



ACADEMIC
PRESS

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Learning and Motivation 34 (2003) 52–67

**Learning
and
Motivation**

www.elsevier.com/locate/l&m

The joint contributions of shape and color to variability discrimination[☆]

Michael E. Young^{*} and Michelle R. Ellefson

*Department of Psychology, Southern Illinois University, 271F LSII, Mailcode 6502,
Carbondale, IL 62901-6502, USA*

Received 8 April 2002; received in revised form 5 July 2002

Abstract

Two experiments examined college students' discrimination of complex visual displays that involved different degrees of variability or "entropy." Displays depicted 16 line drawings of various shapes and colors; the participants needed to learn to classify a display in terms of its variability in shape or color while discounting variability in the other dimension (Experiment 1), or to classify based on both shape and color variability (Experiment 2). The participants' accuracy and reaction time scores on a two-alternative forced-choice discrimination disclosed that people can learn to discount variability in an irrelevant dimension when necessary, but this variability does affect performance. Our data further suggest that variability discrimination depends on degrees of similarity across multiple dimensions thus underscoring the shortcomings of a measure of categorical variability (e.g., entropy) that only considers whether items are identical or different.

© 2002 Elsevier Science (USA). All rights reserved.

Keywords: Variability; Entropy; Relational learning; Categorization; Relevant and irrelevant

Consider the drawings shown in Fig. 1. People can readily discriminate between the collection of identical stimuli in the upper left portion of Fig. 1 and the collection of different stimuli in the upper right corner of Fig. 1. This judgment is a function of

[☆] We would like to thank Dr. Edward Wasserman for useful discussions about this research and Yue Wang for his assistance in conducting the experiments.

^{*} Corresponding author.

E-mail address: meyoung@siu.edu (M.E. Young).

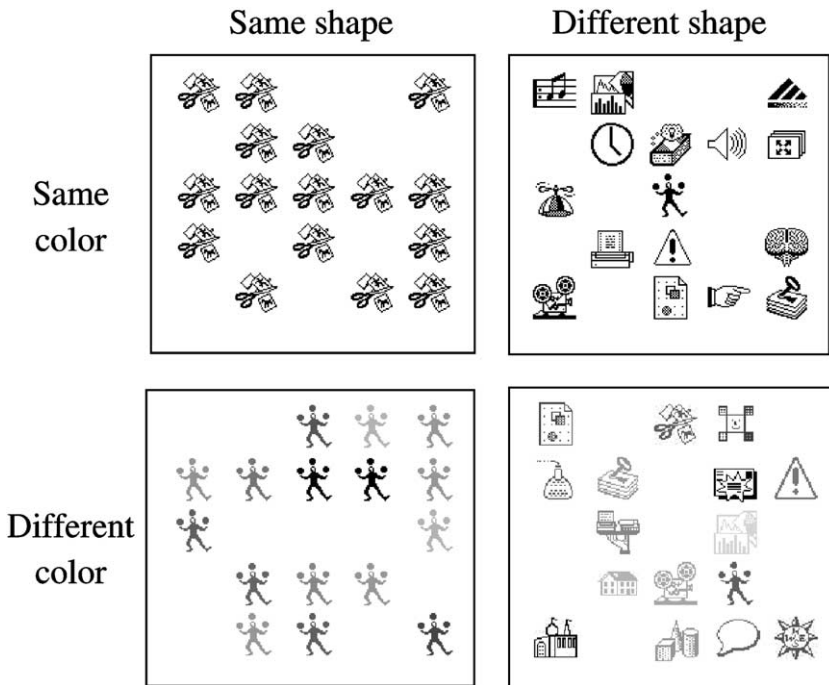


Fig. 1. Examples of the 16-icon same and different arrays used in the present study. The actual arrays involved up to 16 different colors, not the shades of gray depicted here.

the relationships among the display items and not the specific items that constitute the display. For example, a display with no variability may comprise 16 brains, 16 faces, or 16 compasses and still be readily recognized as involving 16 identical items. The relational nature of variability discrimination distinguishes the categorization of stimuli as a function of their relational similarity from the categorization of stimuli as a function of their perceptual similarity (e.g., by placing pictures of cats in one category and those of horses in another).

People judge among levels of categorical variability in the service of a range of everyday tasks. Judgment of variability arises in choosing a grocery store (does the store have a variety of brands and products?), judging the difficulty in providing computer support to a department (how many different computer platforms are in use?), and assessing racial diversity in a group of people (do I want to attend a school with such low diversity?). Psychologists have documented that people have preferences for certain levels of variety (Berlyne, 1960; Munsinger, 1966; Munsinger & Kessen, 1966) and marketing researchers have reported that perceived variety has a strong effect on decisions of where to shop (e.g., Baumol & Ide, 1956; Broniarczyk, Hoyer, & McAlister, 1998; Hoch, Bradlow, & Wansink, 1999). Despite the everyday relevance of categorical variability, only recently have researchers begun to examine how people perceive this variability.

