

# A theory of variability discrimination: Finding differences

**MICHAEL E. YOUNG**

*Southern Illinois University, Carbondale, Illinois*

**EDWARD A. WASSERMAN**

*University of Iowa, Iowa City, Iowa*

AND

**MICHELLE R. ELLEFSON**

*University of Pittsburgh, Pittsburgh, Pennsylvania*

Visual variability discrimination requires an observer to categorize collections of items on the basis of the variability in the collection; such discriminations may be vital to the adaptive actions of both humans and other animals. We present a theory of visual variability discrimination that aggregates localized differences between nearby items, and we compare this finding differences model with a previously proposed positional entropy model across several data sets involving both people and pigeons. We supplement those previously published data sets with four new experiments, three of which involve arrays comprising items entailing systematic, quantitative differences. Although both theories provide strong and similar fits of the published data sets, only the finding differences model is applicable to investigations involving quantitative item differences, providing excellent fits in these new experiments.

---

People judge variety in a wide range of everyday tasks. We judge the variety of products on a store shelf, racial diversity in our workplace, and restaurant fare in our community. Psychological scientists have found that people exhibit preferences for certain levels of variety (Berlyne, 1960; Munsinger, 1966; Munsinger & Kessen, 1966), and marketing researchers have found 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). As well, nonhuman animals are sensitive to variability when they forage for food (Caraco, 1981; Caraco, Martindale, & Whittam, 1980) or choose mates (McGregor, Krebs, & Perrins, 1981; Searcy, 1984). Yet, despite the clear importance of variability discrimination to adaptive behavior, only recently have researchers begun to explore how we and other species make such discriminations.

Improving our understanding of variability discrimination may change people's behavior. A grocer who knows that customers' shopping preferences are affected by perceived variety can organize store shelves or choose particular flavor combinations to increase the apparent variety of merchandise. The director of an art museum has multiple dimensions of variability to manage (e.g., artist, style, color, and type) and might choose the works to display and their spatial organization in order to maximize perceived variety and thereby avoid ennui in patrons.

Given that overeating is more likely with greater variety in food choice (Kahn & Wansink, 2004), a school cafeteria might opt to increase the variety of fruit and vegetable options and to reduce the variety of dessert options. Perceived variability may also underlie judgments of creative behavior (e.g., paintings, music, or fashion), pleasantness, and relevance (highly variable stimuli contain more information). For example, judgments of creativity may be highest for moderate degrees of variability, and the optimal level of variability may increase as a function of experience (cf. Munsinger & Kessen, 1966).

Historically, psychological scientists have focused on a species' ability to filter out variability in both stimuli and responses. The goal of learning was believed to be the isolation of environmental regularities (variability is noise) and the production of consistently reinforced behavior. Some theorists, however, have emphasized the importance of both recognizing and producing variability. For example, Skinner (e.g., 1981) stressed the involvement of response variability in selection by consequences. Other theorists (Ashby & Gott, 1988; Fried & Holyoak, 1984) have stressed the importance of perceived variability to optimal categorization.

In a recent review, Neuringer (2004) examined the *production* of variability; this work focused on the actor. Our own line of research has examined the observer's *discrimination* of variability, variability that may be inher-

---

M. E. Young, meyoung@siu.edu

---

ent in the environment or produced by other organisms. Successfully determining the amount of variability in a set of stimuli (including, potentially, observed behavior) can improve an organism's foraging, categorization, judgments of creativity, and decision making. Interestingly, the ability to discriminate variability is present in species as diverse as pigeons, baboons, and people (for a review, see Wasserman, Young, & Cook, 2004).

Our discrimination tasks require organisms to extract a statistical property, variability, from collections of items that differ along one or more categorical or continuous dimensions and to actively ignore other stimulus attributes, such as average size, color, and shape. Recent work by Ariely (2001) and Chong and Treisman (2003, 2005) required humans to extract a different statistical property, average, from collections of circles that varied in size. Subjects judged the average size of simultaneously presented 4- to 16-item collections of circles. These collections varied in the average size of the circles in the display and the variability of those sizes. Subjects could rapidly estimate the average size of the circles in a display, but their behavior was systematically biased by the variability of sizes (more variability produced less accurate judgments of the average). These results document the ability of humans to rapidly assess a statistical property, average, of a collection of items that vary along a single continuous dimension (size) while actively ignoring another property, variability (see Parkes, Lund, Angelucci, Solomon, & Morgan, 2001, for an analogous effect).

Figure 1 shows two examples of typical displays used in the research of Young, Wasserman, and colleagues. These displays exhibit the minimal and maximal degrees of variability that are possible in collections of 16 items. Not only do pigeons, baboons, and people all respond discriminatively to these two endpoints of the variability dimension (i.e., they learn to choose one response for *same* arrays and a second response for *different* arrays), but also their responding to intermediate values of variability is a monotonic function of display variety.

Young and Wasserman (1997) found that the information theoretic measure of categorical variability, *entropy*, nicely captures the functional relationship between stimulus variability and discriminative responding. To quantify entropy (measured as bits of information required for minimal coding), Shannon and Weaver (1949) used the following equation:

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

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.

Although entropy provided an excellent first fit of pigeons' discriminative performance, a modulating effect of item location was discovered in a later study involving a positional variant of entropy (Wasserman, Young, & Nolan, 2000). When identical items were clustered together (left side of Figure 2), pigeons were slightly more likely to choose the *same* report option than when identical items were more widely distributed (right side of Figure 2). Wasserman et al. (2000) proposed a measure of positional entropy, in which individual entropy scores were calculated for all possible contiguous subsets (e.g., all nine possible groups of contiguous items in a  $2 \times 2$  arrangement) and later averaged. This positional variant of entropy accounted for the effect of the organization of same and different items in mixtures like those shown in Figure 2.

In more recent work, Young and Ellefson (2003) found that the *degree* of difference is important to discrimination performance, not just whether two items are identical or nonidentical. In nearly every published study, the

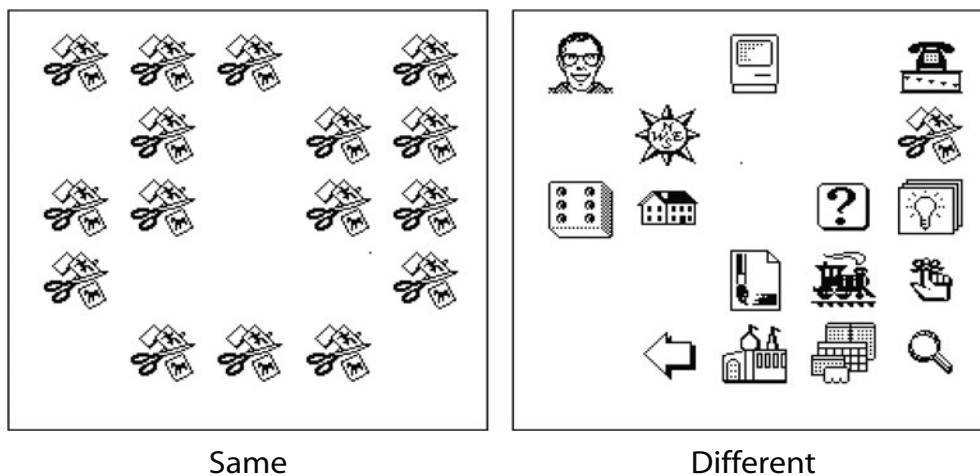
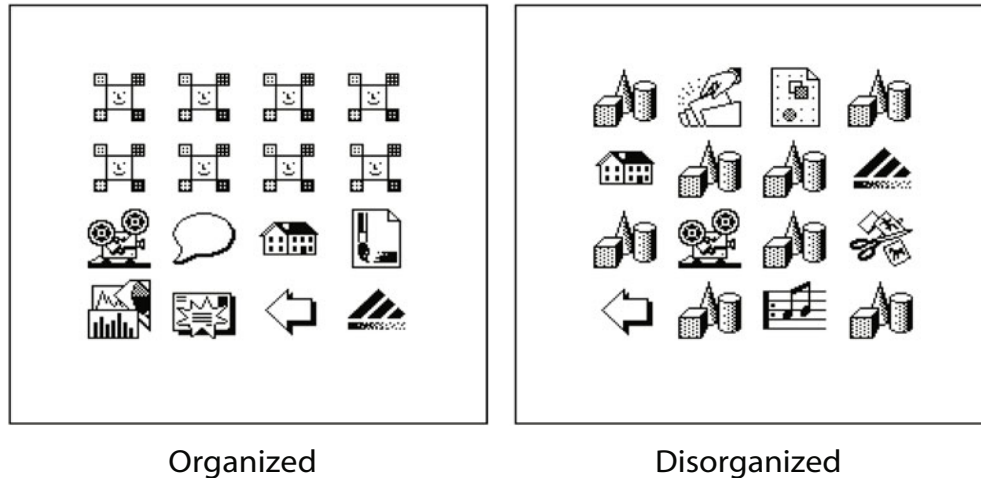


Figure 1. Examples of the 16-icon arrays used during training in many of the studies by Young, Wasserman, and colleagues. These arrays consisted of 16 icons chosen from a set of 24 to create *same* and *different* arrays with these 16 chosen icons randomly located in a  $5 \times 5$  grid.



**Figure 2.** Examples of the clustered displays (identical items clustered together) and distributed displays (identical items widely distributed) used in Wasserman, Young, and Nolan (2000). These arrays involved eight identical items and eight different items.

multiple-stimulus displays have involved variability in shape (brightness and approximate size have been held constant). Young and Ellefson used displays in which the items varied in shape and color (the same set of shapes shown in Figure 1, but involving a palette of 20 distinct colors). After receiving training that required discriminative responding to *same* and *different* arrays, college student subjects were tested with displays in which one of the dimensions was held constant (e.g., a color-constant display would involve 16 different items of the same color). Although people could be trained to largely ignore variability in a dimension such as color or shape (Experiment 2 of that study), reaction times (RTs) continued to be affected by variability in the irrelevant dimension, thus implicating multidimensional similarity in variability discrimination.

These effects of similarity on discrimination performance pose problems for the original entropy account. Entropy, as calculated by Equation 1, requires that items are either in the same category or not; there is no provision for *degree* of category membership.

#### **A NEW THEORETICAL ACCOUNT: FINDING DIFFERENCES**

Although we could have continued to search for other possible extensions of entropy (e.g., by positing a multidimensional, positional entropy), we opted to approach the problem from an altogether different perspective. Perhaps visual variability discrimination leverages stimulus information that has proven to be useful in performing other discrimination tasks. Rather than proposing a new mechanism that is dedicated specifically to detecting variability, we considered the possibility that a redescription of the stimulus that has been used in the service of another common task, *visual search*, might provide the foundation for variability discrimination.

In a typical visual search task, a rich array of items is presented: One item is the target of the search, and the rest are distractors. For example, the subject might receive instructions to find a T among Ls or to find a red X among red Os and blue Xs (e.g., Treisman & Gelade, 1980). The similarity of the target to the distractors strongly affects the efficiency of the search: The more similar the target is to the distractors, the harder it is to find (Duncan & Humphreys, 1989).

