# Variations on variability: effects of display composition on same–different discrimination in pigeons

Leyre Castro · Edward A. Wasserman · Michael E. Young

**Abstract** Discriminating *same* from *different* multiitem arrays can be represented as a discrimination between arrays involving low variability and arrays involving high variability. In the present investigation, we first trained pigeons with the extreme values along the variability continuum (arrays containing 16 identical items vs. 16 nonidentical items), and we later tested the birds with arrays involving intermediate levels of variability; we created these testing arrays either by manipulating the combination of same and different items (mixture testing) or by changing the number of items in the *same* and *different* arrays (number testing). According to an entropy account (Young & Wasserman, Journal of Experimental Psychology: Animal Behavior Processes 23:157–170, 1997), the particular means of changing variability should have no effect on same–different discrimination performance: Equivalent variability should yield equivalent performance. In this critical test of an entropy account, we found that entropy could explain a large portion of our data, but not the entire collection of results.

**Keywords** Comparative cognition · Concept learning · Pigeon

Discriminating collections of same and different items has generally been considered to involve a *categorical* distinction:

L. Castro (✉) · E. A. Wasserman (✉)
Department of Psychology, The University of Iowa,
E11 Seashore Hall,
Iowa City, IA 52242, USA
e-mail: leyre-castroruiz@uiowa.edu
e-mail: ed-wasserman@uiowa.edu

M. E. Young
Department of Psychology, Southern Illinois University,
Carbondale, USA

Two or more items are either the same as or different from one another. However, recent research has suggested that animal and human same–different discrimination may be better represented as a *continuous* distinction involving different degrees of sameness and differentness (e.g., Castro & Wasserman, 2011; Young & Wasserman, 1997, 2001, 2002; for a review, see Wasserman & Young, 2010).

If we examine the *same* and *different* arrays that are depicted in Fig. 1, we see that all of the items in the *same* arrays are identical and that all of the items in the *different* arrays are nonidentical; so, the *same* and *different* arrays actually represent the two ends of a continuum of variability. *Same* arrays represent minimal variability, with all 16 icons the same as one another, whereas *different* arrays represent maximal variability, with all 16 icons different from one another. We can also create arrays depicting a combination of some same and some different icons (mixture arrays), representing intermediate degrees of variability.

Information theory offers a single quantitative measure of categorical variability—*entropy*. To quantify entropy, one can use the following equation (Shannon & Weaver, 1949):

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

where $H(A)$ is the entropy of array $A$, $a$ is a type of item in $A$, and $p_a$ is the proportion of items of that type within the array. *Same* arrays entail 16 identical icons, so there is only one category with a probability of occurrence of 1.0, yielding an entropy of 0.0. *Different* arrays entail one occurrence of each of 16 icons (in our examples), yielding an entropy of 4.0. The entropy of mixture arrays may be closer to 0.0 or to 4.0, the two *endpoints* of the entropy dimension, depending on the relative frequency and number of icon types. Thus, a same–different discrimination can be represented as a discrimination of variability.
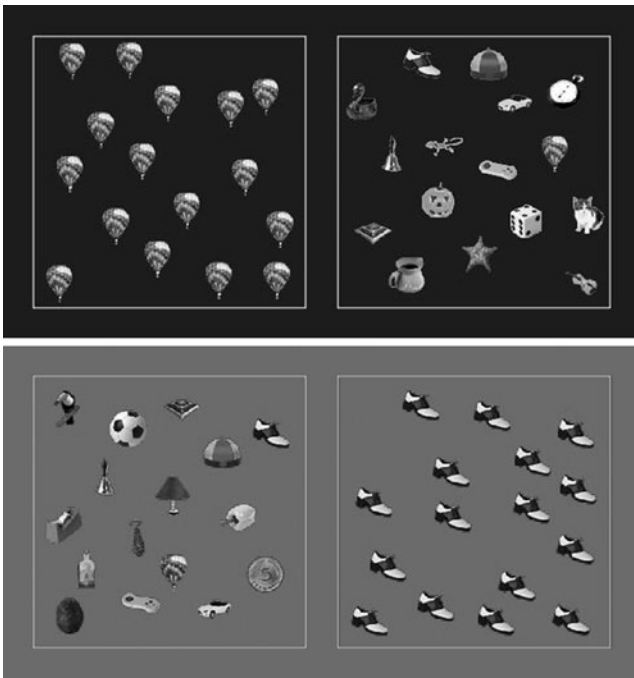
**Fig. 1** Examples of the two types of training trials that were used in the present study. Each of the trials contains one 16-item *same* array and one 16-item *different* array. The color of the background indicates whether the pigeons must choose the *same* array or the *different* array. For example, in the case of the red background (dark gray in this reproduction), the correct choice is the *same* array, whereas in the case of the blue background (light gray in this reproduction), the correct choice is the *different* array. The border of the arrays was black on the computer screen