Young and Wasserman (2001a) examined people's acquisition of a discrimination based on visual variability. In Experiment 1 of that study, people observed a series of displays like those shown in the upper half of Fig. 1 (all of the icons were black) and were trained (through corrective feedback) to press one key for same displays and a second key for different displays. When these same participants were later forced to classify displays with intermediate degrees of variability, most participants (80%) had a strong tendency to respond "different" to displays comprising any degree of variability and to reserve the "same" response for same displays—they showed a categorical distinction between no variability and some variability. The minority of participants (20%) showed a gradual transition from responding on the "different" key to responding on the "same" key as variability decreased; this transition was reminiscent of a standard psychophysical curve. In Experiment 2 of their study, however, all participants evidenced a psychophysical curve when Young and Wasserman required one response for a range of low variability displays (including those with some degree of differentness) and a second response for a range of high variability displays.

Measuring categorical variability

The sensitivity to categorical variability evidenced by people in the Young and Wasserman (2001a) study has also been observed in baboons (Wasserman, Fagot, & Young, 2001) and pigeons (e.g., Young & Wasserman, 1997). In all three species, Young, Wasserman, and colleagues have observed that behavior is strongly correlated with a statistical measure of categorical variability: *entropy*. To quantify entropy, they used the following equation (Shannon & Weaver, 1949):

$$H(A) = - \sum_{a \in A} p_a \log_2 p_a, \quad (1)$$

where $H(A)$ is the entropy of categorical variable A , a is a category of A , and p_a is the proportion of observed values within that category. When a display has 16 identical icons, there is only one category (the single icon) with a probability of occurrence of 1.0. Because $\log_2(1.0) = 0.0$, the entropy of a same display is 0.0. A different display consists of one occurrence of each of 16 icons (i.e., 16 shape categories), yielding an entropy of: $-.0625 \times \log_2(.0625) \times 16$, or 4.0. Displays with intermediate degrees of variability have intermediate levels of entropy.

To apply Eq. (1), all one needs to know is the category of each item (e.g., juggler, hand), and the relative frequency of each category. Entropy, however, is silent about *degrees* of category membership—any two items are either members of the same category (i.e., are the same) or they are not. The studies by Young, Wasserman and colleagues used displays comprising a collection of differently shaped icons that were chosen to be distinctly dissimilar. In contrast, a study of consumers' perception of variety (Hoch et al., 1999) used collections of 16 items that could differ in up to three dimensions (shape, color, and name). This study targeted judgments of product variety and their effect on buying decisions, a task quite different from those used in the

other studies, but the results suggest that degree of similarity is important to variability discrimination.

Multiple dimensions of variability

Hoch et al. (1999) presented participants with a series of collections of “jinkos” representing the assortment of products in a series of stores. The jinkos could be one of four colors and four shapes, and have one of four names (shown on the product label). Participants were asked to either judge the variety and organization of the assortment (the *variety orientation* group) or to indicate their satisfaction with the product assortment and the likelihood that they would shop at a store with that assortment (the *choice orientation* group). Participants examined collections of 16 jinkos on shelves (4 per shelf) which were either organized (by color, shape, and name) or not, and contained various levels of variability (accomplished by either holding one attribute constant, biasing the distribution toward certain values of a dimension, or allowing duplicates—there were 10 different types of variability examined).

In accord with expectations, they found that increasing the variability of jinkos increased the perceived variety (up to a point), and that organized assortments were judged less variable than random assortments (cf. Wasserman, Young, & Nolan, 2000). They also reported that variation in color was more salient than variation in shape, which in turn was more salient than variation in product name. Hoch et al.’s (1999) study provides a rich set of data on people’s explicit judgments of variability and how these perceptions influence consumer behavior. We, however, were interested in exploring somewhat different issues.

First, we sought to understand the *acquisition* of variability information across multiple dimensions. How effectively can people learn to attend to one dimension of variability while ignoring a second? Can irrelevant variability be ignored or will it continue to be distracting? This issue is particularly relevant in light of a proposal by Young, Ellefson, and Wasserman (in press). They suggested that variability discrimination can be accomplished by examining an activation map like that proposed in Wolfe’s (1994) theory of visual search. Wolfe suggested that an item receives bottom-up activation to the degree that it differs from nearby items. Young et al. suggested that the average activation level of items in this activation map would reflect the number of (nearby) differences that are present in the display and could thus support a judgment of variability. If people can direct their attention to only one dimension, then top-down activation would alter the activation map, and thus the perceived variability. We examined this possibility in Experiment 1.

Second, we sought direct evidence of the additive effects of variability when more than one dimension varied. Research on categorization has found that two items which differ across multiple dimensions are classified as “different” more accurately and quickly than items that differ in only one (e.g., Nickerson, 1967; Nosofsky, 1986; Shepard, 1964). If multiple dimensions of variability affect behavior in a manner similar to that observed in other studies, then arrays involving items that share two features (e.g., 16 blue trains) should be easier to classify as same than those that involve

items sharing only one feature (e.g., 16 different blue objects). Experiment 1 of the present study provided some evidence of this anticipated result; participants were able to rapidly redirect their attention toward relevant dimensions and away from irrelevant dimensions (cf. Gottwald & Garner, 1975; Stone, 1971). In Experiment 2, we investigated this issue more directly by using a task in which both dimensions remained relevant throughout (cf. Gottwald & Garner, 1975; Nickerson, 1967; Stone, 1971).