Wolfe (1994) proposed that each item in the array is *active* to various degrees. An item's level of activation establishes the degree to which it stands out from its neighbors, thus determining the difficulty of finding it during a search task. In Wolfe's model, Guided Search 2.0 (GS2), each item receives two sources of activation: bottom up and top down. Bottom-up activation is produced by the array items themselves; those items that are different from their neighbors will be more active (i.e., stand out). Top-down activation is produced by expectations usually induced by task demands or instructions. For example, if an observer is reinforced for finding red items or is told to search for a red item, color will receive extra attention; greater attention to color (a top-down influence) will make bottom-up color differences more salient.

In GS2 (Wolfe, 1994), the bottom-up sources of activation are captured in an activation map, which includes an activation score for each item in the display. Those items that differ from their neighbors receive greater activation than do those items that are similar to their neighbors. This redescription of the original display thereby captures *local differences*. For a visual search task, the location of salient differences helps to direct attention to those items that stand out from the background.

In our explorations of this theory, we noted that the *average* activation in GS2's activation map provides a plausible index of the degree of variability in a collection of visual items. If the map contains many highly active items, the

display must comprise items that are different from their neighbors (i.e., it involves high variability). On the other hand, if the map contains few active items, the display must comprise items that are very similar to their neighbors (i.e., it involves low variability). We were attracted to the parsimony of using Wolfe's (1994) activation map for both visual search (where the particular locations of differences are important) and variability discrimination (where only the average level of activation is important). No additional *entropy detection* mechanism need be posited, and the broader applicability of GS2 is established—an important step toward theoretical parsimony.

This approach has three other potential benefits as a theoretical account of performance in visual variability discrimination tasks. First, Wolfe's (1994) activation map computes activation on the basis of local differences, thus providing a possible account of the effect of organization on variability discrimination (see Figure 2). Second, the similarity of an item to its neighbors is computed in producing the map, thereby providing a mechanism for the degree of similarity to affect discrimination performance. Third, the role of top-down processes in GS2 allows the modeling of differential dimensional salience to be produced by perceptual or task demands (e.g., if subjects were reinforced for responding to color variability, but not to shape variability, differences in color should be emphasized, thereby producing the greatest activation for those items that differ in color from their neighbors; Young & Ellefson, 2003).

Given the promise of GS2's activation map as a basis for variability discrimination, we have developed a new model that can capture these functional relationships. Here, we explore whether the model can account for variability discrimination performance in pigeons and people, and we compare this *finding differences model* with the *positional entropy model* offered by Wasserman et al. (2000).

### Finding Differences Model

The activation maps used in our simulations were either  $4 \times 4$  or  $5 \times 5$  matrices, each value of which represented the activation (i.e., *differentness*) of a particular item in the original display. The activation map represents a computational redescription of the display that provides significant utility for modeling the functional relationships between the properties of the stimulus and the behavior of the observer.

The activation of an item is a function of two factors: (1) the degree of difference between it and each of its neighbors and (2) the distance between it and each of its neighbors. Wolfe (1994) posited that nearby differences should be more influential than distal differences. We agree. He incorporated this hypothesis by computing an item's activation as a function of feature differences between it and its "neighbors," where the neighborhood was prespecified (e.g., only adjacent items). Rather than define an item's neighborhood as only those items that are adjacent to or within some prespecified distance, we *scaled* the degree of difference between items as a function of their Euclidean distance in space. Thus, we posited that the organism would be more influenced by nearby dif-

ferences than by distal differences and that this influence would be graded (as a function of Euclidean distance), rather than all-or-none (in the neighborhood or not). Other than this change, the activation map that we have used in the finding differences model is isomorphic to that presented by Wolfe.

To measure the degree of difference between items  $x$  and  $y$ , we used an extended version of the Minkowski power model:

$$d = \left( \sum_i |x_i - y_i|^r \right)^{q/r}, \quad (2)$$

where  $x$  and  $y$  are vectors of features,  $d$  is the degree of difference between the vectors, and  $r$  and  $q$  are free parameters that determine the shape of the difference function. We adopted a city block metric ( $r = 1$ ), because our items were composed of separable dimensions (Shepard, 1991). Large values of  $q$  would accentuate large differences between array items (e.g., with a  $q$  of 1, two stimuli that differ in four dimensions would be twice as different as two items that differ in only two dimensions, but four times as different with a  $q$  of 2). In the absence of variability experiments specifically identifying the best-fitting  $q$  value, we set  $q$  to 1 for simplicity. Thus, Equation 2 reduces to

$$d = \sum_i |x_i - y_i|. \quad (3)$$

For our tasks, the difference between array items  $x$  and  $y$ , was, therefore, the sum of their feature differences. In our simulations, we assumed that differences in size, brightness, shape, or orientation each produced a difference score of 1.0. For example, when comparing a large, black square with a vertical bar with a small, black circle with a vertical bar (see Figure 3 for examples), the resulting difference score,  $d$ , is 2.0. These two stimuli thus should be judged to be more similar than that same large, black square with a vertical bar is to a small, white circle with a horizontal bar (difference score,  $d$ , of 4.0). Equation 3 also incorporates degrees of difference within each dimension (a white circle should be more similar to a gray than to a black circle).

To incorporate differences in salience among the dimensions (due to bottom-up or top-down influences) requires a small modification of Equation 3 (Nosofsky, 1986):

$$d = \sum_i s_i |x_i - y_i|, \quad (4)$$

in which  $s_i$  is the salience of dimension  $i$ . Differences in salience may be due to the prior reinforcement contingencies, the perceptual abilities of the observer, or the instructions provided to the observer (in human studies).

When these difference scores are aggregated for an item and scaled by the Euclidean distance between the item and its neighbors, we get the activation score for a particular item,  $x$ , in the display:

$$a_x = \sum_{y \in N_x} \left( e^{-\text{dist}(l_x, l_y)} \sum_i s_i |x_i - y_i| \right), \quad (5)$$

where  $l_x$  and  $l_y$  represent the spatial locations of items  $x$  and  $y$ ,  $\text{dist}(l_x, l_y)$  represents the Euclidean distance between



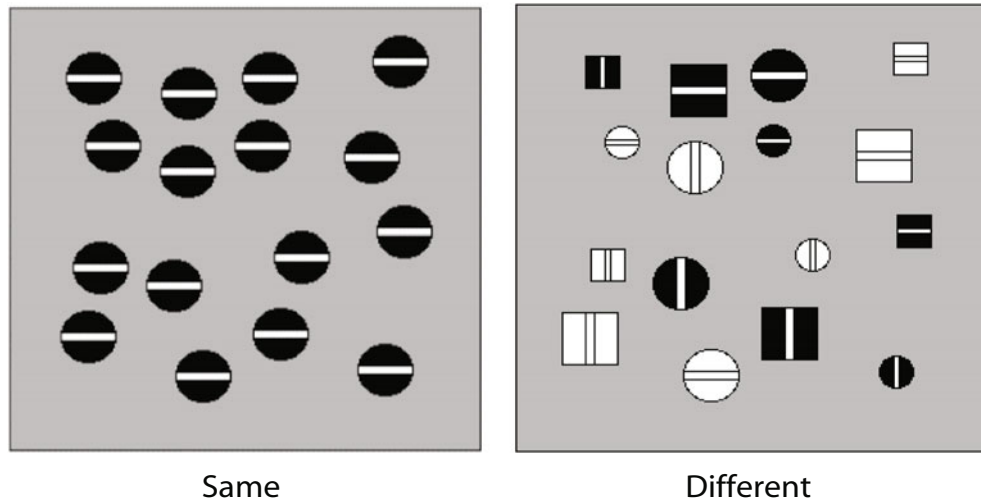


Figure 3. Examples of the *same* and *different* displays used in Young, Castro, and Wasserman (2002). These arrays involved up to 16 possible items created from all possible combinations of two levels of brightness, size, orientation, and shape ( $2 \times 2 \times 2 \times 2 = 16$ ).

these locations (i.e., the square root of the sum squared positional differences along the vertical and horizontal axes),  $c$  represents the effect of spatial distance (a large  $c$  value produces less influence of distant neighbors; see Figure 4 for an example),  $N_x$  represents the neighborhood of  $x$  (in our simulations, the neighborhood incorporated every item in the display other than  $x$ ), and  $a_x$  represents the activation of item  $x$  in the display.

After the original display was redescribed as an activation map (see Figure 5 for an illustrative stimulus array and its corresponding activation map), we simply averaged across the  $a_x$  values in the map to produce an *aggregate* measure ( $\bar{a}$ ) of differences present in the array (note that in arrays comprising fewer items, these items will have fewer neighbors and, thus, generate lower activa-

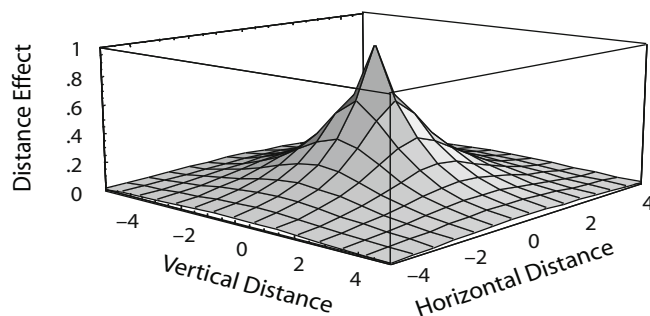


Figure 4. The effect of spatial distance on similarity when  $c$  is .65. For an icon located in the center of this figure, the effect of a difference between it and a distant icon decreases as the distance between the icons increases. Distance was measured on the basis of row and column position. Thus, adjacent items in the same row or column are 1.00 unit apart, and diagonally adjacent items are 1.40 units apart. The greatest possible distance in our  $5 \times 5$  arrays was between items at opposite corners, 5.70 units. We used positive and negative distance values in the figure to designate left/right and above/below.

tions). This measure could then be mapped to a response through the reinforcement contingencies of the task. This measure functions as the predictor of performance.

Finally, the probability that an observer judges the display to be different is given by

$$P(R = \text{different}) = \frac{u}{1 + e^{-m(\bar{a}-b)}} + l, \quad (6)$$

in which  $m$  and  $b$  are free parameters that determine the sharpness of the response function and the point of subjective equality, respectively, and  $l$  and  $u$  alter the asymptotes of the sigmoid ( $l$  is the lower asymptote, and  $u$  is the upper asymptote minus  $l$ ). This same equation will be used later to map positional entropy to the probability of a response, so it is not a distinctive feature of the new model. We needed the four-parameter version of the logistic primarily to capture response biases in the pigeon data.

It should be noted that ours is not the first model that has tried to account for the effects of mixture, organization, and similarity in a variability discrimination task. Hoch et al. (1999) examined people's judgments of product variety on store shelves and offered a model that bears a family resemblance to the finding differences model. Their approach was quite different from ours, however, since it used a model with many more free parameters to empirically determine (1) the shape of the generalization function, (2) the class of functions that might model the effect of spatial distance, and (3) the utility of assuming uniform or varied saliences across dimensions. Given the limited data from their single study, Hoch et al. could make only broad conclusions concerning these functional relationships. Our preference was to base many of our modeling decisions on prior research, thus opting for a city block distance metric, exponential effects of spatial distance, and varied dimensional salience (Hoch et al., 1999; Nosofsky, 1986). Furthermore, Hoch et al. used a dichotomous measure of spatial distance (contiguous vs.

