment 1, subjects learned to distinguish "long" durations (greater than 500 ms) from "short" durations (less than 500 ms). All subjects performed 12 training blocks followed by four test blocks; each block consisted of 40 trials. In the fading condition, the long and short durations began at values selected to be easily distinguished (600 vs. 400 ms, respectively), and the difference was adjusted after every trial using a staircase algorithm that maintained a high level of subject performance. In the difficult condition, the durations were 530 and 470 ms, respectively. These same values were used in the test for all subjects.

Method

Participants and design. Subjects were 84 undergraduate students at the University of California, San Diego, who participated in a lab experiment for course credit.

Design. Subjects were randomly assigned to the fading condition or the difficult condition (between-subjects design).

Stimuli. The stimulus on each trial was a 600-Hz tone played through loudspeakers that varied from trial to trial only in its duration.

Procedure. Subjects were told the following:

> This is an experiment on the perception of brief time intervals. On every trial, you will hear a tone. The duration of the tone will be approximately 1 half second. But sometimes it will be a little longer than 0.5 s—this makes it a "long tone"—and sometimes it will be a little shorter than 0.5 s—this makes it a "short tone." Your job is simply to try to decide if you heard a short tone or a long tone on each trial, and respond accordingly.

Subjects took as long as they wanted to respond, and they were also advised that the difficulty of the task (proximity to the 500-ms boundary) might vary.

There were 12 blocks of 40 training trials followed by four blocks of 40 test trials. On each trial, a fixation cross appeared for 500 ms, followed by a blank period of 500 ms, followed by the playing of the tone. The subject responded by pressing the *N* or *M* keys for short and long duration, respectively. After the subject responded, feedback was provided (a display of *Correct* or *Error* on the screen for 1,000 ms), followed by a 1,500-ms pause before the fixation cross for the next trial.

For subjects in the difficult condition, the time durations used were 530 and 470 ms throughout both the training and test phases of the experiment. For subjects in the fading condition, the time durations were adjusted from their starting point of 600 and 400 ms, respectively, with a staircase algorithm. When the subject made a mistake, the gap between the long and short durations was expanded by 20 ms; when the subject made correct responses on three consecutive trials, the gap was contracted by 20 ms with the constraint that the gap could not get any smaller than that used in the difficult condition. Before the final test, subjects were told, “In the final set of four blocks, there will be no feedback after each trial. Please continue to do your very best.” The final test used the 470 ms/530 ms distinction for all subjects.

Results and Discussion

Sixteen participants failed to complete the study within the 1-hr time period during which they were available for testing, and their partial data were not included in the final analysis. The top panel of Figure 1 shows the mean gap between the long and short durations as a function of training block for the two conditions. The difficult condition intervals remained at their preset values throughout training, while the differences in the fading condition tended on average to decline over the first half of training, followed by what appears to be a rough plateau.

The middle panel of Figure 1 shows average accuracy over the training blocks for the two conditions. As expected from the nature of the staircase algorithm, accuracy stayed close to 80% in the fading condition. The lower panel shows performance on the final four test blocks. Overall performance was 67.8% in the fading condition and 66.1% in the difficult condition; the 1.7% difference was not significant, $t(66) = 0.82$, $p = .41$, $d = 0.20$. The 95% confidence interval (CI) on the 1.7% difference ranged from –2.5 to 6.0.

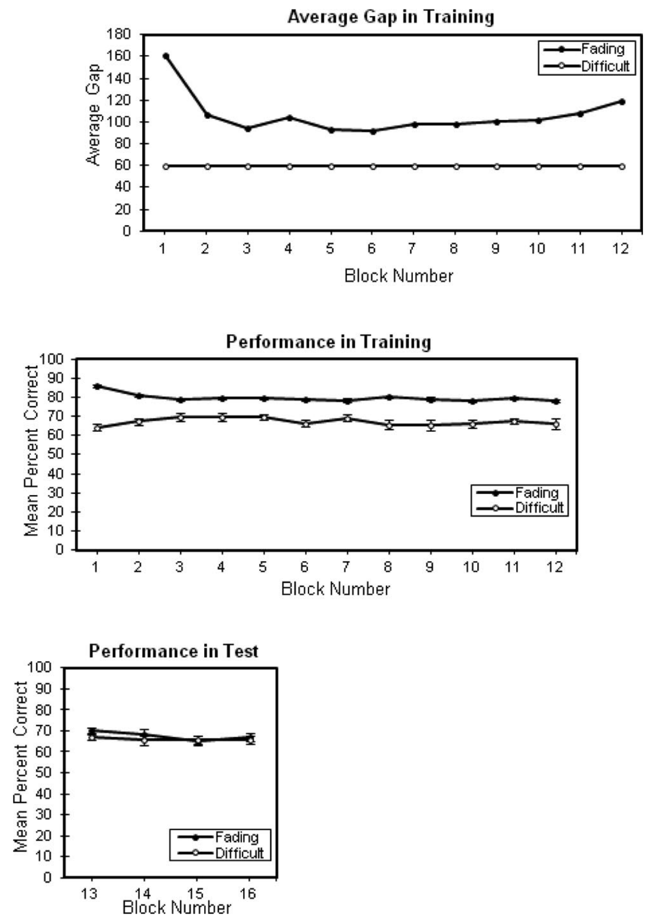

Figure 1. Results of Experiment 1. Top panel: Average gap in milliseconds as a function of block and condition (gap is fixed at 60 for the difficult condition, whereas it starts at 200 and is adjusted adaptively for the fading condition). Middle panel: Average percentage correct for the two groups during the 12 training blocks. Bottom panel: Average percentage correct for the two groups during the final four test blocks (using the same gaps as in the difficult condition).

Experiment 2: Line Length Discrimination

Experiment 2 had the same design as that of the previous experiment, except that the discrimination involved the length of lines. All aspects of the method were the same as in Experiment 1 except as noted in the following section.

Method

Twenty-eight subjects from the same population participated in the experiment. The stimuli on each trial were vertical lines 5 pixels in width. In the difficult condition, subjects saw either a long or a short line (104 or 96 pixels long, respectively—the same values used in the test). In the fading condition, the two lengths were initially set to 120 and 80 pixels, respectively, and adjusted using the staircase algorithm. In the final test, the difficult values were used for all subjects. The position of lines on the screen was randomly jittered from trial to trial so that subjects would need to assess length rather than position of the end point of the lines.