Third, we were interested in the extent to which Hoch et al.'s (1999) results would generalize to a task like that of Young and Wasserman (2001a). Hoch et al. examined assortments that all contained moderate to high variability, used a limited palette of shapes, colors, and names (four of each), told people that the task involved the perception of variety, and asked direct questions about perceived variability. Young and Wasserman examined assortments that spanned the full range of variability (from none to maximal), used a large palette of shapes (24), did not inform participants which aspect of the displays would be important, and used choice behavior and reaction times as indices of perceived variability. Replicating the basic effects of Hoch et al. in a decidedly different task would bolster their conclusions.

We focused on the two most salient dimensions of variability studied by Hoch et al. (1999), color and shape (see Fig. 1). Our experiments were designed to disclose people's classification of color and shape variability when both could vary within a collection. In Experiment 1, we found that participants' classifications of variability in a relevant dimension were affected by variability in an irrelevant dimension, but this effect dissipated with training. In Experiment 2, participants revealed a similarity structure consistent with the findings of the first experiment—arrays that differed in two abstract dimensions (shape and color variability) were judged less similar than arrays that differed in only one of these abstract dimensions. These findings have profound consequences for the utility of entropy as a measure of categorical variability.

Experiment 1

Method

Participants

Eighty-two students enrolled in an introductory psychology course at Southern Illinois University at Carbondale served as voluntary participants. They received course credit for their participation.

Visual stimuli and experimental design

Twenty-four highly distinguishable icons were chosen as the total shape pool. The icons were the same as those used in Young and Wasserman (2001a). Twenty colors were chosen as the total color pool; the color pool was initially chosen by using 0%, 50%, or 100% saturation of the red, blue, and green channels. This process produced an initial palette of 27 different colors (e.g., red, dark blue, black, orange, and grass

green) from which we removed colors that were largely indiscriminable from at least one of the other colors or did not exhibit sufficient contrast against our white background (e.g., yellow). This process resulted in our final palette of 20 discriminable colors.

There were four different stimulus classes, same color/same shape, same color/different shape, different color/same shape, and different color/different shape (examples of each are shown in Fig. 1). For any given display comprising same shapes, a single icon from the appropriate set was randomly chosen and was used to make up an array of 16 identical icons. For any given display comprising different shapes, 16 of the 24 icons were used with no repetitions. An analogous strategy was used for producing displays with same or different colors. Thus a same color/same shape display would involve 16 identical icon shapes of the same color (e.g., 16 lime green jugglers or 16 pink dice). The 16 colored icons were randomly distributed over 25 locations arrayed in a 5×5 grid. Thus, 16 of the 25 locations contained icons and 9 were blank. The size of the grid on the computer screen was 8.5×8.5 cm. Subjects sat approximately 60 cm from the screen.

Each display contained 16 items because earlier work (e.g., Young & Wasserman, 1997) used 16 items thus permitting easier comparison of the results of the current study to those of earlier studies. Our use of 16 items also will enable subsequent research to examine displays in which some of the items were different shapes and/or colors and some were identical.

Procedure

Training. The participants entered the experimental room and were seated at one of four identically configured Macintosh PowerPC G4 computers. The experimenter read the participants a set of general instructions and asked the participants to put on a pair of headphones connected to the computer (the headphones were used to present auditory feedback). In addition to the text explaining the mechanics of the experiment and exhorting the participants to make their choices as quickly as possible while still being accurate, the participants were read the following introductory paragraph:

You will be observing a series of displays and attempting to learn which response is correct for each display. You will make your best guess as to which of *two* responses is correct. You will then be provided feedback in the form of a pleasant tone (if correct) or an unpleasant tone (if incorrect). This information will assist you in improving your chances of being correct for subsequent displays. Your goal is to accurately predict the correct response for each display by the end of training.

No information was provided that could have directed the participant toward any particular aspect of the displays. At no point in the instructions were the words “same” or “different” used. By requiring participants to learn which aspects of the display were important, we could better examine the learning process and how attention to color and shape might change over time. This method has the added benefit that it can be used with non-linguistic participants (e.g., infants, pigeons, or baboons). Once each participant indicated an understanding of the procedure, the experimenter started each of the programs.

The experiment comprised 120 trials: 30 randomized blocks of 4 trials each. Each block comprised 1 trial each of same color/same shape, same color/different shape, different color/same shape, and different color/different shape.

Participants were assigned to one of two conditions, *shape variability* or *color variability*. Participants in the *shape variability* condition were trained to make one response for the displays involving identical shapes (same color/same shape, different color/same shape) and a second response for those involving different shapes (same color/different shape, different color/different shape). Participants in the *color variability* condition were trained to make one response for the displays involving identical colors (same color/same shape, same color/different shape) and a second response for those involving different colors (different color/same shape, different color/different shape). For half of the participants in each condition, the “1” key was correct for one category and the “3” key was correct for the other category; for the other half of the participants, the key assignments were reversed. The presentation of stimuli and the processing of responses were controlled by PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993).