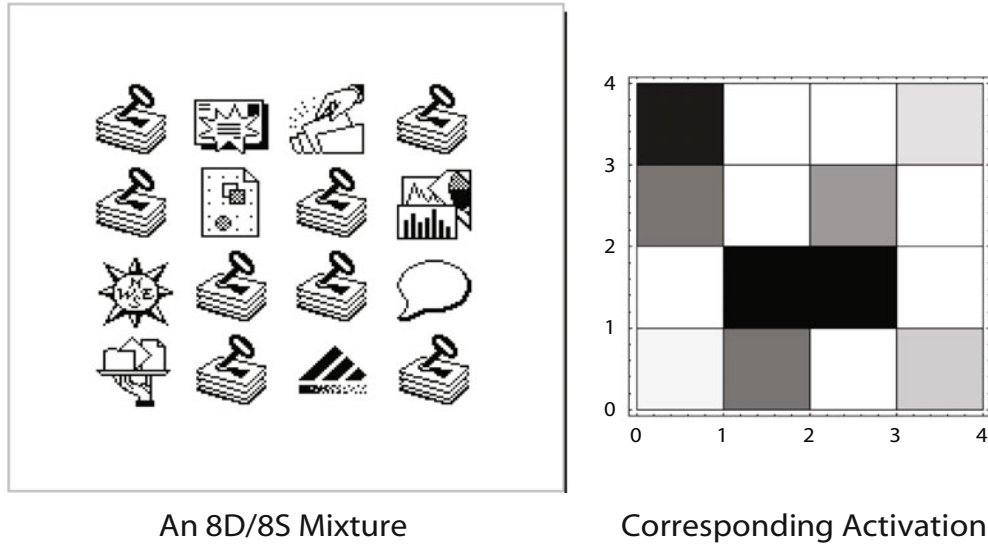


Figure 5. An example array (“8D/8S” designates a display with eight different items and eight same items) and its corresponding activation map, in which white designates high activation, black designates low activation, and shades of gray designate intermediate levels of activation.

not), but they did acknowledge that Euclidean distances should be explored as an alternative. Although their model is not a superset of ours, their model and the finding differences model share a number of features; each model could be generalized to encompass the other.

### FITTING THE DATA

After a decade of conducting experiments across a variety of situations and species, we sought to develop a theoretical account that captures most, if not all, of the effects that we have observed on variability discrimination in people and pigeons. In this section, we will evaluate our proposed finding differences model against the seminal studies on variability discrimination and will compare it with an extant model. In comparing models, it is important to consider two factors: (1) the fit of the model to the data and (2) the complexity of the model (Myung, 2000; Pitt, Myung, & Zhang, 2002).

The basic entropy formula contains no free parameters. However, in order to fit the data disclosing an influence of spatial organization on discrimination behavior (Wasserman et al., 2000), a “window of attention” had to be introduced with two free parameters, the *height* and *width* of the window, to yield the positional entropy model. The finding differences model captures the effect of localized differences through the use of a single parameter,  $c$ , that scales the influence of distance on similarity.

The finding differences model also includes a set of additional free parameters for the salience of each dimension. Salience is relevant only for experiments involving multiple dimensions of variability, so it is not used (i.e., it is held constant) in most of our simulations. A new, multidimensional entropy formula would be needed for experiments involving several dimensions of variability and, hence, would require its own set of unspecified pa-

rameters. Equation 6 represents the mapping of the extracted variability to an actual response and is common to both models.

To optimally compare the complexity of two models, it would be necessary not only to compare the number of free parameters, but also to compare the complexity of the functions that each parameter could generate (e.g., their minimum description length, or MDL; Myung, 2000). Our simulations, however, clearly revealed numerous situations in which the positional entropy model was simply unable to account for performance by people and pigeons. Rather than complicating the presentation by providing specious comparisons of the two models, we will present only the results of our fits of the finding differences model to demonstrate both its strengths and its weaknesses. Detailed results of likelihood ratio analyses comparing the two models are available from the first author.

In the present examination of the utility of the finding differences model, we will focus on fitting the model to a series of data sets obtained in several separate studies of variability discrimination by pigeons and people, as well as to new data sets presented here. Six results will be examined: (1) As display variability increases, pigeons and people are more likely to make *different* report responses (Young & Wasserman, 1997, 2002); (2) when identical items are clustered together (compare the left and right columns in Figure 6), pigeons and people may report the display to be less variable than when the items are distributed (Wasserman et al., 2000, and the new data presented here); (3) when similar but not identical items are clustered together, people report the display to be less variable than when the items are distributed, and this effect is magnified when the items are very similar (new data); (4) discriminative performance decreases as the number of items decreases, but only for *different* displays (Young & Wasserman, 2001a; Young, Wasserman, & Garner, 1997);

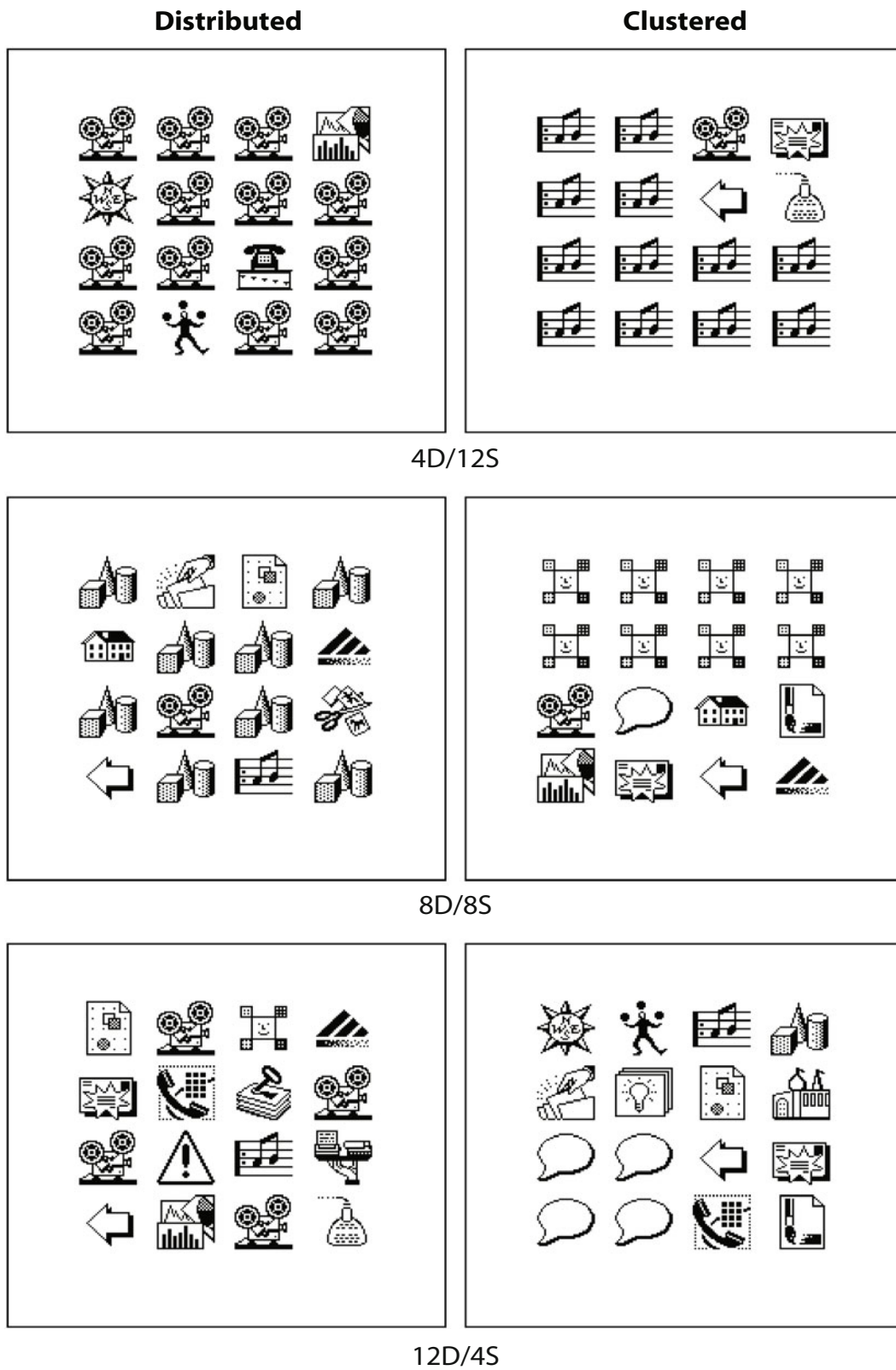


Figure 6. Examples of distributed versus clustered 4D/12S, 8D/8S, and 12D/4S displays. D, different; S, same.

(5) when the display comprises multidimensional items of the sort shown in Figure 3, as the items are made more similar (e.g., by holding brightness constant), the displays are more likely to be reported *same* (Castro, Young, & Wasserman, 2002, and new data presented here); and (6) when the items are made less distinguishable through blurring, the displays are more likely to be reported *same* (new data presented here).

The spatial organization effects reported by Wasserman et al. (2000) represent the only published data requiring the deployment of the positional entropy model, instead of the original entropy model. Wasserman et al. (2000) reported evidence showing that a  $2 \times 3$  attentional window provided the best fit by the positional entropy model. To determine the optimal  $c$  value for pigeons in the finding differences model, we ran simulations of Wasserman et al. (2000). Subsequent simulations of pigeons' performance in other experiments used this optimal  $c$  value to fit data from the other studies; we assumed that the effect of distance was constant across studies. Likewise, the optimal  $c$  for humans was determined using a human analogue (new data presented later in this article). Although treated here as a constant, the value of  $c$  may vary as a function of the size of the items, their proximity, practice with the task, and distracting stimuli, *inter alia*.

Because of the very large number of possible configurations of items in some of the displays, we ran 100 simulations for each display type to sample the range of configurations; we then computed the average positional entropy and average activation for these 100 configurations. These averages served as the predictors for the model fits. Optimal parameter values of Equation 6 were determined by maximum likelihood estimation (MLE). Although we would have preferred analyzing individual-subject estimates of the parameters, these estimates were often unstable (e.g., if a subject showed a step function discrimination, a range of bisection points are equally descriptive of the performance). Conversely, fitting only the average subject results in small sample sizes. We made a strong assumption by treating subjects as exact replications and fit the entire data set using a single set of parameters. In those situations in which clear individual differences were revealed by a cluster analysis (see below), we fit the major clusters separately.

Although  $R^2$  values are commonly used to describe the quality of a fit, they can be inflated by extreme values. In lieu of  $R^2$ , we will show figures of the means and standard errors of the pigeon and human data and superimpose the fitted value on these figures. Only the finding differences model's fits will be shown, because this model was nearly always superior and adding the positional entropy fits merely obscured the figures to no purpose.