Young and Wasserman (1997) trained pigeons to make one report response to 16-icon *same* arrays and a second report response to 16-icon *different* arrays. After the birds had mastered the task, they were presented with arrays containing different *mixtures* of identical and nonidentical icons. In three experiments, the pigeons showed a smooth transition in *same* and *different* responding as the mixture arrays changed from almost all same to almost all different. Entropy fit the data remarkably well. Moreover, baboons (Wasserman, Fagot, & Young, 2001a) and even some humans (Young & Wasserman, 2001) showed the same graded changes in their *same* and *different* reports as a function of changes in the mixture of same and different items. Entropy can thus account for pigeons', baboons', and some humans' responding to mixture arrays.

In addition, pigeons, baboons, and humans were tested with arrays comprising fewer than 16 icons; these testing arrays could contain 2, 4, 8, 12, or 14 items, which were either all the same as or all different from one another. The pigeons and baboons exhibited progressive decrements in same–different discrimination accuracy as the number of items was reduced from the training value of 16; accuracy was especially poor with 2- and 4-item arrays (Wasserman, Young, & Fagot, 2001b; Young, Wasserman, & Garner,

1997). Most human subjects' accuracy stayed high for arrays containing different numbers of items, but their reaction time scores disclosed that the 2-item discrimination was much more difficult than was the discrimination of arrays containing more items (Young & Wasserman, 2001; see also Castro, Young, & Wasserman, 2006). Curiously, the pigeons' decrement in performance with fewer items was strongest on *different* trials; their performance on *same* trials remained high.

The notion of visual display entropy can actually explain the curious pattern of behavior that was observed when the number of items involved in the discrimination was decreased. That analysis suggests that, when organisms have been trained to discriminate 16-icon *same* arrays from 16-icon *different* arrays, they will have learned to make one response (*same*) to displays with minimal entropy (0.0) and a second response (*different*) to displays with maximal entropy (4.0). In testing, organisms should distribute their responses as a function of entropy; arrays with entropies closer to 0.0 should be more likely to be classified as *same*, whereas arrays with entropies closer to 4.0 should be more likely to be classified as *different*. The entropy of a 2-item *different* array, 1.0, is thus more similar to the entropy of 16-item *same* arrays, 0.0, than it is to the entropy of 16-item *different* arrays, 4.0. Entropy discrimination should thus prompt classification of 2-item *different* arrays as *same* rather than *different*; empirical findings accord with this unique prediction. These results thus represent an important and counterintuitive confirmation of organisms' use of entropy in this task.

Thus, we see that visual display variability has profound effects on same–different discrimination performance. As was documented above, changes in the variability of multi-item arrays can be brought about in two entirely different ways: (1) by varying the type and frequency of same and different items in the array and (2) by reducing the number of items in *same* and *different* arrays. However, no one has directly compared performance in a same–different discrimination involving the same levels of variability created either by manipulating the combination of same and different items or by changing the number of items in *same* and *different* arrays. According to the entropy account, the particular means of changing variability should have no effect on same–different discrimination performance: Identical variability should yield identical discrimination performance no matter how that level of variability is produced. A direct comparison of these two disparate ways of manipulating variability should allow us to conduct a critical test of the entropy account.

In this investigation, we used a same–different discrimination task in which pigeons were initially trained to distinguish *two* arrays of 16 items that were either all identical to one another (*same* arrays) or all nonidentical

to one another (*different* arrays). Both *same* and *different* arrays appeared simultaneously on the screen, and pigeons had to choose one or the other, depending on the background color of the screen (e.g., Castro, Kennedy, & Wasserman, 2010; Castro & Wasserman, 2011; Flemming, Beran, & Washburn, 2007; examples of these displays are depicted in Fig. 1).

Once the birds had mastered the discrimination between Entropy 0 (*same*) and Entropy 4 (*different*) arrays, we presented them with arrays involving different levels of entropy: 1.0, 2.0, and 3.0 (see examples in Fig. 2 and all of the combinations in Table 1). These levels of entropy were generated either by manipulating the number of items in the arrays containing all same and all different items (number testing) or by maintaining 16 items in the arrays but changing the proportion of same and different items in the array (mixture testing).

If the birds' discrimination behavior is based solely on entropy, then the larger the entropy disparity between the arrays, the better the discrimination should be. Critically, the two different ways of creating different levels of variability should have no material effect on discrimination behavior: If variability is the same, discrimination performance should be the same as well.