Each trial began with illumination of the display area with a white field containing a black cross in the center. The participant controlled the onset of a trial; pressing the “2” key turned on the icon array as a picture on a white background. The participants then made a choice response (“1” or “3” key). A correct choice response was followed by a pleasant tone. An incorrect choice response was followed by an unpleasant tone and the stimulus remained on the display; the participant was not allowed to continue until the correct response was made. These correction trials were not scored in data analyses. The computer did not respond to an illegal key press (any response other than a “1” or “3”).

Results

Most of the participants mastered the discrimination by the end of training, but many did not. For scoring purposes, we divided the learning session into five 24-trial blocks. To avoid including participants who failed to learn, we dropped the 11 worst performers in each condition based on Block 5 performance (their mean Block 5 accuracies were 41% and 49% for the *shape variability* and *color variability* conditions, respectively). This criterion was used in lieu of a specific performance criterion (e.g., 70% accuracy) to avoid differential attrition. Among those who were deemed to have mastered the discrimination (30 participants in each condition), learning was strong; during the final block, these participants averaged 89% correct (84% in the *shape variability* condition and 94% in the *color variability* condition). The accuracy varied across conditions, but this effect was statistically significant only during the final block of training, $F(1, 58) = 7.88$, $MSE = .4319$. Although we did not specifically manipulate the relative salience of color and shape variability, our observation that variability in color tended to be easier to assess than variability in shape is consistent with the results of Hoch et al. (1999) and previous demonstrations of the high salience of color in visual search tasks (e.g., Pomplun et al., 2001; Theeuwes, 1991; Theeuwes, 1992).

A preliminary examination of the data revealed that the effects under investigation varied significantly as training progressed. As a result, we examined the data across blocks to get a clear picture of participant behavior. Fig. 2 shows the percentage of “different” responses for the two conditions broken down by array type and block. This figure reveals three pertinent results. First, early in training, adding variability in the irrelevant dimension (color variability in the shape condition, shape variability in the color condition) increased the percentage of “different” responses for both the displays comprising items that varied in the relevant dimension and displays comprising items that were identical in the relevant dimension. Second, the effect of irrelevant variability was much more evident early in training than it was late in training and largely disappeared by Block 3. Third, the effect of adding irrelevant variability appears to have a larger impact on responding to the displays involving items that vary in the relevant dimension than to those involving items that are constant on the relevant dimension.

To confirm these observations, we performed a $4 \times 2 \times 5$ repeated measures, mixed ANOVA of percentage of “different” responses as a function of Stimulus Class (same color/same shape, same color/different shape, different color/same shape, and different color/different shape), Condition (*shape variability vs. color variability*), and Block (1–5). A significance level of .05 was chosen for this and all subsequent tests. The analysis revealed a main effect of stimulus, $F(3, 174) = 110.87$,

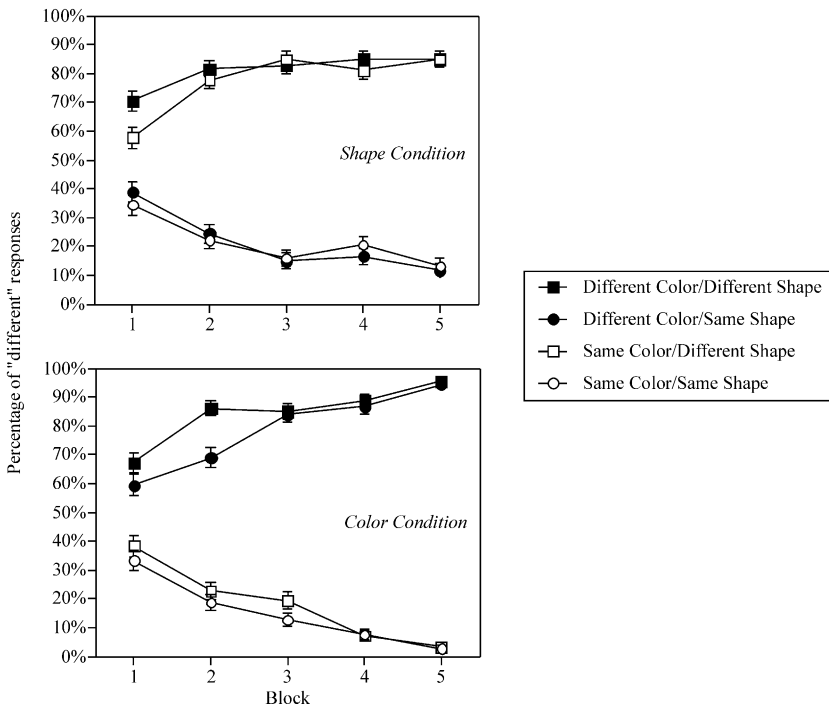


Fig. 2. Mean percentage of “different” responses in the two conditions examined in Experiment 1.

$MSE = 1.1304$, a Stimulus \times Block interaction, $F(12, 696) = 12.26$, $MSE = .1772$, and a Stimulus \times Block \times Condition interaction, $F(12, 696) = 21.07$, $MSE = .1772$.