Results and Discussion

The top panel of Figure 2 shows the line length values used in training, and the middle panel of Figure 1 shows mean performance in training. In training, the fading condition did somewhat better on average (86% for fading vs. 83% for difficult). This trend was not significant, however, $t = 1.6$, $p = .12$, $d = 0.62$. However, in the final test (Figure 1, bottom panel), the fading condition was outperformed by the difficult condition, though the 3.2% difference was not significant, $t(25) = –1.23$, $p = .23$, $d = –0.47$ (95% CI [−8.62 to 2.17]).

Experiment 3: Linear Fading

Is the failure to find a substantial benefit of fading due to the use of a staircase algorithm to control stimulus differences? One

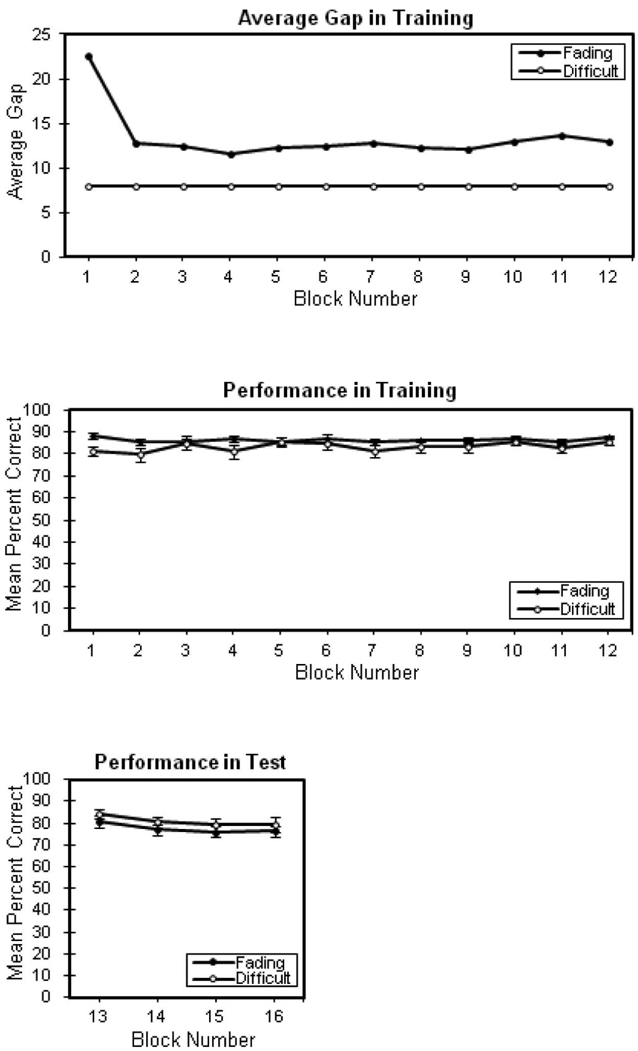

Figure 2. Results of Experiment 2. Top panel: Average gap in pixels as a function of block and condition (gap is fixed at 8 for the difficult condition, whereas it starts at 40 and is adjusted adaptively for the fading condition). Middle panel: Average percentage correct for the two groups during the 12 training blocks. Bottom panel: Average percentage correct for the two groups during the final four test blocks (using the same gaps as in the difficult condition).

consequence of the staircase algorithm is that for the average subject, the gaps between stimuli never reach the same difficulty level as that used in test (although for some individual subjects, they did.). One might surmise that for this reason, this adaptive form of fading might not work best.

Method

In Experiment 3, we used the same task as in Experiment 2, but here the fading group experienced a predetermined schedule for reducing the length difference over time: the difference began at 52 pixels and was reduced by 4 pixels every block, ending up with a block of practice at 8 pixels (the same degree of difference used in test, and in the difficult condition throughout training).

Results and Discussion

Thirty-four subjects from the same population participated in the experiment. The top panel of Figure 3 shows the line length values used in training, and the middle panel shows average performance during training. As expected, the fading group outperformed the difficult group, although the difference disappeared in the final blocks of training when the two conditions converged. However, in the final test (top panel), the fading group performed comparably (75.8%) to the difficult group (74.9%), $t(32) = 0.28$, $p = .78$, $d = 0.10$. Again there was a minor difference in degree of learning seen in the fading group, but the difference (0.96%) is small (95% CI [–5.89, 7.80]).

Experiment 4: Length Embedded in Noise Dimensions

The results thus far show that fading produces major enhancements in performance during training—as expected since it makes

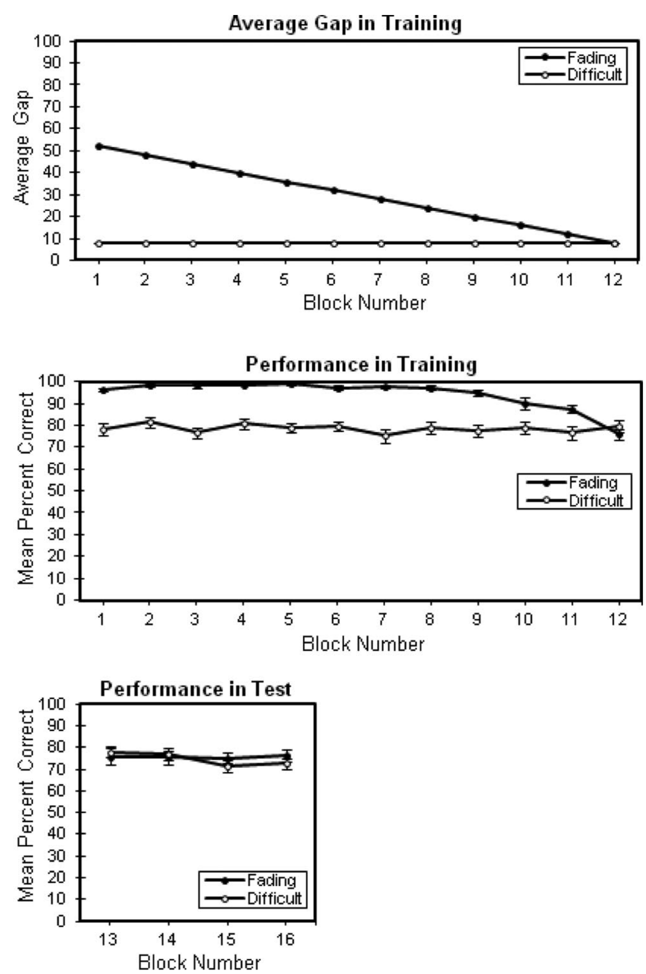

Figure 3. Results of Experiment 3. Top panel: Average gap in pixels as a function of block and condition (gap is fixed at 60 for the difficult condition, whereas it started at 52 and was adjusted linearly for the fading condition). Middle panel: Average percentage correct for the two groups during the 12 training blocks. Bottom panel: Average percentage correct for the two groups during the final four test blocks (using the same gaps as in the difficult condition).

the discrimination easier. However, these changes in performance have not yielded any enhancement of learning, as assessed in the final difficult test. In the next experiment, the task involved the same length discrimination that was used in Experiment 2, but the stimuli varied on multiple additional dimensions, all of which were task-irrelevant.