## Method

### Subjects

We studied 4 feral pigeons kept at 85% of their free-feeding weights by controlled daily feedings. The pigeons had served in unrelated studies prior to the present investigation. Those studies included categorization of natural and artificial categories, discrimination of geons rotated to different degrees, discrimination of geometrical shapes from the background surrounding them, and learning of symmetrical associations.

### Apparatus

The experiment used four $36 \times 36 \times 41$ cm operant conditioning chambers detailed by Gibson, Wasserman, Frei and Miller (2004). The chambers were located in a dark room with continuous white noise. Each chamber was equipped with a 15-in. LCD monitor located behind an AccuTouch® resistive touchscreen (Elo TouchSystems, Fremont, CA). The touchscreen activation force is typically less than 113 g, according to Elo TouchSystems' technical specifications. The portion of the screen that was viewable by the pigeons was
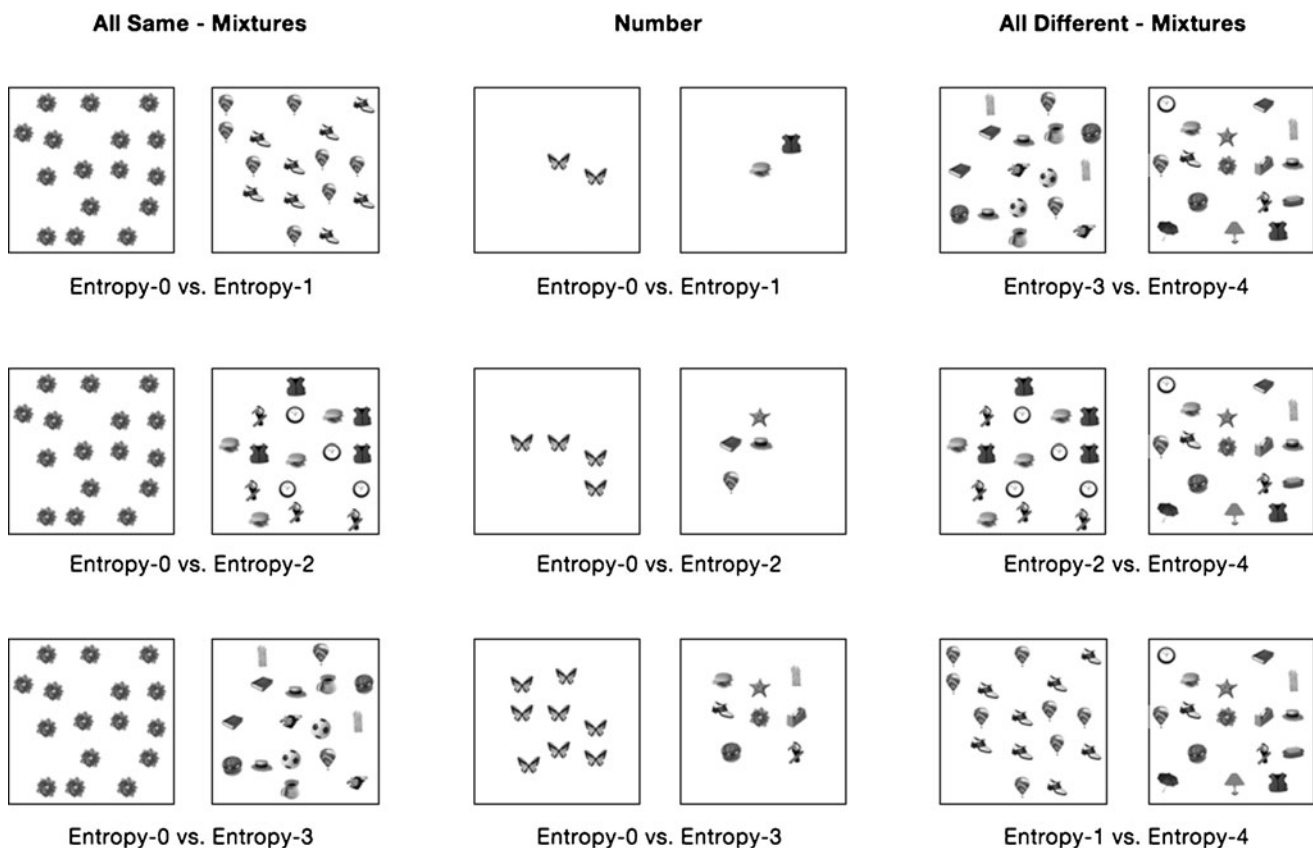