We performed planned contrasts of the Stimulus \times Block \times Condition interaction to identify whether adding variability in the irrelevant dimension significantly increased the percentage of “different” responses in both conditions for each block. In the *shape variability* condition, the different shape/different color arrays produced significantly more different responses than the different shape/same color arrays only in Block 1 (13% more), and the same shape/different color arrays failed to produce significantly more “different” responses than the same shape/same color arrays at any time (all differences were less than 5%). In the *color variability* condition, the different shape/different color arrays produced significantly more “different” responses than the same shape/different color arrays in Blocks 1 (8% more) and 2 (17% more), but the different shape/same color arrays failed to produce significantly more “different” responses than the same shape/same color arrays (all differences were less than 7%).

To determine whether reaction times (RTs) would reveal any residual effects of the irrelevant variability (and to confirm the absence of speed-accuracy tradeoffs), we performed a $4 \times 2 \times 5$ repeated measures, mixed ANOVA of $\log(\text{RT})$ as a function of Stimulus Class (same color/same shape, same color/different shape, different color/same shape, and different color/different shape), Condition (*shape variability* vs. *color variability*), and Block (1–5). Only the main effects of stimulus, $F(3, 174) = 4.00$, $MSE = .4656$, and block, $F(4, 232) = 79.73$, $MSE = .2575$, were statistically significant. RTs decreased across blocks, as would be expected. More importantly, RTs for the individual stimuli differed systematically: different shape/different color (807 ms), same shape/same color (799 ms), different shape/same color (843 ms), and same shape/different color (820 ms). Thus, stimuli for which both dimensions varied or both dimensions were constant produced faster reaction times than stimuli for which one dimension varied and one was held constant (confirmed by a planned comparison, $\alpha = .05$). Because stimulus differences did not vary significantly across blocks (as revealed by non-significant interactions), irrelevant variability retarded responses throughout training.

Finally, we examined individual differences to determine if the results were a by-product of different participants showing sudden acquisition at different times rather than the gradual learning curves suggested by Fig. 2. Twenty-three participants did show very rapid learning at some point during the first two blocks of the experiment and achieved perfect performance during the last three blocks of training (11 participants in the *color variability* condition and 12 in the *shape variability* condition). These participants thus could not have contributed to the observed effects on choice behavior during the final three blocks. However, a repetition of the above RT analysis for these 23 participants in the final three blocks still revealed significant main effects of stimulus, $F(3, 63) = 3.05$, $MSE = .0112$, and block, $F(2, 42) = 3.48$, $MSE = .0156$, and a repeat of the planned comparison was also still statistically significant. Therefore, even when the participants had fully mastered the task (as evidenced by their perfect discriminative performance), they still showed residual effects of variability in the irrelevant dimension.

Discussion

Consistent with reports elsewhere that document the ability to shift attention away from irrelevant dimensions and toward relevant dimensions (e.g., Goldstone, 1994; Kruschke, 1992; Nosofsky, 1986; Sutherland & Mackintosh, 1971; Theeuwes, 1992), our participants quickly learned to respond to variability in the relevant dimension while discounting variability in an irrelevant dimension. The effect of irrelevant variability was detectable in the choice data only for a brief period early in training; reaction time differences, however, persisted through all five blocks of training.

The ease with which our participants were able to focus on the relevant dimension and discount the irrelevant dimension likely reflects our use of psychologically separable dimensions (Garner, 1970; Gottwald & Garner, 1975). Gottwald and Garner found that a *filtering task*, in which only one dimension is relevant to classification, was relatively easy when the dimensions were separable (e.g., color and size) but more difficult when the dimensions were integral (e.g., chroma and saturation). Their task required participants to sort items based on their sharing some visual feature (e.g., color) whereas ours required participants to sort items (through discriminative responding) based on their sharing a relational feature (homogeneity vs. heterogeneity). Although we anticipate that irrelevant variability would be harder to ignore if our dimensions of variability were integral rather than separable, the current study did not examine this issue.

Our data suggest that a measure of categorical variability like entropy is insufficient to capture the richness of our participants' variability discrimination behavior. Participants were able to learn to attend to entropy in one dimension (e.g., shape or color) while discounting another, but the effects of the irrelevant dimension remained. These results suggest that stimulus similarity, modulated by attention, must be incorporated into a complete model of variability discrimination.

Furthermore, our data parallel those of Hoch et al. (1999) showing that variability along multiple dimensions produces more perceived variety than variability along only a single dimension. Although the effect on choice behavior was transitory, adding redundant variability (adding more heterogeneity to different displays or more homogeneity to same arrays) produced shorter reaction times throughout the task. Although Hoch et al. asked their participants for explicit judgments of variety in collections of items that vary along multiple dimensions, our results demonstrate that even when you train participants to ignore one source of variability, they have a persistent difficulty doing so.

Experiment 2

In Experiment 1, we found that arrays involving items that varied along two dimensions (one relevant and one irrelevant) were easier to classify as “different” than those that varied along only the relevant dimension. But, this effect was relatively small. In Experiment 2, we reexamined the contrast between displays involving one or two dimensions of variability by using a task in which both dimensions

remained relevant throughout. This new task is analogous to the *condensation task* of Gottwald and Garner (1975) in which two dimensions must be attended in order to produce correct responses, a task that they documented to be considerably more difficult than their filtering task.