Method

Subjects were told that they would learn to distinguish "New World demons" from "Old World demons" (see Figure 4). The demons varied in four dimensions: horn height, eye diameter, lightness of the head, and presence of a nose. Horn height was the only dimension that predicted category membership, and the distribution of horn height values in this experiment exactly tracked (on average) the distribution of line lengths used in Experiment 2. The values on the other three dimensions were drawn (randomly and independently on each trial) from either a uniform distribution (eye diameter and lightness of the head) or by a coin flip (presence of a nose) without regard to whether the demon was Old World or New World. Thus, the variation in horn height—the sole predictor of category membership—was embedded in variability on three nonpredictive dimensions. Subjects saw 480 individually created demons in training and 160 in the test.

Results

Thirty-five subjects from the same population participated in the experiment. The top panel of Figure 5 shows the horn height values used in training, and the middle panel shows average performance during training. The difficult group showed scarcely any learning whatsoever. By contrast, the fading group performed quite well (averaging 86.1% from Blocks 3 to 12). In the final test (top panel), the fading group maintained its superiority, performing at 69.0%, compared with 49.0% for the difficult group, $t(33) = 7.72$, p < 001, $d = 2.58$.

Discussion

The benefit of fading seen in the final test averaged 20.1%, an order of magnitude larger than the largest fading advantage seen in any of the first three experiments (1.75% in Experiment 1).

Experiment 5: Expanding the Length Difference

Experiment 4 showed a dramatic benefit of fading training when the learner was confronted with a task requiring the identification of the relevant dimension from among a number of varying (and task-irrelevant) dimensions. However, the difficult group appears to have learned virtually nothing in Experiment 4, which perhaps makes the advantage of the fading group arguably less meaningful. To see if the advantage of fading remained when the discrimination task confronting the difficult group in training (and both groups in test) was one that could be acquired to some degree, we increased by a factor of 2.5 the discrimination difference used in the difficult training (as usual, the same values as in the difficult condition were used in the final test for all subjects). The methods followed those of Experiment 4 except as noted.

Method

Thirty-six subjects participated. For the difficult group (and in the final test for all subjects,) horn heights were 90 and 110 pixels for Old World and New World demons, respectively. The starting values for the fading group were the same as in Experiment 4.

Results

Two subjects did not complete the study within 1-hr timeframe, and one subject did not follow instructions (he left the experimental booth on three occasions, asking questions focused on personal difficulties he was having that were unrelated to the study). Removing these subjects left 33 participants in the experiment. The top panel of Figure 6 shows the horn height values used in training, and the middle panel shows average performance during training. The difficult group showed substantial learning, with performance averaging 61.8% over the second half of the training phase. By contrast, the fading group performed even better than in Experiment 4 (averaging 98.1% in the second half). In the final test (top panel), the fading group maintained its superiority, performing at 97.8%, compared with 67.1% for the difficult group, $t(31) = 5.52$, $p < .001$, $d = 1.95$.

Discussion

Although both groups showed learning in this experiment, the fading group showed an even larger benefit (30.6%) than in the previous study (95% CI [18.3, 42.5]).

Experiment 6: Erasing the Fading Advantage Through Verbal Instructions

The preceding two experiments showed dramatic benefits of fading in the context of a trial-and-error learning task in which the learner had to figure out what dimension was relevant. If it is indeed this opacity that produced the fading advantage—rather than, say, the variability across dimensions per se—we should expect that the advantage could be eliminated by telling all learners what dimension was relevant. The current study followed the methods of Experiment 4, except as noted.

Method

Thirty-four subjects participated. After the subject had completed Block 6 (the halfway point in training), all subjects, regardless of condition, saw a message presented on the computer screen.

> Be advised: The key difference between Old World and New World demons relates to the height of the horns. There are other dimensions on which the stimuli vary randomly, but horn height is the one that helps to distinguish Old World versus New World.

Before proceeding, subjects had to affirm, "I understand that horn height is the only dimension that distinguishes Old World and New World demons."

Results

The top panel of Figure 7 shows the horn height values used in training, and the middle panel shows average performance during