### Spatial Organization

**Pigeons.** The first set of simulations involved a systematic exploration of the optimal value of  $c$  that was necessary to fit the data from Experiment 2 in Wasserman et al. (2000). We ran a series of simulations across a range of  $c$  values from .20 to 1.00 in steps of .10.<sup>1</sup> Smaller  $c$  values indicate little effect of spatial distance, whereas larger  $c$

values indicate greater effects of spatial distance. The optimal value for  $c$  was .65. Figure 4 illustrates the effect of spatial distance for this value; at this value, items that are more than three "units" apart have little influence on each other (adjacent items in the same row or column are one unit apart; items in the same row or column with one item between them are two units apart, etc.).

The finding differences model captured the overall effect of clustering the identical items together (clustered displays); these displays were rated as more *same* than were those displays in which the same collection of items involved no clustering (distributed displays). Although small details of these pigeon data were not fully captured by either model, the overall fits were good (see the top of Figure 7).

**People.** To obtain data on the effect of spatial distance on people's discrimination of variability, we performed a new experiment in which subjects learned to discriminate arrays with *low* variability (entropy of 1.0) from those with *high* variability (entropy of 3.0), rather than *same* (entropy of 0.0) from *different* (entropy of 4.0), in order to prevent the use of a simple strategy in which arrays of identical items were distinguished from arrays in which any items were different (Young & Wasserman, 2003).

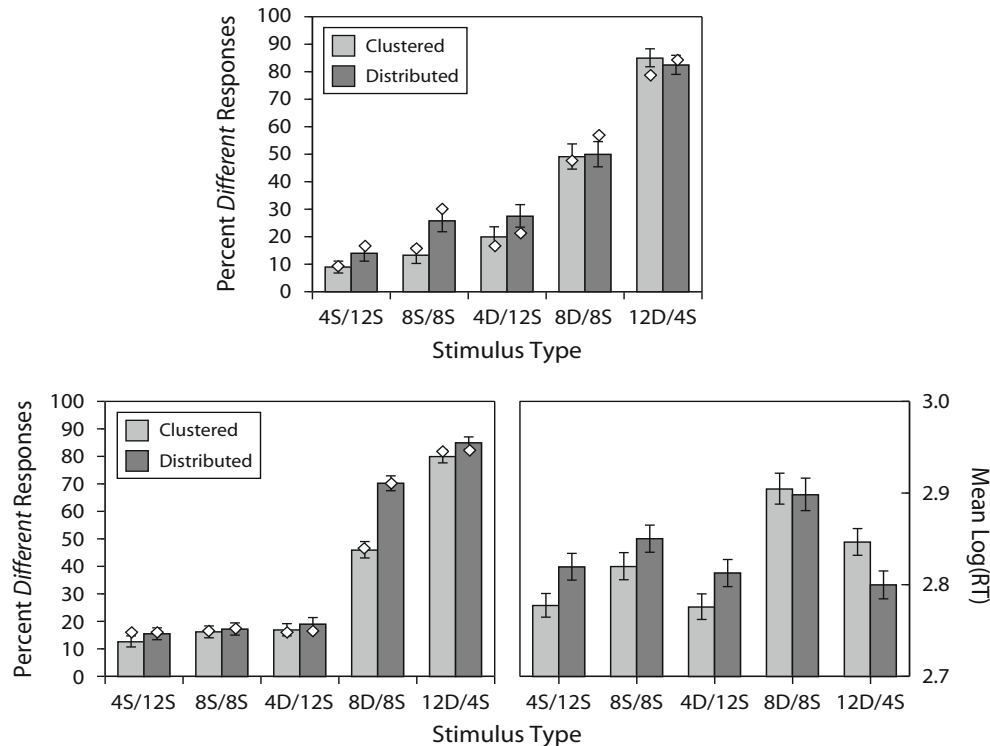
After initial training, we evaluated the effect of item proximity on people's variability discrimination by introducing various mixture arrays in which identical items were either distributed or clustered (Figure 6). These testing arrays involved *same* and *different* icons of various mixtures: 4S/12S, 8S/8S, 4D/12S, 8D/8S, and 12D/4S, where *S* indicates a group of identical items and *D* indicates a group of nonidentical items.

People were increasingly likely to report *different* as the entropy of the display increased (cf. Young & Wasserman, 2001a). The mixture displays involved entropies of 0.8 (4S/12S), 1.0 (8S/8S), 1.3 (4D/12S), 2.5 (8D/8S), and 3.5 (12D/4S). Averaged across the distributed and clustered displays of this visual information, the percentage of trials with a *different* response progressively rose: 14%, 17%, 18%, 58%, and 82%, respectively.

Changing the organization of the display produced reliable changes in the subjects' responding (see the left side of the bottom of Figure 7). Clustering the identical items together produced more *same* responses, although the magnitude of this effect differed reliably across mixtures. Apparently, identifying similarities is made easier by spatial proximity. The clustering effect was largest for the mixture that produced an intermediate *different* response probability. This effect may be due to the greater sensitivity of behavior that is further from the floor or ceiling.

Organization effects were also revealed in the mean log RT data (right side of the bottom of Figure 7; a logarithmic transform was used to normalize the RT distribution). For those displays with low entropy (4S/12S, 8S/8S, and 4D/12S), organizing the display (i.e., making it appear to have even less entropy; see the left side of the bottom of Figure 7) produced shorter RTs. For the display with high entropy (12D/4S), organizing the display (i.e., again making it appear to have less entropy; see Figure 7) produced longer RTs. And, for the display with an intermediate degree of entropy (8D/8S), the RTs were not reliably





**Figure 7. Spatial organization. Mean percentage of *different* responses for the pigeon study (top) and mean percentage of *different* responses and  $\log_{10}$  reaction times (RTs) for the human study (bottom). The diamonds indicate the fitted values from the finding differences model.**

altered by changes in display organization. Here, effects of organization that were not evident in choice responding were evident in RTs.

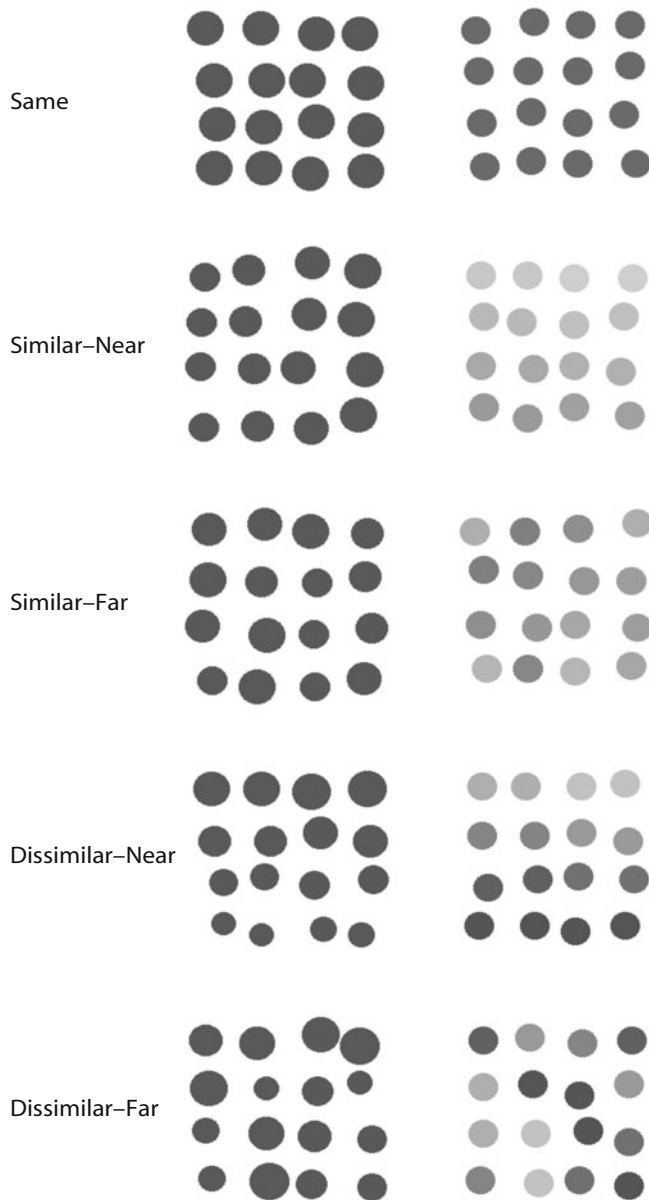
The simulations involved a systematic exploration of the optimal value of  $c$  that is necessary to fit this data set. We again ran a series of simulations across a range of  $c$  values from .20 to 1.00 in steps of .10, optimizing the parameters of Equation 2 using MLE for each run. The optimal value for  $c$  was .50, suggesting that people were less affected by the distance between items than were the pigeons. However, likelihood ratio tests of the fit of the finding difference model as a function of the  $c$  parameter revealed that a moderately broad range of  $c$  values produced similar fits. The finding differences model captured the overall effect of clustering the identical items together (see the bottom of Figure 7).

### Spatial Organization and Continuous Similarity (People Only)

To compute entropy, items in the display are judged to be in the same category or in different categories. To determine the extent to which relative similarity plays a role in categorical variability judgments, we tested people's classification performance, using displays comprising items that varied along a continuous stimulus dimension, either size or color, and revisited the issue of display organization using these stimuli (see Figure 8). Displays involving different but very similar items should generate higher error rates and longer RTs than will displays involving

different dissimilar items, despite the fact that the entropy of these displays is identical (3.0). We also expected to replicate the effects of spatial organization, but with the proximity of similar, not identical, items producing an increase in *same* responses. Furthermore, we predicted the effect of spatial organization to be larger when the items were more similar, because grouping together very similar items should produce small difference scores; in contrast, grouping together the most similar items in a display of dissimilar items should still result in adjacent items producing relatively large difference scores.

In the *different* displays, we studied (similar–near, similar–far, dissimilar–near, and dissimilar–far; see Figure 8), the 16 items comprised two copies of each of 8 items. *Near* displays (in which the most similar items were adjacent) were programmed to appear in the following configuration: 1122/3344/5566/7788 (in which the number designates stimulus rank along the relevant dimension in each row of the display; see Figure 8). The configuration was randomly rotated 90°, 180°, or 270° to create four possible configurations (e.g., 2468/2468/1357/1357 is a 90° rotation of the configuration above). *Far* displays simply randomized the location of each item within the display (although it is possible that this randomization procedure could produce a display in which similar items are proximal, this outcome would be very rare and would reduce our effect sizes). The range of sizes and colors was relatively narrow for *similar* displays and relatively broad for *dissimilar* displays. For the *same* displays, an item was



**Figure 8.** Examples of the stimuli used in the size-varying (left column) and color-varying (right column) conditions. The actual size-varying items were a shade of blue, and the color-varying stimuli were shades of blue, green, and various blue/green mixtures.

randomly chosen from the legal items for the appropriate condition (size or color varying) and replicated 16 times.

We performed two analyses of discriminative performance: The first examined accuracy on the *different* trials, and the second examined RTs on the *different* trials. The variables of distance between similar items and their similarity were irrelevant for *same* trials. Performance on *same* trials was uniformly strong ( $M = 92\%$ ).