In our second experiment, one type of display (same shape/same color in the *same* condition; different shape/different color in the *different* condition) required one response whereas the other three types of displays required a second response. Making a correct response necessitated attention to both color and shape. We were interested in differences in performance for the three stimulus classes in the “other” category to determine whether extra variability (for the three ‘not all same’ displays in the *same* condition) or extra homogeneity (for the three ‘not all different’ displays in the *different* condition) would increase response accuracy. Measures of similarity like the Minkowski metric (Shepard, 1964) suggest that stimuli which differ in two features should be judged less similar than those that differ in only one. For example, in the *same* condition, the same color/same shape displays belong in one category whereas the other three display types belong in the other category; displays involving two differences (different color/different shape) should appear more different than those involving only one difference (different color/same shape and same color/different shape). Evidence of systematic effects of similarity will further reveal the role of multi-dimensional similarity in variability discrimination.

Experiment 2 also provides the opportunity to reexamine another question—is there something special about complete identity (i.e., same color and same shape) that makes it especially distinct? Young and Wasserman (2002) provided evidence that identity is special in that it was easy to discriminate displays comprising 16 identically shaped items from displays in which even one or two items were different from the others, but it was more difficult to discriminate displays comprising 16 differently shaped items from displays in which two or more of the items were identical. In the present experiment, the discrimination required in the *same* condition thus might be easier because the distinctive display (same color/same shape) is placed in a different category than the others whereas in the *different* condition there is no such benefit (same color/same shape displays are placed in the same category as same color/different shape and different color/same shape displays).

Method

Participants

Fifty-six students enrolled in an introductory psychology course at Southern Illinois University at Carbondale served as voluntary participants. They received course credit for their participation.

Procedure

Visual stimuli. The stimuli were identical to those used in Experiment 1.

Training. The experiment comprised 120 trials: 30 randomized blocks of 4 trials each. Each block comprised 1 trial each of same color/same shape, same color/different shape, different color/same shape, and different color/different shape.

Participants were assigned to one of two conditions, *same* or *different*. Participants in the *same* condition were trained to make one response for the displays involving identical shapes and colors (same color/same shape) and a second response for the other stimulus types. Participants in the *different* condition were trained to make one response for the displays involving different colors and icons (different color/different shape) and a second response for the other stimulus types. For half of the participants in each condition, the “1” key was correct for one category and the “3” key was correct for the other category; for the other half of the participants, the key assignments were reversed.

Results and discussion

Most of the participants mastered the discrimination by the end of training, but some did not. For scoring purposes, we again divided the learning session into five 24-trial blocks. To avoid including participants who failed to learn, we dropped the four worst performers in each condition based on Block 5 performance (their mean Block 5 accuracies were 61 and 52% for the *same* and *different* conditions, respectively). Among those who were deemed to have mastered the discrimination (24 participants in each condition), learning was strong; during the final block, these participants averaged 87% correct (90% in the *same* condition and 84% in the *different* condition). Only 4 of the 48 participants showed perfect discrimination during the final three blocks of the study (recall that 23 of the 60 participants in Experiment 1 had done so), and all of these were in the *same* condition. Apparently, making a unique response to items that are identical in both shape and color was easier than making a unique response to items that differed in both shape and color. The special status of sameness has been observed as a byproduct of studies of categorization (e.g., Evans & Smith, 1988; Kemler, 1979, 1983) and directly observed in two recent studies of variability discrimination involving various mixtures of same and different shapes (Young & Wasserman, 2001a, 2002).

Fig. 3 reveals that our participants had a much more difficult time choosing the correct key for the minority category in the *same* condition (same shape/same color) and in the *different* condition (different shape/different color). This result was almost certainly due to our desire to hold the frequency of each display type constant (thus paralleling Experiment 1) which made for a disparity in the frequency of each category (25% for the minority category vs. 75% for the “other” category) thus creating a response bias toward the category that occurred more often. Be that as it may, our interest was in the relative accuracy of the three display types in the “other” category thus making performance on the minority category irrelevant for our purposes (except as a reflection of the unique character of perfect identity).

In the *same* condition, the displays that varied in two dimensions (different shape/different color) were classified more accurately than the displays that varied in only one dimension (different shape/same color and same shape/different color). In the *different* condition, the displays that matched in both dimensions (same shape/same color) were classified more accurately than the displays that matched in only one dimension (different shape/same color and same shape/different color). In contrast