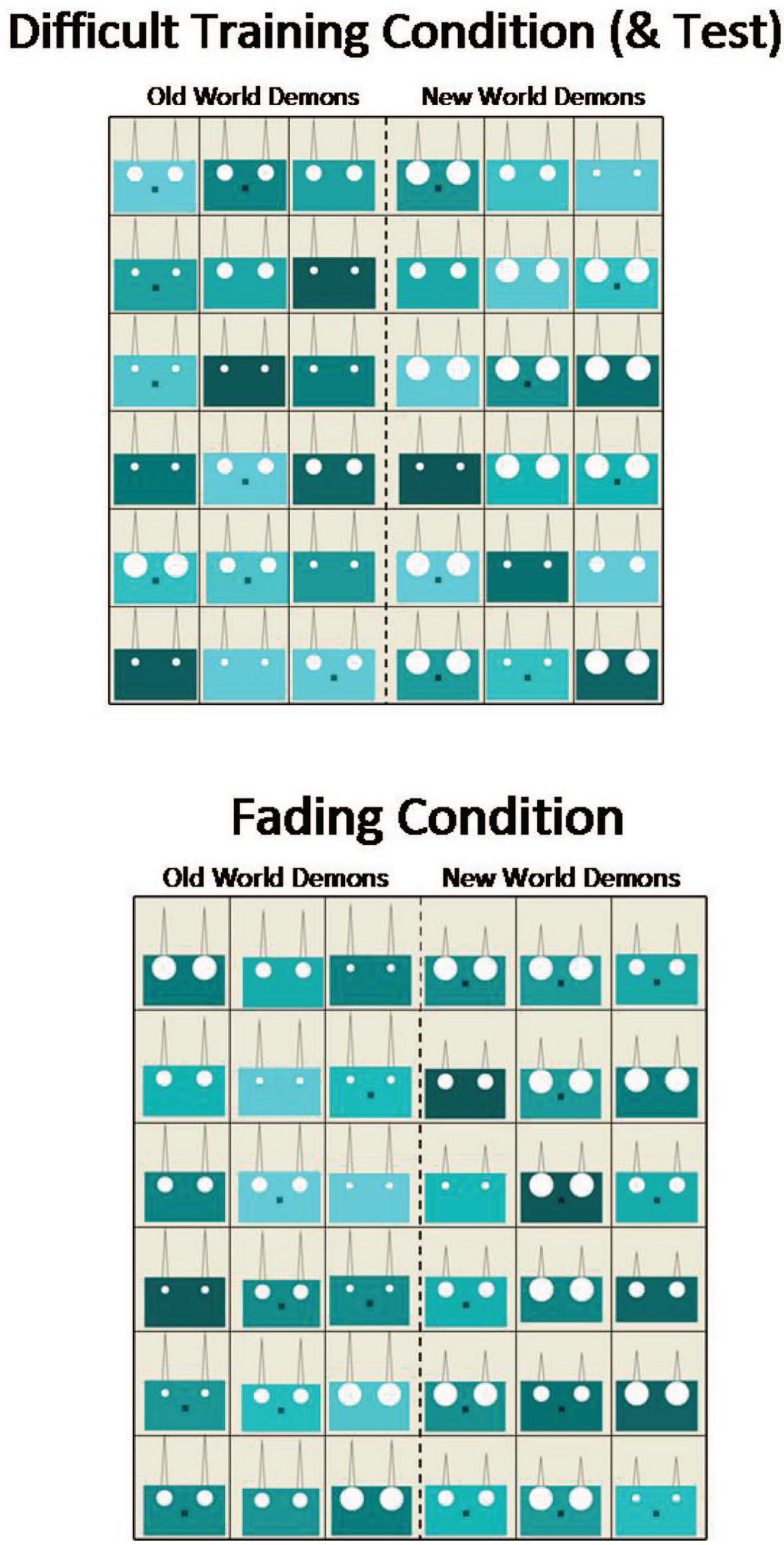

Figure 4. Examples of stimuli used in Experiment 4. Top panel shows stimuli from difficult condition. Bottom panel shows stimuli from fading condition (initial value). Horn height is the relevant dimension, but brightness, eye size, and presence of a nose all vary.

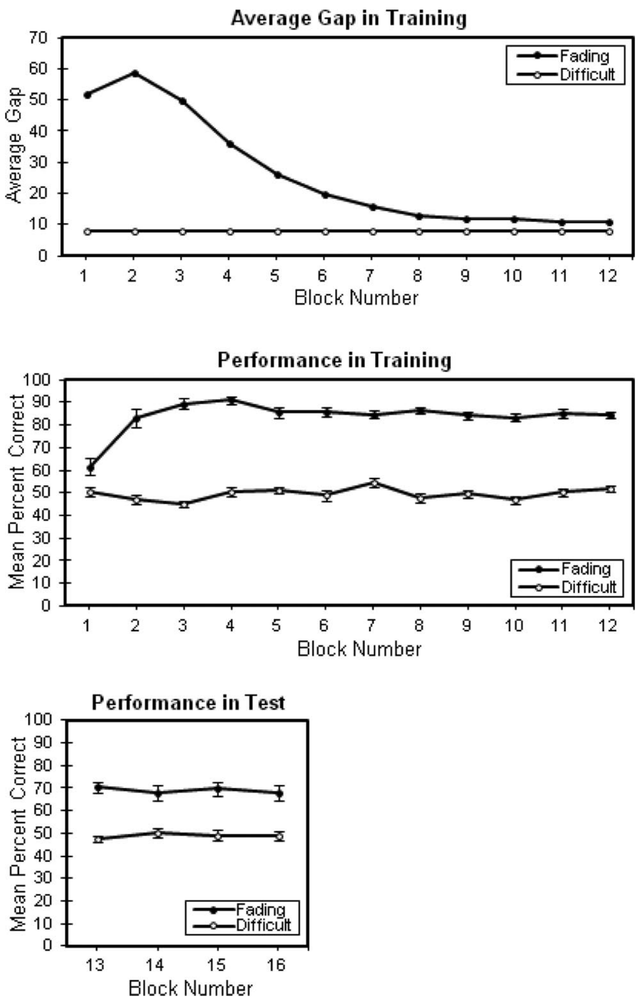

Figure 5. Results of Experiment 4. Top panel: Average horn height gap in pixels as a function of block and condition in a "demon categorization" task with demons that varied in horn height and four other dimensions. Middle panel: Average percentage correct for the two groups during the 12 training blocks. Bottom panel: Average percentage correct for the two groups during the final four test blocks (using the same gaps as in the difficult condition).

training (with an arrow indicating the point at which the special instruction was provided to both groups). As in Experiment 4, the difficult group showed no detectable learning in the first portion of training (averaging 49.5% for Blocks 1–6). However, after the message was displayed, this group showed an immediate jump in performance. Interestingly, so did the fading group. In the final test (bottom panel), the fading group performed worse at 69.5%, compared with 75.1% for the difficult group, a nonsignificant difference, $t = 1.04$, $p = .31$, $d = -0.36$.

Discussion

The results are consistent with the idea that the increase in learning due to the fading condition seen in Experiments 4 and 5 reflected a conscious discovery of the predictive relationship by this group. The explicit coaching provided here yielded an immediate advantage to both groups, leaving no residual learning advantage for the fading condition.

Experiment 7: Fading with a Hard-to-Verbalize Dimension

We have seen that fading can help the learner to discover discrete and simple and easily verbalized predictors of category membership that are obscured by the presence of variation in other dimensions that are not related to category membership. What would happen if the subject must learn about a predictor that is not so easily verbally labeled? To explore this issue, we used stimuli that varied in hue, saturation, and brightness, with only saturation determining category membership. Learners were advised merely that "color" was the relevant dimension.