The results are shown in Figure 9. Generally, the data followed the anticipated ordering from the most varied *different* display (the leftmost bar in each graph corresponds to displays containing generally dissimilar items, with the

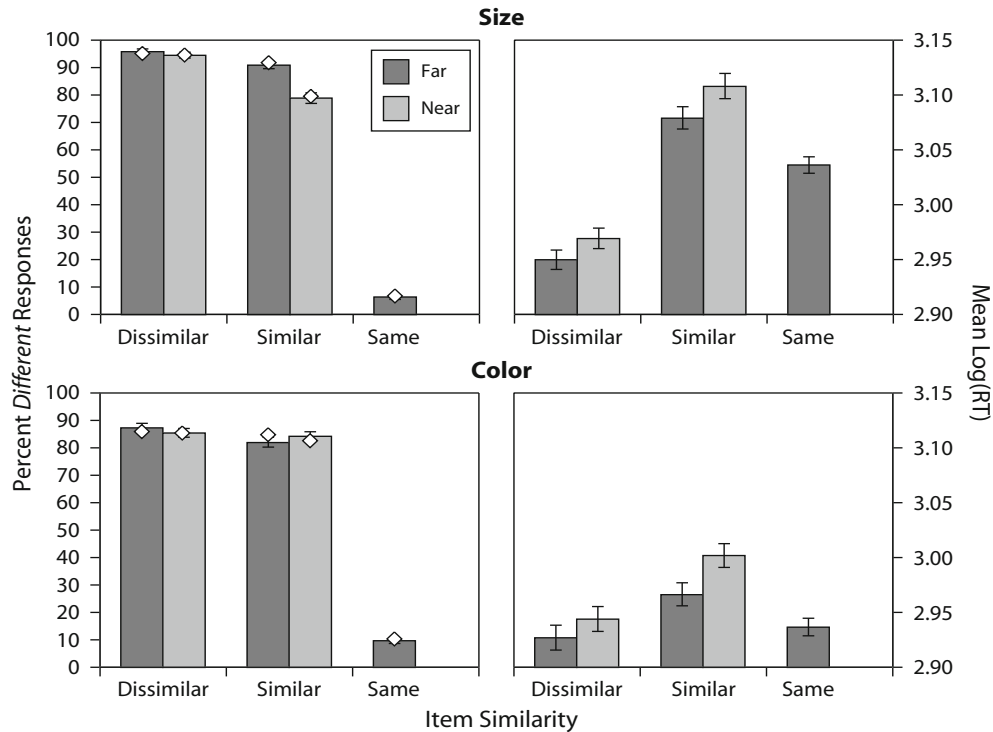
most similar items located far apart) to the least varied *different* display on the right (which corresponds to displays containing generally similar items, with the most similar items located near each other). Judged variability generally decreased and mean log RTs increased as predicted judged variability decreased. The effect was stronger for the task involving variability in size than for the task involving variability in color. The subjects seemed to show generally lower accuracy for the color-varying displays, but with shorter RTs; there was no reason to suspect that the subjects would sacrifice accuracy for speed in this condition.

The positional entropy model has no basis for predicting the effects of similarity observed in Figure 9, because each of the *different* displays has the same entropy (3.0). To test the finding differences model, we coded the size and color values of each stimulus as a point on a continuous scale between .30 and 1.00 in .10 steps for dissimilar items and between .65 and 1.00 in .05 steps for similar items. Note that this coding presupposes a linear ordering of differences along color and size, but other scales (e.g., a logarithmic scale and a scale in which values were estimated from the data) produced similar fits. As long as the coding roughly matched the ordinal value of the observations, the fit was good. The simulations clearly indicate that the finding differences model can capture the relationship between similarity along continuous dimensions and discriminative behavior in a variability discrimination task (see Figure 9 to compare fitted and observed values).

### Sensitivity to Mixtures

**Pigeons.** The original study documenting an effect of item mixture (how many of each kind of item was present within a display) on discrimination behavior involved several different types of icon mixtures across three different experiments (Young & Wasserman, 1997). A more recent study involved 11 different mixtures within a single experiment that spanned an even larger entropy range than did the original study (Young & Wasserman, 2002). We used the data from the more recent study to test the generality of the two focal models. The displays were  $5 \times 5$  arrays in which 9 of the 25 locations were empty (and thus were not involved in the calculations, except by altering the distance between items). The finding differences model reproduced the overall effect of increasing stimulus variability (see the top of Figure 10). The fit was very strong given that the value of  $c$  was derived from a different set of data.

**People.** We used the data from Experiment 1 of Young and Wasserman (2001a) to test the effect of mixtures. That study showed significant individual differences in two major clusters of subjects as a function of the number of items in the display. The larger cluster (categorical) evidenced little effect of number for either *same* or *different* arrays; the smaller cluster (continuous) evidenced an effect of number for *different* arrays, but not for *same* arrays. When separately fit to the two clusters of subjects, the finding differences model captured the effect of icon mixture in both clusters (see the bottom of Figure 10), although the “4444” stimulus (four icons of each of four types) in the continuous cluster was significantly underpredicted.



**Figure 9. Spatial organization and continuous similarity.** Mean percentages of accurate responses and  $\log_{10}$  reaction times (RTs) for the size-varying (top) and color-varying (bottom) conditions. The diamonds indicate the fitted values from the finding differences model.

### Number of Items

**Pigeons.** The next set of simulations involved fitting the data from Experiment 1 of Young et al. (1997), which investigated the effects of varying the number of items on *same* and *different* trials. After initial training with 16-item *same* and *different* displays, pigeons' responding to *same* displays was unchanged across a range of smaller display sizes, whereas *different* displays were increasingly likely to be reported as *same* as the number of items was reduced, with 2-item *different* displays consistently reported as *same*. This unanticipated asymmetry was captured by the entropy model, and is worthy of revisiting in the present context. The simulation procedure was identical to that used for the mixture simulations, because Young et al. (1997) also used  $5 \times 5$  arrays, in which the icons were randomly placed.

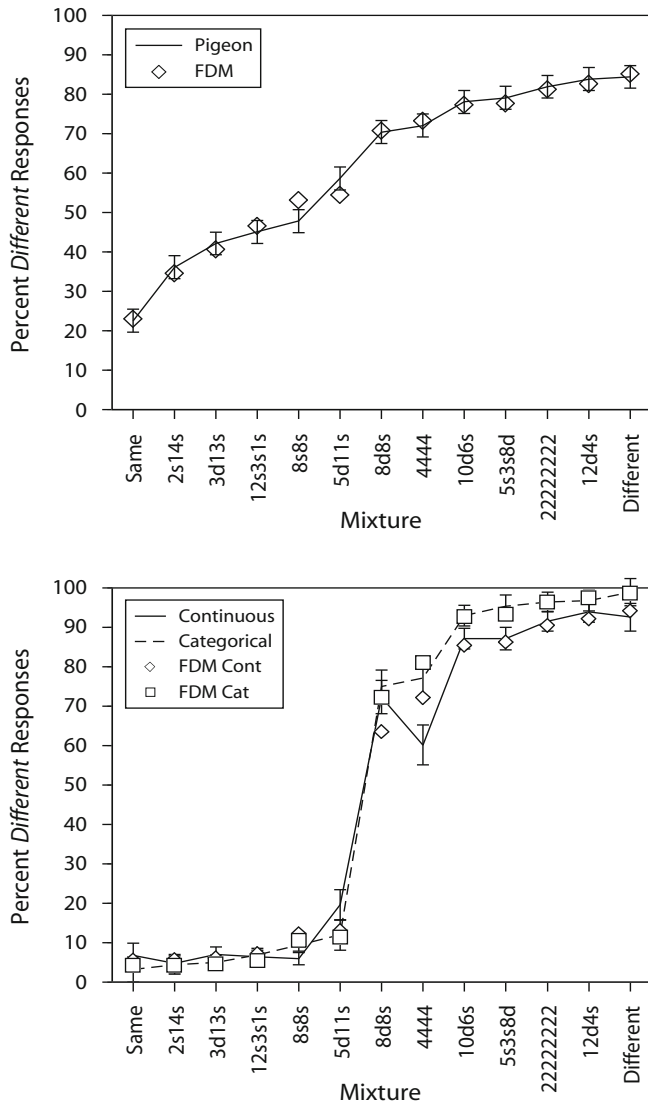
The finding differences model captured the asymmetry in the effect of item number on *same* and *different* arrays (see the top of Figure 11). The asymmetrical effect of item number on entropy occurs because entropy is based on the distribution of category probabilities. Increasing the number of items in a *same* display does not affect the probability distribution, whereas increasing the number of items in a *different* display increases the number of categories that are present. The asymmetrical effect of number on the finding differences model is the product of a choice in the construction of the underlying GS2 model. The model accumulates evidence of differentness, not sameness, for each item in the display. With fewer items, there are fewer possible differences to identify.

**People.** The simulations of human performance involved fitting the data from Experiment 1 of Young and Wasserman (2001a), in which we investigated the effects of varying the number of items on *same* and *different* trials. Recall that Young and Wasserman (2001a) found two clusters of subjects based on their response to item number: The continuous group behaved like the pigeons by showing a strong asymmetry in the effect of number on *same* and *different* trials, whereas the categorical group showed a much weaker effect of number. We independently fit the model to these two clusters of subjects. The simulation procedure was otherwise identical to that used for the number simulations involving pigeons. The finding differences model captured the overall asymmetry of the effect of item number on *same* and *different* arrays for both clusters of subjects (see the bottom of Figure 11).

### Multidimensional Similarity (Four-Dimension Studies)

All of the studies considered thus far involved items that were categorically different in shape or varied along a single continuous dimension (size or color). In the next set of experiments, we systematically varied the relative similarity of items that were composed of four separable dimensions (brightness, shape, color, and planar orientation; see Figure 3) and manipulated the relative similarity of the items in a display by holding one or more of these dimensions constant.

**Pigeons.** In a pair of experiments, Young, Castro, and Wasserman (2002) trained pigeons to discriminate the vari-



**Figure 10. Sensitivity to mixtures.** Mean percentages of *different* responses for the pigeon study (top) and the human study (bottom). The diamonds indicate the fitted values from the finding differences model (FDM). The 3d13s types of notation designate a display including 3 different random icons (e.g., a compass, document, and phone) and 13 identical icons (e.g., 13 copies of a train). The 4444 display included 4 copies of each of four icon types, whereas the 22222222 display included 2 copies of each of eight icon types.

ability of displays like those shown in Figure 3. The pigeons' behavior was controlled primarily by differences in brightness, with little effect of the other dimensions. These results, however, are still problematic for the original entropy account, because the entropy of all of the displays was identical (3.0). The positional entropy model falls prey to the same problem, because entropy is based on identity versus nonidentity; there is no provision for degrees of similarity.