**Fig. 2** Examples of the critical testing combinations. The labels under each pair of arrays indicate the entropy of each of the two arrays. In the actual experiment, the arrays could appear on either a red or a blue background

**Table 1** Entropy and icon arrangement of training and testing trials

|  | Entropy Disparity | Entropy Array 1 | Entropy Array 2 | Arrangement Array 1 | Arrangement Array 2 |
|---|---|---|---|---|---|
| Training | 4 | 4 | 0 | 16 icons | 1 icon × 16 |
| Mixture testing | 1 | 4 | 3 | 16 icons | 8 icons × 2 |
|  | 1 | 3 | 2 | 8 icons × 2 | 4 icons × 2 |
|  | 1 | 2 | 1 | 4 icons × 2 | 2 icons × 8 |
|  | 1 | 1 | 0 | 2 icons × 8 | 1 icon × 16 |
|  | 2 | 4 | 2 | 16 icons | 4 icons × 2 |
|  | 2 | 3 | 1 | 8 icons × 2 | 2 icons × 8 |
|  | 2 | 2 | 0 | 4 icons × 2 | 1 icon × 16 |
|  | 3 | 4 | 1 | 16 icons | 2 icons × 8 |
|  | 3 | 3 | 0 | 8 icons × 2 | 1 icon × 16 |
| Number testing | 1 | 1 | 0 | 2 icons | 1 icon × 2 |
|  | 2 | 2 | 0 | 4 icons | 1 icon × 4 |
|  | 3 | 3 | 0 | 8 icons | 1 icon × 8 |
|  | 3.6 | 3.6 | 0 | 12 icons | 1 icon × 12 |
|  | 4.32 | 4.32 | 0 | 20 icons | 1 icon × 20 |
|  | 4.58 | 4.58 | 0 | 24 icons | 1 icon × 24 |

*Note.* Composition of all training and testing array combinations. The specific type and frequency of icons in each of the two arrays simultaneously presented on the screen (Arrangement Array 1 and Arrangement Array 2) result in specific entropy levels for each of the two arrays (Entropy Array 1 and Entropy Array 2). Entropy disparity is the difference between the entropy levels of the arrays.

28.5 × 17 cm. Pecks to the touchscreen were processed by a serial controller board outside the box. A rotary dispenser delivered 45-mg pigeon pellets through a vinyl tube into a Plexiglas food cup located in the center of the rear wall opposite the touchscreen. Illumination during the experimental sessions was provided by a houselight mounted on the upper rear wall of the chamber. The pellet dispenser and houselight were controlled by a digital I/O interface board. Each chamber was controlled by an Apple® iMac® computer. The program to run the experiment was developed in MATLAB®.

Stimuli and experimental design

A total of 72 color icons selected from an image database (Hemera Technologies Inc.) constituted the total item pool. Each of the icons occupied an area of 1 × 1 cm. Special care was taken to ensure that the icons were readily distinguishable from one another and that no set contained a disproportionate number of icons of a specific color, shape, or size. For any given *same* array, a single icon was randomly chosen from the pool and then used to create an array of 16 identical items; for any given *different* array, sixteen icons were randomly chosen from the pool and then were used to create an array of 16 nonidentical items. The icons were randomly distributed to 16 of 25 possible locations in an invisible 5 × 5 grid composed of 2 × 2 cm cells (so the size of each array was 10 × 10 cm). Thus, 16 of the 25 locations contained icons, and 9 were blank. In addition, each of the 1 × 1 cm icons could be positioned in any part of each of the 2 × 2 cm cells. This

distribution procedure resulted in spatially disordered stimulus arrays in which vertical or horizontal alignment of the icons was disrupted (see Fig. 1). This distribution method was used during both training and testing.

For testing different numbers of items, because the 5 × 5 matrix into which the icons would be placed might contain as few as 2 items, we used a method of placement in which the icons were always immediately adjacent to one another—vertically, horizontally, or diagonally—in order to prevent the icons from being located far apart. This grouping technique used the central matrix position in all of the arrays. To create arrays of 2, 4, or 8 icons, 1, 3, or 7 added icons, respectively, were randomly located in the remaining eight positions of the central 3 × 3 area. To create arrays of 12, 16, 20, or 24 icons, the central 3 × 3 area was filled with 9 icons; 3, 7, 11, or 15 added icons, respectively, were randomly located in the remaining positions of the 5 × 5 grid.