Method

The task again involved sorting demons that varied in hue, brightness, and saturation (the demons had fixed values of eye

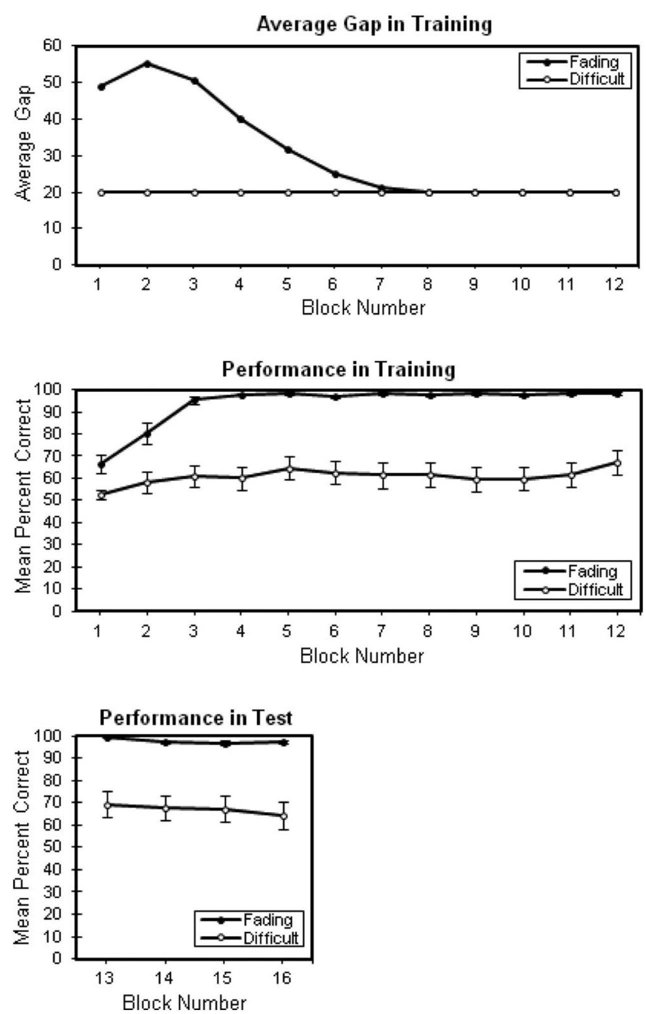

Figure 6. Results of Experiment 5 (similar to Experiment 4 but with greater difference on the relevant dimension of horn height). Panels show average horn height gap (top), performance in training (middle), and test (bottom).

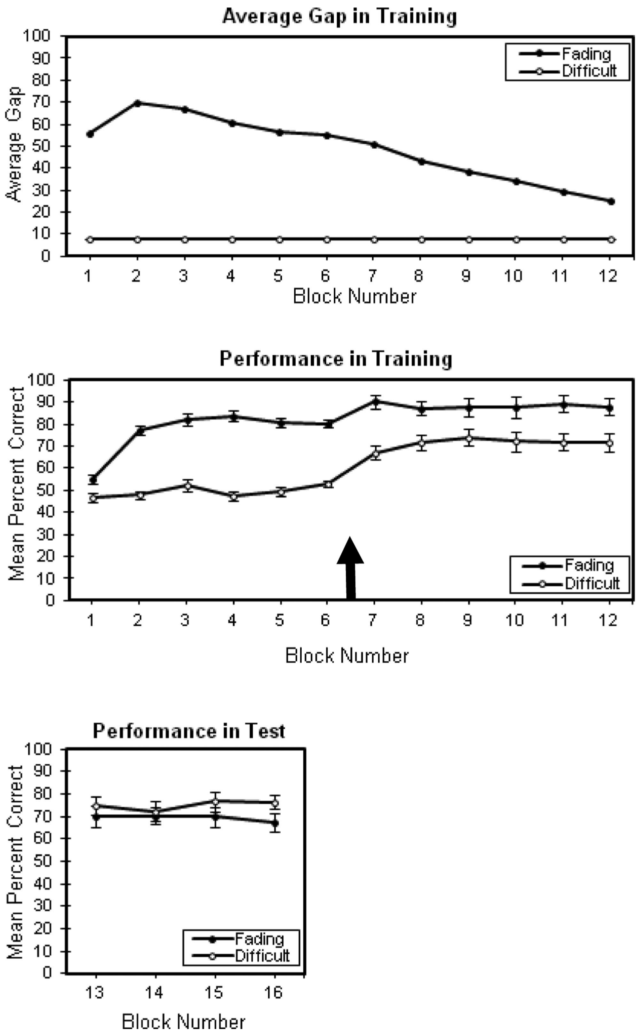

Figure 7. Results of Experiment 6 (like Experiment 4 but with explicit cuing of relevant dimension, provided between Blocks 6 and 7 and marked by arrow). Panels show average horn height gap (top), performance in training (middle), and test (bottom). Cuing benefits both groups in training, reversing the fading advantage seen in test.

height and horn height, and they always had noses). Participants were told, "[T]he difference between Old World and New World depends upon something about the color of the demon. However, it is not a simple rule such as 'Old World demons are red' or anything like that." Hue values were chosen uniformly over the entire range (0–1), and brightness values were chosen from the range between 0.5 and 0.9 on a 0–1 scale. In the difficult group training and in the final test, the saturation values were .55 for Old World demons and .45 for New World demons on a 0–1 scale. For the fading condition, the saturation values began at .65 and .35 (a 2× stretching of the saturation range) and were adjusted by .01 with the staircase algorithm used in prior experiments. All color values were represented in MATLAB's (MathWorks, Natick, MA) hue, saturation, and brightness (HSB) color space, and MATLAB was used to convert from HSB to red, green, and black (RGB) values.

Results

The top panel of Figure 8 shows the saturation difference between Old and New World demons used in training, and the middle panel shows average performance during the training phase. The difficult group showed no detectable learning (averaging 50.8% for Blocks 7–12). The fading group, on the other hand, reached 77.4% for Blocks 7–12. In the final test (bottom panel), the fading group performed better (60.9%) than the difficult group (51.0%), a significant difference, $t = 4.34$, $p < .001$, $d = 1.64$. During the test, the fading group appeared to lose some of its gains (averaging 68.2% in Block 13 falling to 57.1% in Block 16, a significant drop, $t = 3.5$, $p < .005$.).

Discussion

Here in a task requiring people to acquire a subtle discrimination based on a relatively unfamiliar and hard-to-verbalize aspect of