In fitting these data with the finding differences model, we initially chose to use a systematic but arbitrary group of salience parameters that were intended to reflect the relative ordering of salience, not the particular differences:

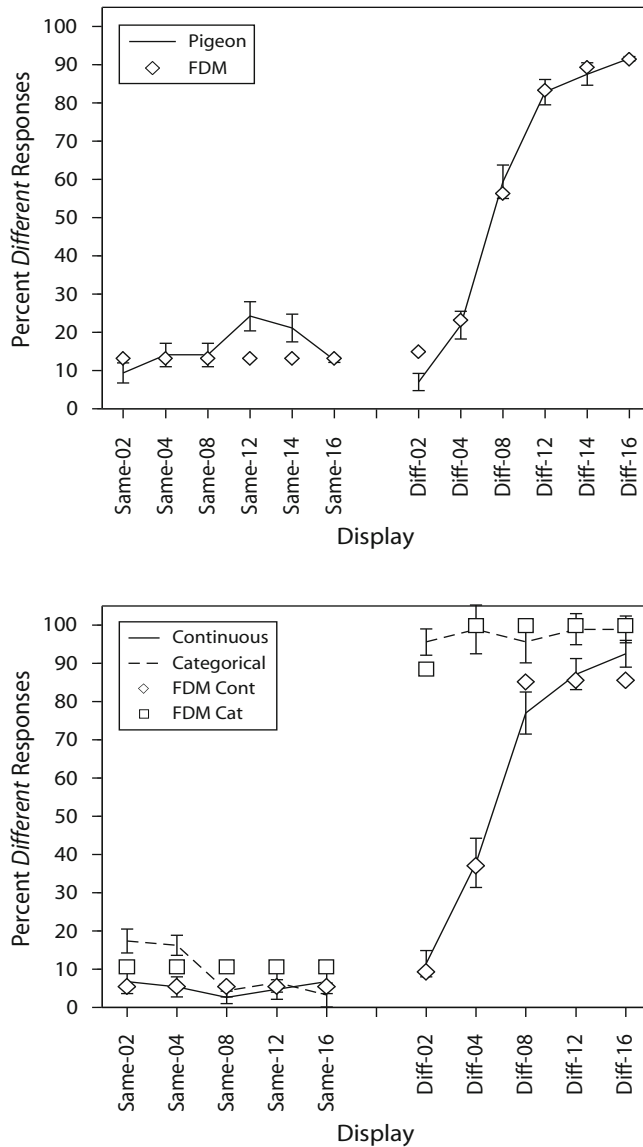
1.00, .90, .80, and .70, for brightness, size, orientation, and shape, respectively. Fits involving other values had no discernible impact on the fit as long as the brightness dimension received the highest salience value. Although the finding differences model provided an excellent fit, discriminative performance was almost entirely determined by brightness, thus giving only limited evidence of sensitivity to multidimensional variability by pigeons (i.e., the data primarily show evidence of sensitivity to unidimensional variability).

**People.** To examine multidimensional similarity in people, we conducted a new experiment with the same 4-D stimuli as those used in the pigeon study. The human experiment, however, used a design that required attention to all four dimensions in order to optimize discriminative performance. The subjects were assigned to one of two conditions: *same versus others* or *different versus others*. In the *same versus others* condition, half of the displays were *same* arrays, and the other half were displays in which one (*mixture*) or zero (*different*) dimensions were held constant while the others varied. In the *different versus others* condition, half of the displays were *different* arrays, and the other half were displays in which one (*mixture*) or four (*same*) dimensions were held constant while the others varied (see Figure 12). We were concerned that *same versus others* training might prove too easy and result in ceiling effects, so we included *different versus others* training for half of the subjects to provide a second test of the effect of similarity.

The *same versus others* discrimination was indeed much easier; during the 144 trials of discrimination training, the subjects averaged 88% and 66% correct in the *same versus others* and *different versus others* conditions, respectively. This finding parallels that of Young and Wasserman (2002), who reported that subjects easily discriminated displays with an entropy of 0.0 (*same*) from displays with an entropy of 0.5 or more, but subjects found it difficult to discriminate displays with an entropy of 4.0 (*different*) from those with an entropy of 3.5 or less. The profiles of performance in each condition are shown in Figure 13.

The subjects' behavior indicated that the salience of the four dimensions was ordered from size (least salient), to orientation, to shape, to brightness (most salient). This ordering can be inferred from the observation that, when a low salience dimension was held constant (e.g., in the *size* mixture), the subjects in the *same versus others* condition predominately classified the array as *different* (with a negligible increase in mean log RT relative to the *different* arrays; see the top of Figure 13), whereas the subjects in the *different versus others* condition had the greatest difficulty classifying the array as *same* (see the bottom of Figure 13). Similarly, when a high salience dimension was held constant (e.g., in the *brightness* mixture), the subjects in the *same versus others* condition were more likely to misclassify the array as *same* (with a significant increase in RT; see the top of Figure 13), whereas the subjects in the *different versus others* condition predominately classified the array as *same* (with little increase in RT relative to the *same* arrays; see the bottom of Figure 13). Because of the clear behavioral disparities produced by the





**Figure 11.** Number of items. Mean percentages of *different* responses for the pigeon study (top) and the human study (bottom). The diamonds indicate the fitted values from the finding differences model (FDM).

new design, these data present a greater challenge to our models. Although the positional entropy model does not predict the similarity effects observed here, the finding differences model provided a good fit for the different versus other group and for the same versus other group (see Figure 13).

### Blurring

In our final two experiments, one with pigeons and one with humans, we examined the effect of stimulus blurring on responding. These experiments served two purposes: (1) to determine whether responding in our task is due to a low spatial frequency discrimination and (2) to provide a second assessment of the effect of item similarity on performance. A low-spatial-frequency account of our data suggests that

observers of these displays are performing the equivalent of a texture discrimination task. If visual variability discrimination is mediated by low spatial frequencies, blurring the icons should have little or no effect on responding, because blurring degrades higher spatial frequencies and should retain the basic texture of these displays.

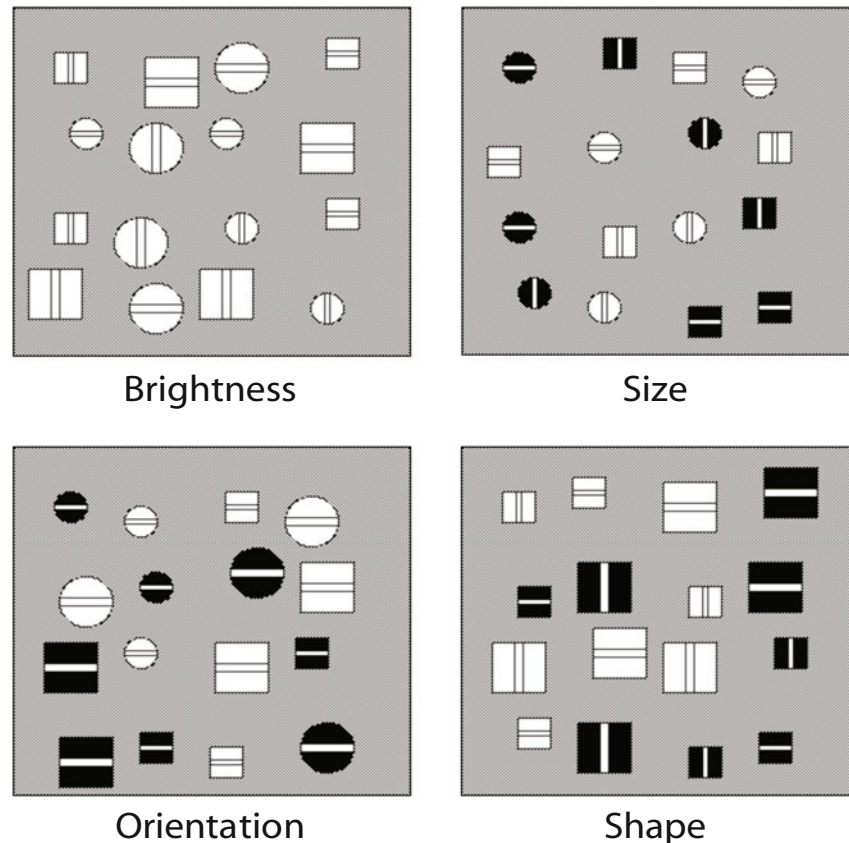
Young and Wasserman (2001b) obtained evidence that the pigeon's variability discrimination performance is not due to low spatial frequency discrimination by showing that factors that altered the lower spatial frequencies had no effect on behavior. The present experiments take the opposite tack by determining whether factors that affect the higher spatial frequencies (while leaving low frequencies intact) do alter discriminative behavior (cf. Nothdurft, 1991, in which spatial frequency filtering had different effects on item detection and texture segregation).

In contrast to a spatial frequency account, the finding differences model predicts that blurring will make the individual items more similar and thus produce more *same* responding on trials involving nonidentical items. Note that this effect of blurring should be decidedly asymmetric; *different* displays should look more same-like, whereas *same* displays should be unaffected because the items are already identical. As well, this asymmetry should increase as the level of blurring is increased.

**Pigeons.** To blur the stimuli, each icon was loaded into Adobe Photoshop and subjected to a Gaussian blur. The original size of each icon "box" ( $32 \times 32$  bits) was maintained during the blurring process, so that the icons were never so large as to abut one another. The blurring levels (1.0, 2.0, 3.0, and 4.0, in terms of pixel widths) were chosen to represent a range of readily discernible differences. Because of the large effects of blurring that we observed in pigeons in our first phase of testing, we performed a second testing phase involving more subtle blurring. The stimuli included the originally tested values plus the subtler levels (0.3 and 0.6); a sample complete profile is shown in Figure 14.

Discrimination of *same* from *different* trials during the 5 days directly preceding testing averaged 90% for the 4 birds. All the pigeons underwent 8 days of Phase 1 testing and 4 days of Phase 2 testing. During Phase 1, blurring the icons had a profound effect on the pigeons' discriminating *same* from *different* arrays (Figure 15, solid bars). Although accuracy to *same* and *different* training arrays was uniformly high ( $M = 86\%$ ), accuracy to blurred arrays was near chance ( $M = 53\%$ ). Importantly, the effect of blurring was asymmetrical; the *same* arrays were unaffected by blurring, whereas the *different* arrays were consistently classified as *same* when blurred.

During Phase 2, the behavioral effects of the smaller blurring levels were considerably attenuated (Figure 15, hashed bars). Although accuracy to *same* and *different* training arrays was uniformly high ( $M = 85\%$ ), accuracy to the blurred arrays decreased as the blurring level increased ( $M = 67\%$ ,  $56\%$ , and  $51\%$  for blurring levels of 0.3, 0.6, and 1.0, respectively). Again, the effect of blurring was asymmetrical; the *same* arrays were unaffected by blurring, whereas the *different* arrays were classified as *same* when blurred.



**Figure 12.** Examples of the 4-D mixture arrays. These arrays were created by holding one of the four dimensions (brightness, size, orientation, or shape) constant while allowing the others to vary. This procedure resulted in four types of arrays, depending on which dimension was held constant.

Like the results from the 4-D studies, these results are problematic for the positional entropy model because there is no provision for degree of similarity. In fitting these data using the finding differences model, we chose to grade differences as a function of the degree of blurring. The grading values were somewhat arbitrary but were initially chosen to reflect a linear relationship between similarity and the level of blurring and to appreciate that the blurred items were still discriminable. Differences in the model were multiplied by 1.00, .97, .94, .90, .80, .70, and .60, for blurring levels of 0.30, 0.60, 1.00, 2.00, 3.00, and 4.00, respectively. When we tested a logarithmic scaling of blurring levels, the fit was largely unaffected. The finding differences model again provided a good fit (see Figure 15). The simulations clearly indicated that the model can capture the ordinal relationship between similarity (as affected by stimulus blurring) and discriminative performance in a variability discrimination task—most important, the observed asymmetry in the effects of blurring to *same* and *different* displays.