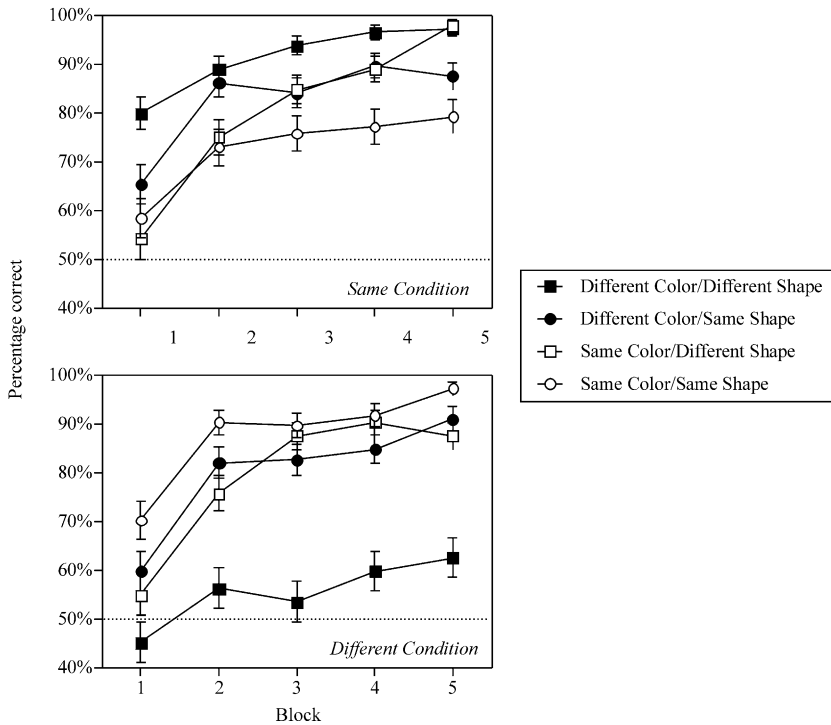


Fig. 3. Mean accuracy in the two conditions examined in Experiment 2. Accuracy was used as the dependent variable (rather than percentage of “different” responses) to make it easier to compare the accuracy of responding to the majority and minority categories and of the two conditions.

to Experiment 1, these effects did not tend to dissipate as training progressed (compare Figs. 2 and 3).

To confirm these observations, we performed a $4 \times 2 \times 5$ repeated measures, mixed ANOVA of percentage of “different” responses as a function of Stimulus Class (same color/same shape, same color/different shape, different color/same shape, and different color/different shape), Condition (*same* vs. *different*) and Block (1–5). The analysis revealed a main effect of stimulus, $F(3, 138) = 3.63$, $MSE = .5831$, a main effect of block, $F(4, 184) = 55.03$, $MSE = .2291$, a Stimulus \times Block interaction, $F(12, 552) = 2.99$, $MSE = .1605$, and a Stimulus \times Condition interaction, $F(3, 138) = 28.28$, $MSE = .5831$. Because we were interested in differences in performance for the three stimulus classes in the “other” category for each condition, we performed planned contrasts ($\alpha = .05$) of the Stimulus \times Condition interaction to determine whether extra variability (in the *same* condition) or extra homogeneity (in the *different* condition) would increase response accuracy.

In the *same* condition, the different shape/different color arrays produced significantly more accurate responses (91%) than both the different shape/same color (80%) and same shape/different color arrays (83%). In the *different* condition, the same

shape/same color arrays produced significantly more accurate responses (88%) than the different shape/same color (79%) and same shape/different color arrays (80%), but the latter difference was only marginally significant ($p = .055$).

To determine whether reaction times (RTs) would reveal any additional effects (and to confirm the absence of speed-accuracy tradeoffs), we performed a $4 \times 2 \times 5$ repeated measures, mixed ANOVA of $\log(\text{RT})$ as a function of Stimulus Class (same color/same shape, same color/different shape, different color/same shape, and different color/different shape), Condition (*same vs. different*), and Block (1–5). The analysis revealed main effects of stimulus, $F(3, 138) = 5.82$, $MSE = .0595$, and block, $F(4, 184) = 74.82$, $MSE = .1806$, and a Condition \times Stimulus interaction, $F(3, 138) = 11.99$, $MSE = .0595$. RTs again decreased across blocks as expected. The Condition \times Stimulus interaction required closer inspection.

For the reasons mentioned earlier, we were interested in any differences among the stimulus types in the “other” category in each condition (examined via planned comparisons, $\alpha = .05$). In the *same* condition, the different shape/different color stimuli produced a much faster mean RT (695 ms), than the different shape/same color (788 ms), and same shape/different color (782 ms) stimuli. In the *different* condition, the same shape/same color stimuli produced a much faster mean RT (665 ms), than the different shape/same color (758 ms), and same shape/different color (724 ms) stimuli. Thus, stimuli in the “other” category for which both dimensions varied (in the *same* condition) or were constant (in the *different* condition) produced much faster reaction times than stimuli for which one dimension varied and one was held constant. These stimulus differences did not vary significantly across blocks (as revealed by non-significant block interactions) thus paralleling the accuracy data.

Our results confirmed those of Experiment 1—identifying display variability is more than just identifying whether items are identical or not. Degree of similarity affects perceived variability, although the higher overall accuracy for the *same* condition relative to the *different* condition does suggest that perfect identity is particularly salient.

General discussion

Our data clearly document people’s ability to discriminate variability in a collection of items when that variability is defined across multiple dimensions. This result extends Young and Wasserman’s (2001a) observation that people can identify the degree of variability along a single dimension (shape) when that variability was altered by changing the relative frequency of the items (with 16 identical and 16 different items at either end of a continuum). Furthermore, the results of Experiment 1 clearly indicate that attention can be rapidly shifted away from a salient but irrelevant dimension thus documenting the effect of top-down processing on perceived variability. Despite this shift in attention, irrelevant variability continued to affect response latencies throughout the training period. Consequently, although entropy as a measure of categorical variability involving a single dimension nicely captured the behavior of Young and Wasserman’s (2001a) participants, it fails to account

for variability across multiple dimensions. A multivariate extension of entropy would address the behavior we observed in Experiment 2, but the results of Experiment 1 suggest that any such measure would need to incorporate the psychological process of shifting attention among the dimensions.