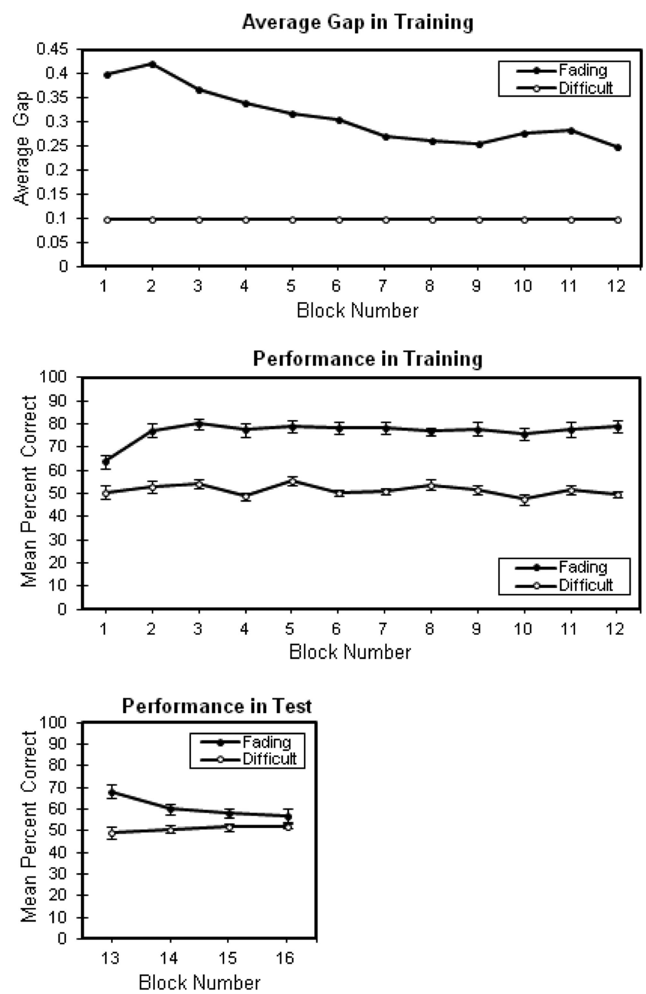

Figure 8. Results of Experiment 7. Panels show average gap in hue (top), performance in training (middle), and test (bottom) for a color-based "demon categorization" task in which the relevant variable is saturation, with task-irrelevant variation in hue and brightness. Here, too, fading proved beneficial, although the gains appeared to diminish during the (feedback-free) test trials.

color, fading again enhanced learning. (In a follow-up study, we allowed saturation to vary as in the current experiment, but with hue and brightness values fixed throughout for a given subject. To our surprise, the difficult group still showed almost no learning in that experiment. The advantage of fading was again confirmed in the final test, albeit apparently somewhat smaller in magnitude.)

As in Experiments 4 and 5, the color stimuli used in this experiment were multidimensional, and the learner's task was not only to identify a category boundary along one dimension but also to identify the relevant dimension. From a generative perspective, the color stimuli are three dimensional, but there are not only multiple potential encodings of these three dimensions—HSB; RGB; long, medium, and short (LMS); and so on—but there are many potential verbal dimensions that might be used for color classification (pastel, neon, typical car colors, and so on). From the subject's perspective, the task in this experiment may have involved selection from a very high dimensional, redundant space of possible encodings to identify the critical dimension.

General Discussion

The research described here was motivated by the observation that although fading has been used extensively over the years with some success in training animals and impaired human populations, little was known of when it is and is not effective in enhancing perceptual category learning for typical, healthy adults. The results can be fairly easily summarized: "stretching" the critical distinction does not seem to help in the acquisition of a difficult discrimination that relies upon a simple perceptual dimension that can be (and is) well specified to the learner. (Naturally, we cannot rule out the possibility that some different form of fading than either of the types examined here might provide big advantages in unidimensional discriminations. For example, it might be interesting to examine an adaptive fading that ends with several blocks of training at the difficult level used in test.)

However, where fading did prove extremely helpful in our experiments was when the dimension was not specified in advance, and category-relevant variation was accompanied by category-irrelevant variation on other "noise" dimensions. Moreover, fading also proved helpful when the discrimination relied upon a feature that was not easily verbalized (color saturation). One rather natural suggestion would be that fading may be helping the learner to achieve a more effective attentional selection of the relevant dimension (see Kruschke, 2005, for discussion of attention and category learning.) This interpretation seems consistent with the fact that when a verbal specification of the relevant dimension was provided partway through training on a multidimensional stimulus (Experiment 6), the benefit of fading was augmented for the remainder of the training session but with no advantage for fading appearing in the final test.

Theoretical Implications

There has been one interesting theoretical effort to develop a Hebbian learning account of why fading might be useful for the special case in which the learner has been extensively pretrained in one classification boundary and now seeks to relearn a different boundary which reclassifies familiar stimuli—as happens for example, when native Japanese speakers attempt to master the [r]–[l] phonemic distinction. McClelland, Thomas, McCandliss, and Fiez (1999) offered a Hebbian account—basically suggesting that a deeply ingrained percept (e.g., a sound category in one's native language) can function as an "attractor" that interferes with the ability to perceive distinctions lying within the boundaries of the familiar category. McCandliss et al. (2002) compared the effects of multiple training sessions conducted using an adaptive training (fading) procedure versus a fixed training procedure for training Japanese natives in this category. When feedback was provided, the adaptive training enhanced learning somewhat more than the fixed training procedure. (Without feedback, neither procedure was particularly effective.) On the other hand, 3 days of training produced considerable learning even in the fixed training procedure, which the authors acknowledged to be something of a challenge for their model (see also McClelland et al., 2002).

Obviously, the situation of the native speaker of Japanese trying to acquire a phonological distinction not present in their native language may differ in important ways from any of the cases studied in the current study. While some of our tasks involved difficult discriminations obscured by noise, none of them would seem to involve the dramatic potential for negative transfer that was present in the phonemic training case, where the learner had already had a vast amount of experience applying a different decision boundary to the very same stimuli.

According to many influential theoretical analyses, human category learning is completely or mostly driven by the making and detection of errors (e.g., Gluck & Bower, 1988; Kruschke, 1992; Nosofsky et al., 1994). According to this analysis, when the learner classifies an object and gets feedback saying that the judgment was wrong, mental representations of the category are modified. On the other hand, when no error is made, category representations are generally left unmodified. From this perspective, one would expect that fading—which seeks to minimize errors—would not be a very helpful training procedure.

Essentially the same prediction arises from the more abstract perspective of computational theories in the field of machine learning. Consider n dimensional training stimuli as points in an n-dimensional space. The goal of learning is to discover the *decision boundary* that separates examples of one category from examples of the other category. Different types of machine-learning classifiers make different assumptions about the shape of the decision boundary. The simplest decision boundary is a hyperplane (or a two-dimensional line) that cuts the input space in half, and such a boundary is achieved by a linear classifier. Regardless of the form of the classifier, the examples close to the category boundary provide the strongest constraint on the boundary. Thus, the early examples of difficult training are more useful to the learner than the early examples of fading due to their proximity to the category boundary. For this reason, traditional linear classifiers tend to work better when trained on difficult stimuli. The same is even more dramatically true for the sophisticated modern classification algorithms widely used in computer applications, such as support vector machines (SVMs; Vapnik, 1995). SVMs are explicitly adjusted on the basis of the "marginal examples" found to lie on or near the classification boundary. The fading intervention reduces both the near-boundary density and number of the cases presented near the margin.