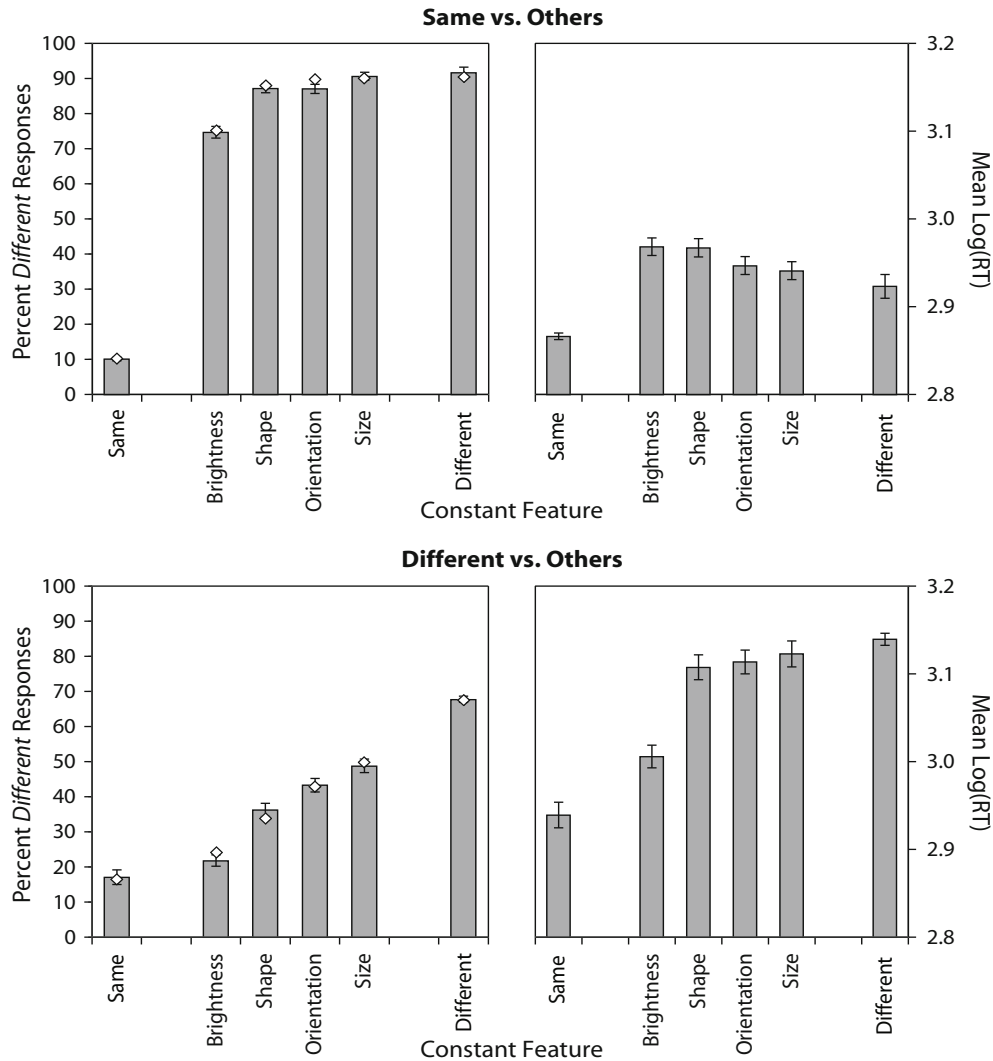
**People.** The subjects initially received 60 trials of *same–different* discrimination training using intact stimuli. We established a training criterion of 70% correct for the last 10-trial block, because a few of the subjects showed complete inattentiveness to the task (performing

near chance throughout training on all of the displays); given that the training discrimination was quite easy, the poor performance of these nonlearners may have been due to external college demands at this time in the semester. A total of 15 of the 20 subjects met the 70% criterion; all subsequent analyses exclude the nonlearners. During the last block of training trials, the retained subjects averaged 90% correct. This high level of accuracy on the training displays was maintained in testing ( $M = 95\%$ ). Testing stimuli comprised only the 0.0, 1.0, 2.0, 3.0, and 4.0 blurring levels shown in Figure 14.

During testing, the subjects showed poorer discrimination as the blurring level increased; the effect was much larger for *different* trials than for *same* trials (left side of Figure 16). The mean log RTs showed a similar trend, with a significant increase in RTs with increased blurring for the *different* trials but little evidence of an increase for *same* trials (right side of Figure 16). The finding differences model again provided a good fit, reproducing the asymmetrical effects of blurring quite well.

### Summary

For three of the six data sets, both focal models could account for most of the variance in mean discriminative performance. The clearest differences in the predictive



**Figure 13. Multidimensional similarity.** Mean percentages of *different* responses and  $\log_{10}$  reaction times (RTs) during the testing phase of the *same versus others* condition (top) and the *different versus others* condition (bottom). The diamonds indicate the fitted values from the finding differences model.

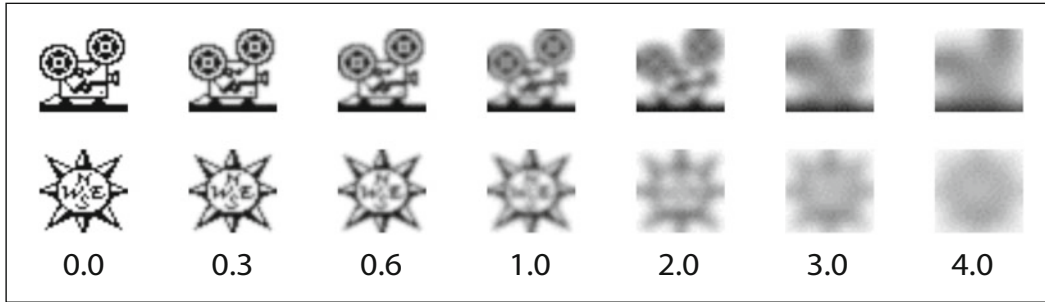
power of the models were disclosed in fitting the similarity (4-D and continuous) and blurring data sets. Only the finding differences model could incorporate the effects of degree of similarity and the differential salience of each dimension. The finding differences model provided fits that are generally as good as those provided by the positional entropy model for stimuli that differed categorically, but it offers something more: the ability to fit new data sets that are problematic for the original model.

## DISCUSSION

After over a decade of empirical research on variability discrimination in pigeons and people (Wasserman et al., 2004), the limitations of the informational concept of entropy as a complete account of our results have become strikingly apparent. In the present article, we have integrated cognitive models from the realms of attention, categorization, and

choice to create a single, unified account—the finding differences model—as a new explanation of the wide array of behavioral data that we have previously reported with both pigeons and people as subjects of study. The new model did a creditable job of accounting for the existing data; as well, the results of the new experiments reported here filled gaps in the literature, and the experiments also included investigations of item similarity that provided unique support for our latest account of the observed behavior.

Importantly, this breakthrough was accomplished without positing a new model developed expressly to explain variability discrimination; we simply adapted a well-validated model of a very different cognitive task, visual search. This approach avoids adding yet another model to the pantheon of task-specific models that populate the field of cognitive psychology. As will be demonstrated below, we also believe that this model can be applied to still more tasks in the comparative cognition domain.



**Figure 14.** Examples of the effect of Gaussian blurring on two of the icons used in the arrays. Blurring levels of 0.0, 1.0, 2.0, 3.0, and 4.0 were used in the human experiment, and all blurring levels were used in the pigeon study.

The present project served to evaluate a model from the same general class as Wolfe’s (1994) GS2 and to contrast it with entropy-based models. The parameter details of the Minkowski metric, exponential decay as a function of distance, and dimensional salience were not systematically explored. Although answering questions about the precise nature of these parameters is worth pursuing, our simulations showed that the majority of the variance in behavior can be predicted without appeal to these particulars. The performance of the finding differences model suggests that this type of model captures the stimulus features that control discriminative responding in our family of tasks: the number and degree of localized differences in a collection of items.

Entropy had been successful as an account of discriminative performance because the collections involved items that did or did not match one another along a single dimension. The new model is not so constrained; *degrees* of difference are central to the new model, and they can, in principle, be assessed through any measure of similarity. Thus, any manipulation that affects the similarity of the presented items is predicted to affect discriminative performance. This notion was confirmed in the present

studies of multidimensional similarity, blurring, and continuous dimension similarity.

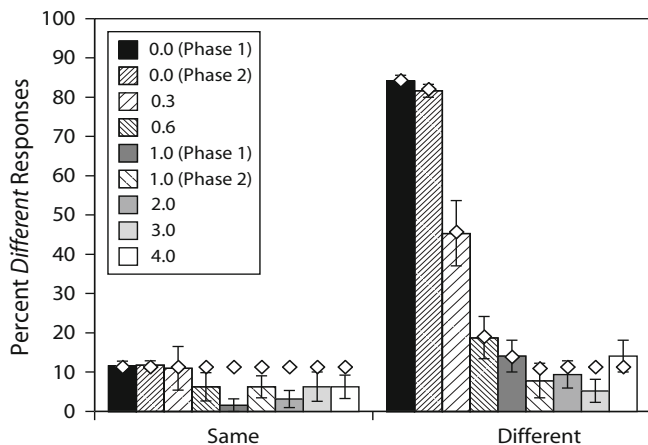
Conversely, we would infer that any manipulation that does not affect discriminative performance in our variability task also does not affect perceived similarity. As one noteworthy example, Young and Wasserman (2001b) discovered that intermixing the identical items in *same* displays at multiple planar orientations (e.g., jugglers presented at 0°, 90°, 180°, and 270° of rotation) had no discernible effect on the pigeon’s discrimination of variability. This result leads us to conclude that pigeons deem such items to be identical despite obvious differences in planar rotation (Delius & Hollard, 1995, however, reported that pigeons could be trained to discriminate items as a function of their planar orientation).

Because of its reliance on similarity, the finding differences model can be applied to any set of stimuli for which a principled measure of similarity exists. Thus far, our stimuli have differed along separable categorical and continuous dimensions. Yet stimulus dimensions are not confined to those that are separable (Shepard, 1991). Different measures of similarity may be necessary for integral dimensions, but we predict that the general model will still apply.

We suggest that the contribution of our new model to an understanding of variability discrimination is considerable. The model suggests new investigations involving (1) the consequences of prior training in which similarity is behaviorally mediated (e.g., through the creation of stimulus equivalence classes), (2) the separable versus integral nature of stimulus dimensions, (3) features that vary in a correlated fashion (e.g., size and color vary together) or uncorrelated fashion (e.g., size and color vary independently), and (4) the effect of differently shaped distributions of dimensional value (e.g., uniform vs. normal) on discrimination. The model may also help to resolve other key issues in the literature by providing an account of the similarities and differences between our variability discrimination task and the oddity discrimination tasks of Cook and his colleagues, described next.