Our findings are consistent with and extend those of Hoch et al. (1999). Recall that their participants were asked to make explicit judgments of the variety of products arrayed on store shelves or to indicate the likelihood that they would shop at a store with a given collection of products. Their task differed in many ways from ours, but both studies demonstrate the influence of additional variability or homogeneity across dimensions even when participants are trained to ignore a source of variability. We do not know how actively or consciously people try to judge product variety on store shelves while shopping, but the results of the present study suggest that people do not have complete control over their ability to ignore irrelevant variability. Thus, marketers are well-served by adding variability in product dimensions (e.g., label color, bottle shape) that should have nothing to do with true product differences.

Given the significance of categorical variability to various choice tasks, we believe it is important to better understand its effects. This information may be used to establish the generality of established cognitive processes (Nosofsky, 1986; Shepard, 1964; Wolfe, 1994), to develop better marketing strategies (Broniarczyk et al., 1998; Hoch et al., 1999), or to examine the development of relational concepts in human (Young & Wasserman, 2001a) and non-human species (Wasserman et al., 2001; Young & Wasserman, 1997, 2001b). Although many believe that “variety is the spice of life,” accurately identifying that variety is another matter altogether.

References

- Baumol, W. J., & Ide, E. A. (1956). Variety in retailing. *Management Science*, 3, 93–101.
- Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. New York: McGraw-Hill.
- Broniarczyk, S. M., Hoyer, W. D., & McAlister, L. (1998). Consumers' perceptions of the assortment offered in a grocery category: The impact of item reduction. *Journal of Marketing Research*, 35, 166–176.
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments & Computers*, 25, 257–271.
- Evans, P. M., & Smith, L. B. (1988). The development of identity as a privileged relation in classification: When very similar is just not similar enough. *Cognitive Development*, 3, 265–284.
- Garner, W. R. (1970). The stimulus in information processing. *American Psychologist*, 25, 350–358.
- Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, 123, 178–200.
- Gottwald, R. L., & Garner, W. R. (1975). Filtering and condensation tasks with integral and separable dimensions. *Perception and Psychophysics*, 18, 26–28.
- Hoch, S. J., Bradlow, E. T., & Wansink, B. (1999). The variety of an assortment. *Marketing Science*, 18, 527–546.
- Kemler, D. G. (1979). Accessing similarity and dimensional relations: Effects of integrality and separability on the discovery of complex concepts. *Journal of Experimental Psychology: General*, 108, 133–150.

- Kemler, D. G. (1983). Exploring and reexploring issues of integrality, perceptual sensitivity, and dimensional salience. *Journal of Experimental Child Psychology*, *36*, 365–379.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.
- Munsinger, H. (1966). Multivariate analysis of preference for variability. *Journal of Experimental Psychology*, *71*, 889–895.
- Munsinger, H., & Kessen, W. (1966). Preference and recall of stimulus variability. *Journal of Experimental Psychology*, *72*, 311–312.
- Nickerson, R. S. (1967). “Same”–“Different” response times with multi-attribute stimulus differences. *Perceptual and Motor Skills*, *24*, 543–554.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.
- Pomplun, M., Sichelschmidt, L., Wagner, K., Clermont, T., Rickheit, G., & Ritter, H. (2001). Comparative visual search: A difference that makes a difference. *Cognitive Science*, *25*, 3–36.
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.
- Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*, 54–87.
- Stone, G. C. (1971). Response latencies in visual search involving redundant or irrelevant information. *Perception & Psychophysics*, *9*, 9–14.
- Sutherland, N. S., & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning*. New York: Academic Press.
- Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception & Psychophysics*, *50*, 184–193.
- Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, *51*, 599–606.
- Wasserman, E. A., Fagot, J., & Young, M. E. (2001). Same–different conceptualization by baboons (*Papio papio*). *Journal of Comparative Psychology*, *115*, 42–52.
- Wasserman, E. A., Young, M. E., & Nolan, B. C. (2000). Display variability and spatial organization as contributors to the pigeon’s discrimination of complex visual stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *26*, 133–143.
- Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202–238.
- Young, M. E., Ellefson, M. R., Wasserman, E. A. (in press). Toward a theory of variability discrimination: Finding differences. *Behavioural Processes*.
- Young, M. E., & Wasserman, E. A. (1997). Entropy detection by pigeons: Response to mixed visual displays after same–different discrimination training. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*, 157–170.
- Young, M. E., & Wasserman, E. A. (2001a). Entropy and variability discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 278–293.
- Young, M. E., & Wasserman, E. A. (2001b). Evidence for a conceptual account of same–different discrimination learning in the pigeon. *Psychonomic Bulletin & Review*, *8*, 677–684.
- Young, M. E., & Wasserman, E. A. (2002). Detecting variety: What’s so special about sameness. *Journal of Experimental Psychology: General*, *131*, 131–143.