A recent analysis based on computational learning theory (Khan, Zhu, & Mutlu, 2011) offers an alternative framework for understanding the benefit to a learner of a particular sequence of training examples. The analysis assumes that the learner is theoretically optimal, meaning that the learner makes the best possible use of examples he or she is provided to home in on the category boundary. The analysis further assumes that the decision boundary is a hyperplane aligned with all axes of the stimulus space but one. That is, the decision boundary can be cast as a threshold on a single dimension of possibly multidimensional stimuli (e.g., line segments longer than 3 cm). Thus, the learner's task involves identifying the relevant stimulus dimension and then determining the threshold value on that dimension. Khan et al.'s (2011) analysis indicates that if the stimulus space is one dimensional—and therefore that there is no need to determine which dimension is relevant—the teacher should select examples just below and just above the threshold, allowing the optimal learner to discover the category boundary with just two examples. However, if the stimulus space is *n* dimensional, the situation is quite different because the learner's first goal is to determine which dimension is relevant.

Kahn et al.'s mathematical analysis is based on the notion of a *hypothesis space*—a set of candidate hypotheses that the learner has not ruled out. Each hypothesis represents one possible category definition and is of the form, "The relevant dimension is *d*, and the threshold on dimension *d* for category membership is θ." The theoretically optimum learner rules out all hypotheses that are inconsistent with the examples he or she has been shown. For example, if the learner is shown a demon with a 4-cm horn and is told that it is an Old World demon, then the learner can eliminate the hypothesis that the relevant dimension is horn length and that any horn length greater than 5 cm is characteristic of an Old World demon. If the plausibility of a dimension is related to the number of remaining hypotheses based on that dimension, then in order to boost the plausibility of the true relevant dimension (e.g., horn height), the teacher should choose initial training examples that are as far from the decision boundary as possible, because these examples preserve the largest number of hypotheses on the true relevant dimension. Because feature values on the other dimensions are assumed to be chosen randomly, hypotheses pertaining to these dimensions will be efficiently ruled out, leaving only hypotheses pertaining to the relevant dimension. Once the relevant dimension has been determined, then clearly the examples should be chosen as close to the boundary as possible. The transition from determining which dimension is relevant to determining the boundary on that dimension yields a fading schedule. Through a formal framework of risk minimization, Kahn et al. can quantitatively specify the optimal fading schedule. The fading rate depends on the effective stimulus dimensionality, with faster fading being optimal the lower the dimensionality.

Although Khan et al.'s (2011) analysis was premised on an optimal learner—a learner with unlimited memory and processing power—the result is qualitatively the same for resource limited learners (X. Zhu, personal communication, May 4, 2012). Kahn et al.'s analysis was not cast in terms of selective attention, but one might consider the amount of attention drawn to a dimension as being proportional to the number of remaining hypotheses on that dimension. In this view, the benefit of fading is to drive the learner's attention to the relevant dimension as irrelevant dimensions are gradually eliminated due to inconsistencies with the examples presented. Sequence effects have also been observed in unsupervised learning (Clapper & Bower, 2002), which have been argued to arise from the build-up of expectations when runs of similar (same category) items are presented in sequence. It would be intriguing to develop a unified theory that covers both supervised learning tasks (as described in this article) and unsupervised tasks.

Other accounts of the fading advantage also deserve consideration. One such account is an alternative analysis of how attention is drawn to the relevant dimension. Selective attention to a dimension may be based on the sheer amount of variability present along this dimension. According to this account, it is not the distance between examples on the relevant dimensions that helps but rather is just the magnitude of the variation on that dimension. This might be expected to occur if, for example, learners search for the relevant dimension by starting with dimensions showing the most salient variation. This account makes an interesting prediction that can be tested in future research: increasing variability within an irrelevant dimension should produce a very marked interference with learning.

Another account, which strikes the authors as perhaps the most plausible of any discussed here, is that stretching the relevant dimension promotes identification of the relevant dimension because the learner is in fact computing a *noisy* correlation between category membership and dimension value for all of the dimensions in parallel. Although it is simplest to imagine this unfolding in parallel across dimensions, certain forms of sequential correlation models could also explain the data; for example, suppose people proceed one dimension at a time, computing a rough estimate of the correlation between that dimension and the category label, proceeding on to the next dimension if the observed correlation does not exceed a very high threshold ("If there is an obvious correlation between the feature and the category label, continue to investigate this dimension—otherwise, check another dimension"). For such a process, fading would have the effect of prioritizing the relevant dimension, thus enhancing both performance during training and also learning as seen on a later test. It seems possible that detailed examination of the data could distinguish the parallel and sequential models.

In the previous discussion, in speaking of a "noisy" computation of correlation, we meant that observations are wrongly encoded or misremembered; alternatively, dimension values may be perturbed by noise, or category labels may be occasionally flipped. The reason it is critical to assume that the calculation is noisy in some way is because if the calculation were accurate and free of random noise, the stretched case would actually yield the same correlation coefficient as the unstretched case (because stretching is a linear transformation that leaves correlation values unchanged). To put the point in crude terms, it is only if people are doing a "bad job" of correlating values on each dimension with category membership that it should help in any way to stretch the values on the relevant dimensions; if they were doing a perfect job of computing this correlation, the relevant dimension would emerge just as fast without the stretching. It will be interesting in subsequent work to see if it is possible to distinguish between this account and the alternative ones mentioned previously. The focus in the current article is on overall levels of performance in training, but the results also point to the likely promise of more fine-grained analysis of trial-by-trial data to shed new light on basic mechanistic questions surrounding perceptual category learning.

Practical Implications

Besides the intriguing theoretical possibilities just described, the current results also offer some guidance for development of training strategies with difficult discriminations that arise in real-world settings. It seems likely that the discriminations that provide a training challenge in fields from dermatology to radiology to bird watching are usually ones that involve learning to utilize cues that are probabilistic and not easily verbalized. Indeed, for many, the feature dimensions that help separate the different categories are not well described, even by experts. Assuming that is the case, the current results offer bits of both encouragement and discouragement regarding the potential for using fading to enhance training. On the one hand, the fact that naturalistic tasks involve hard-to-verbalize dimensions (similar in that regard to the color saturation dimension used in Experiment 7) is encouraging for the application of fading. Further, in contrast to laboratory tasks that utilize relatively low dimensional stimuli, naturalistic tasks involve stimuli that might potentially be encoded in very high dimensional spaces, and identifying the relevant dimension(s) will increase in difficulty with the number of dimensions. On the other hand, the fact that the structure of the stimuli is not generally well understood in practical domains raises the practical question of whether it would really be possible to generate "stretched" examples to use in training. (The fact that fading has been applied to speech stimuli may reflect the fact that this case is unusual in that speech-critical acoustic features are often both impossible to verbalize but are very well understood and able to be incorporated into artificial stimuli.)