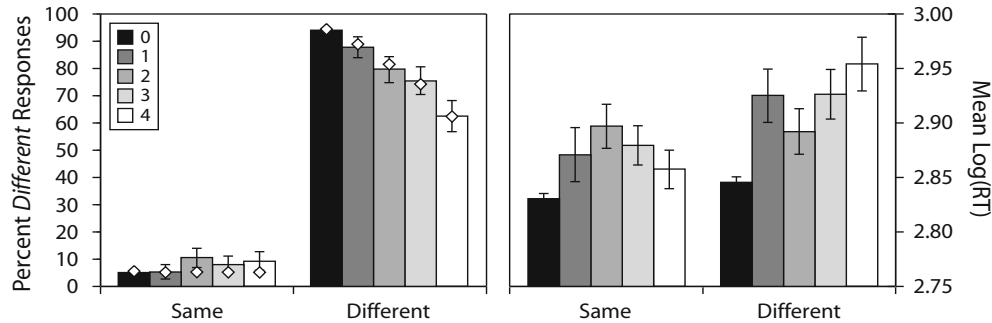
**Oddity Discrimination**

In an oddity discrimination task, pigeons must discriminate displays that contain odd items or areas from displays that are uniform and do not contain odd items or areas (Cook, Cavoto, & Cavoto, 1995; Cook, Katz, &



**Figure 15. Blurring.** Mean percentages of *different* responses during Phases 1 (unhashed bars) and 2 (hashed bars) in the pigeon study. The legend indicates the level of blurring (note that Gaussian blurs of 0.0 and 1.0 occurred in both phases). The diamonds indicate the fitted values from the finding differences model.





**Figure 16. Blurring.** Mean percentages of *different* responses and  $\log_{10}$  reaction times (RTs) during the testing phase in the human study. The legend indicates the level of blurring. The diamonds indicate the fitted values from the finding differences model.

Cavoto, 1997); plus, pigeons must pinpoint the location of any differences. The activation map that underlies our finding differences model may provide precisely the information that a pigeon needs to solve this task. Although the pigeons in our variability discrimination tasks can ignore the location of differences, the pigeons in the oddity detection tasks need to leverage the location information that is present in the activation maps (Gibson, Wasserman, & Cook, 2006, documented pigeons' ability to perform oddity discrimination or variability discrimination with the same set of stimuli, but pigeons show no transfer between the tasks).

A notable disparity between the variability discrimination task and the oddity task is the difference in the effects of item number on performance. Cook et al. (1997) reported that increasing the number of items in an oddity task (e.g., from one odd item with two identical items to one odd item with five identical items) produces more accurate performance when these displays must be discriminated from displays with no odd item. Increasing the number of items, however, produces lower overall entropy (.90 vs. .70 in this example), which should make it harder to discriminate odd displays involving many items from displays with no odd item (i.e., a *same* display with an entropy of 0)—but only if the pigeons were trained to respond on the basis of average activation. In an oddity task, however, pigeons may be learning to respond on the basis of the difference in the activation level of the odd item and that of its background, because odd items differentially stand out depending on the magnitude of this difference.

To explore this possibility, we applied the finding differences model to some of Cook et al.'s (1997) displays. If pigeons are trained to respond on the basis of the variability (i.e., the average activation) of the display, an increase in the number of identical items is predicted to impair performance (the average activation decreases with more identical items). But if the oddity task causes the pigeons to respond on the basis of the maximal difference between the activation levels of the items, an increase in the number of identical items is predicted to improve performance (to a point). Because the odd item's activation level is increased by the addition of more background items (from which it differs), whereas the background items' activa-

tion levels tend to remain relatively constant, the odd item stands out more, relative to its background. Thus, the finding differences model may provide a way to reconceptualize the variability and oddity discrimination tasks by using the same underlying activation map. Pigeons may simply be learning to respond to different aspects of that map.

### Final Thoughts

We propose that it is highly adaptive for organisms to find differences, whether in the service of a visual search task (e.g., locating the tiger in the grass), an oddity discrimination task (e.g., finding the nonconformists in your organization), or a variability discrimination task (e.g., identifying which flock has the greater variety in potential mates). Furthermore, with rather little effort, people find differences when they shop (assessing product variety), look at an audience (assessing racial or age diversity), or visit an art museum (assessing the represented variations in styles, artists, and exhibits). Understanding this vital discrimination process will surely affect how marketers and museum curators organize and package their wares in order to manipulate our perceptions, but it can also inform the development of contemporary theories of the out-group homogeneity effect (Kashima & Kashima, 1993), perceived creativity (Eisenman, Hannon, & Bernard, 1966), and variability preference (Munsinger & Kessen, 1966).

Given that “variety is the very spice of life,” perhaps efforts to increase perceived variability suit our own predispositions. “The deep, inexorable variability of the natural world attracts us. We don't tire of watching a fire or a sunset because it's always different. . . . it is because nature is always *different* that we find it fascinating” (Yoerg, 2001, p. 193).

### AUTHOR NOTE

This research was supported by National Science Foundation Grant 9904569. Correspondence about this article should be addressed to M. E. Young, Department of Psychology, Southern Illinois University, Mail-code 6502, Carbondale, IL 62901-6502 (e-mail: meyoung@siu.edu).

### REFERENCES

- ARIELY, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*, 157-162.  
 ASHBY, F. G., & GOTT, R. E. (1988). Decision rules in the perception and

- categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 33-53.
- 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.
- CARACO, T. (1981). Energy budgets, risk and foraging preferences in dark-eyed juncos (*Junco hyemalis*). *Behavioral Ecology & Sociobiology*, **8**, 820-830.
- CARACO, T., MARTINDALE, S., & WHITTAM, T. S. (1980). An empirical demonstration of risk sensitive foraging preferences. *Animal Behaviour*, **28**, 820-830.
- CASTRO, L., YOUNG, M. E., & WASSERMAN, E. A. (2002, April). *Same-different learning in the pigeon: Entropy and similarity*. Paper presented at the IV Congreso de la Sociedad Española de Psicología Experimental (Fourth Conference of the Spanish Society of Experimental Psychology), Oviedo, Spain.
- CHONG, S. C., & TREISMAN, A. (2003). Representation of statistical properties. *Vision Research*, **43**, 393-404.
- CHONG, S. C., & TREISMAN, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, **45**, 891-900.
- COOK, R. G., CAVOTO, K. K., & CAVOTO, B. R. (1995). Same-different texture discrimination and concept learning by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **21**, 253-260.
- COOK, R. G., KATZ, J. S., & CAVOTO, B. R. (1997). Pigeon same-different concept learning with multiple stimulus classes. *Journal of Experimental Psychology: Animal Behavior Processes*, **23**, 417-433.
- DELIUS, J. D., & HOLLARD, V. D. (1995). Orientation invariant pattern recognition by pigeons (*Columba livia*) and humans (*Homo sapiens*). *Journal of Comparative Psychology*, **109**, 278-290.
- DUNCAN, J., & HUMPHREYS, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, **96**, 433-458.
- EISENMAN, R., HANNON, J. E., & BERNARD, J. L. (1966). Perceived creativity, set, and preference for simple or complex shapes. *Perceptual & Motor Skills*, **22**, 111-114.
- FRIED, L. S., & HOLYOAK, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 234-257.
- GIBSON, B. M., WASSERMAN, E. A., & COOK, R. G. (2006). Not all same-different discriminations are created equal: Evidence contrary to a unidimensional account of same-different learning. *Learning & Motivation*, **37**, 189-208.
- HOCH, S. J., BRADLOW, E. T., & WANSINK, B. (1999). The variety of an assortment. *Marketing Science*, **18**, 527-546.
- KAHN, B. E., & WANSINK, B. (2004). The influence of assortment structure on perceived variety and consumption quantities. *Journal of Consumer Research*, **30**, 519-533.
- KASHIMA, E. S., & KASHIMA, Y. (1993). Perceptions of general variability of social groups. *Social Cognition*, **11**, 1-21.
- MCGREGOR, P. K., KREBS, J. R., & PERRINS, C. M. (1981). Song repertoires and lifetime reproductive success in the great tit (*Parus major*). *American Naturalist*, **118**, 149-159.
- 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.
- MYUNG, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, **44**, 190-204.
- NEURINGER, A. (2004). Reinforced variability in animals and people: Implications for adaptive action. *American Psychologist*, **59**, 891-906.
- NOSOFSKY, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.
- NOTHDURFT, H. C. (1991). Different effects from spatial frequency masking in texture segregation and texture detection tasks. *Vision Research*, **31**, 299-320.
- PARKES, L., LUND, J., ANGELUCCI, A., SOLOMON, J. A., & MORGAN, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, **4**, 739-744.
- PITT, M. A., MYUNG, I. J., & ZHANG, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, **109**, 472-491.
- SEARCY, W. A. (1984). Song repertoire size and female preferences in some sparrows. *Behavioral Ecology & Sociobiology*, **14**, 281-286.
- SHANNON, C. E., & WEAVER, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.
- SHEPARD, R. N. (1991). Integrality versus separability of stimulus dimensions: From an early convergence of evidence to a proposed theoretical basis. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure: Essays in honor of Wendell R. Garner* (pp. 53-71). Washington, DC: American Psychological Association.
- SKINNER, B. F. (1981). Selection by consequences. *Science*, **213**, 501-504.
- TREISMAN, A. M., & GELADE, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, **12**, 97-136.
- WASSERMAN, E. A., YOUNG, M. E., & COOK, R. G. (2004). Variability discrimination in humans and animals: Implications for adaptive action. *American Psychologist*, **59**, 869-878.
- 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.
- YOERG, S. I. (2001). *Clever as a fox*. New York: Bloomsbury.
- YOUNG, M. E., CASTRO, L., & WASSERMAN, E. A. (2002). *Multidimensional variability discrimination in the pigeon*. Unpublished manuscript.
- YOUNG, M. E., & ELLEFSON, M. R. (2003). The joint contributions of shape and color to variability discrimination. *Learning & Motivation*, **34**, 52-67.
- 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, & 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.
- YOUNG, M. E., & WASSERMAN, E. A. (2003). Visual variability discrimination. In S. A. Soraci & K. Murata-Soraci (Eds.), *Perspectives on fundamental processes in intellectual functioning: Visual information processing* (Vol. 2, pp. 171-197). Stamford, CT: Ablex.
- YOUNG, M. E., WASSERMAN, E. A., & GARNER, K. L. (1997). Effects of number of items on the pigeon's discrimination of same from different visual displays. *Journal of Experimental Psychology: Animal Behavior Processes*, **23**, 491-501.

## NOTE

1. Our software would not converge when we tried to use gradient descent to simultaneously find the optimal values for  $c$ ,  $m$ ,  $b$ ,  $l$ , and  $u$ . Given that the fits were least sensitive to changes in  $c$ , we opted for the approach described here. For the other data sets, smaller  $c$  values sometimes produced better fits, whereas larger  $c$  values (especially those greater than 1.0) invariably produced poor fits. It is possible that the effect of distance changes when the items are placed in  $5 \times 5$  arrays, rather than in the  $4 \times 4$  arrays used by Wasserman et al. (2000).

(Manuscript received April 5, 2006;  
revision accepted for publication January 23, 2007.)