The lack of full understanding of the generative process that produces different categories of natural stimuli is not necessarily a fatal problem, however. One potential strategy would be to use classification accuracy of a panel of human judges (who possess at least some competence in the relevant discrimination task) to rank the training stimuli in difficulty. Then, instead of constructing artificial stimuli with stretched intercategory differences, one could simply identify whichever natural stimuli are most easily classified by the panel. These, in turn, could be used early in training. Obviously, the current findings do not guarantee that this form of fading will necessarily be effective in training natural stimuli, but the current findings are consistent with the notion that this approach will succeed. We are currently examining this question using several different categories of natural stimuli. Relatedly, Love and Giguere (2011) conducted an intriguing recent study in which they tried a strategy of "idealizing" training, in which they deleted the members of each category that lay on the "wrong" side of the decision boundary that most readily separated the stimuli. This proved quite useful.

The present experiments explored what seemed like the natural starting point for research in this area, comparing two extreme training regimens (training with stretched stimuli vs. training with the most difficult stimuli). Given the observed benefit of fading, it is worth examining whether the details of the fading schedule matter, either the level of performance that should be maintained during training or the possibility that it might be better to fade on a fixed, performance-independent schedule. Further, it seems possible that the most effective training procedure might actually involve a blend of fading and difficult training. For example, one could use intermittent presentation of stretched stimuli while concentrating most of the learner's efforts in training on the difficult stimuli. It seems possible that the former might aid in periodically reinstating attention to dimensions if that is lost from time to time, while the latter might be most useful in allowing the learner to determine decision boundaries and retain these in memory.

References

Ahissar, M., & Hochstein, S. (1997, May 15). Task difficulty and the specificity of perceptual learning. *Nature, 387,* 401–406. doi:10.1038/387401a0

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56,* 149–178. doi:10.1146/annurev.psych.56.091103.070217

Clapper, J. P., & Bower, G. H. (2002). Adaptive categorization in unsupervised learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 908–923. doi:10.1037/0278-7393.28.5.908

Doran, J., & Holland, J. G. (1979). Control by stimulus features during fading. *Journal of the Experimental Analysis of Behavior, 31,* 177–187. doi:10.1901/jeab.1979.31-177

Gluck, M., & Bower, G. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117,* 227–247. doi:10.1037/0096-3445.117.3.227

Jamieson, D. G., & Morosan, D. E. (1989). Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology/Revue canadienne de psychologie, 43,* 88–96. doi:10.1037/h0084209

Jamieson, D. G., & Rvachew, S. (1992). Remediating speech production errors with sound identification training. *Journal of Speech–Language Pathology and Audiology, 16,* 201–210.

Jones, R. S. P., & Eayrs, C. B. (1992). The use of errorless learning procedures in teaching people with a learning disability: A critical review. *Mental Handicap Research, 5,* 204–212. doi:10.1111/j.1468-3148.1992.tb00045.x

Kamin, L. (1969). Predictability, surprise, attention, and conditioning. In R. Church & B. Campbell (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York, NY: Appleton–Century–Crofts.

Khan, F., Zhu, X., & Mutlu, B. (2011). How do humans teach: On curriculum learning and teaching dimension. *Advances in Neural Information Processing Systems (NIPS) 24.*

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44. doi:10.1037/0033-295X.99.1.22

Kruschke, J. K. (2005). Learning involves attention. In G. Houghton (Ed.), *Connectionist Models in Cognitive Psychology* (pp. 113–140). Hove, East Sussex, England: Psychology Press.

Lawrence, D. H. (1952). The transfer of a discrimination along a continuum. *Journal of Comparative and Physiological Psychology, 45,* 511–516. doi:10.1037/h0057135

Love, B., & Giguere, G. (2011, November). *Idealized training in noisy situations improves generalization.* Paper presented at the annual meeting of the Psychonomic Society, Seattle, Washington.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]–[l] contrast to Japanese adults: Predictions of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience, 2,* 89–108.

McClelland, J. L., Fiez, J. A., & McCandliss, B. D. (2002). Teaching the /r/–/l/ discrimination to Japanese adults: Behavioral and neural aspects. *Physiology & Behavior, 77,* 657–662. doi:10.1016/S0031-9384(02)00916-2

McClelland, J. L., Thomas, A., McCandliss, B. D., & Fiez, J. A. (1999). Understanding failures of learning: Hebbian learning, competition for representational space, and some preliminary experimental data. In J. Reggia, E. Ruppin, & D. Glanzman (Eds.), *Brain, behavioral, and cognitive disorders: The neurocomputational perspective* (pp. 75–80). Oxford, England: Elsevier. doi:10.1016/S0079-6123(08)63068-X

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–79. doi:10.1037/0033-295X.101.1.53

Ploog, B., & Williams, B. A. (1995). Two methods of stimulus fading applied to a simultaneous flicker rate discrimination in pigeons. *Learning and Motivation, 26,* 161–182. doi:10.1016/0023-9690(95)90003-9

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75*(13, Whole No. 517).

Smith, J. D., & Minda, J. P. (2002). Distinguishing prototype-based and exemplar-based processes in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 800–811. doi:10.1037/0278-7393.28.4.800

Strand, S. C., & Morris, R. C. (1986). Discriminations: A comparison of techniques. *Applied Research in Mental Retardation, 7,* 165–181. doi:10.1016/0270-3092(86)90003-2

Sutherland, N. S., Mackintosh, N. J., & Mackintosh, J. (1963). Simultaneous discrimination training of Octopus and transfer of discrimination along a continuum. *Journal of Comparative and Physiological Psychology, 56,* 150–156. doi:10.1037/h0044677

Temple, E., Deutsch, G. K., Poldrack, R. A., Miller, S. L., Tallal, P., & Merzenich M. M. (2003). Neural deficits in children with dyslexia ameliorated by behavioral remediation: Evidence from functional MRI. *PNAS: Proceedings of the National Academy of Sciences of the United States of America, 100,* 2860–2865. doi:10.1073/pnas.0030098100

Terrace, H. S. (1963). Discrimination learning with and without "errors". *Journal of the Experimental Analysis of Behavior, 6,* 1–27. doi:10.1901/jeab.1963.6-1

Vapnik, V. N. (1995). *The nature of statistical learning theory*. Berlin, Germany: Springer–Verlag.

Received June 29, 2012
Revision received September 13, 2012
Accepted September 19, 2012 ■