On each training trial, two arrays—one *same* and one *different*—appeared simultaneously on the screen (see Fig. 1). The arrays were located on the left and the right sides of the screen, 8 cm apart. Both of the arrays were circumscribed by black borders to emphasize their separateness. For half of the trials, the *same* array was on the left, and the *different* array was on the right; the reverse was true for the other half of the trials. *Same* arrays contained 16 identical icons, so their entropy level was 0.0. *Different* arrays contained 16 nonidentical icons, so their entropy level was 4.0. Thus, training involved the discrimination between Entropy 0 and Entropy 4 arrays.

Once the birds mastered this discrimination, we presented them with combinations of arrays involving intermediate levels of entropy: 1.0, 2.0, and 3.0. These three intermediate levels of entropy were created either by manipulating the number of items in the arrays containing all same and all different items (number testing) or by maintaining 16 items in the arrays but manipulating the proportion of same and different items within the arrays (mixture testing). All of the combinations are shown in Table 1, and examples are displayed in Fig. 2.

For number testing, all of the arrays still contained either all same or all different icons. For *same* arrays, entropy was always 0 regardless of the number of items in the array, but entropy changed for 2-item (entropy = 1.0), 4-item (entropy = 2.0), and 8-item (entropy = 3.0) *different* arrays. We also included arrays containing 12 (entropy = 3.6), 20 (entropy = 4.32), and 24 (entropy = 4.58) items, although these arrays yielded entropy disparities that did not have an equivalent value in mixture testing. These combinations yielded six different testing entropy disparities—1.0, 2.0, 3.0, 3.6, 4.32, and 4.58—in addition to the 4.0 training entropy value, between the simultaneously presented arrays.

For mixture testing, 16-item arrays could contain two icons shown 8 times each (entropy = 1.0), four icons shown 4 times each (entropy = 2.0), or eight icons shown 2 times each (entropy = 3.0), in addition to the training arrays (entropy 0.0 vs. entropy 4.0). All possible pairwise combinations of these five different entropy mixture levels were shown to the pigeons: 0–1, 0–2, 0–3, 1–2, 1–3, 1–4, 2–3, 2–4, and 3–4, in addition to the 0–4 training combination. These combinations yielded four different entropy disparities—1.0, 2.0, 3.0 (testing values), and 4.0 (training value)—between the simultaneously presented arrays.

Procedure

*Training* Daily training sessions comprised 144 trials. At the start of a trial, the pigeons were presented with an orienting stimulus: a black cross on a white display in the middle of the screen. Following one peck anywhere on the display, the *same* and *different* arrays appeared on the red or blue colored background. The pigeon's task was to peck directly at either the *same* array or the *different* array. The color of the background served as a conditional cue signaling which array was correct. Half of the birds were required to peck at the *same* array when the background color was blue and to peck at the *different* array when the background color was red; the assignments were reversed for the other half of the birds.

Differential food reinforcement was used to encourage correct responses. If the choice response was correct, food was delivered, and a 5-s intertrial interval (ITI) ensued, during which the screen blackened and the pigeon awaited the next trial. If the choice response was incorrect, there was no food, and a time-out period began. The initial time-out

duration was 5 s; it was increased by 2 s every day until the bird met criterion. After the time-out, correction trials were given until the correct response was made. Correction trials involved repeating exactly the same trial on which the bird had just erred. Only the *first* report response of a trial was scored and used in data analysis. When all of the pigeons reached the 85% correct criterion for both same-cued trials and different-cued trials, they were moved to testing.

*Testing* A total of 24 daily sessions took place in which mixture testing and number testing were given on alternate days. All sessions began with 32 warm-up training trials (which were not included in the statistical analysis). In mixture testing sessions, warm-up trials were followed by 148 additional trials, 36 of which were randomly interspersed testing trials; in number testing sessions, warm-up trials were followed by 136 additional trials, 24 of which were randomly interspersed testing trials. On training trials, arrays of Entropy 0.0 (*same* array, all 16 identical items) and Entropy 4.0 (*different* array, all 16 nonidentical items) were simultaneously presented on the screen. In mixture testing sessions, in addition to the Entropy 0.0 and Entropy 4.0 arrays, we presented mixture arrays displaying intermediate levels of entropy: 1.0 (two icons, 8 times each), 2.0 (four icons, 4 times each), and 3.0 (eight icons, 2 times each), so that different entropy disparities and all possible pairwise combinations of the five different entropy levels were shown: 0–1, 0–2, 0–3, 1–2, 1–3, 1–4, 2–3, 2–4, and 3–4. These nine testing combinations were presented 4 times each in each of the mixture testing sessions. In number testing, in addition to training trials, we presented *same* and *different* testing arrays containing 2, 4, 8, 12, 20, and 24 items. Simultaneously presented number arrays always contained the same number of items, so that, for instance, a 4-item *same* array was always presented along with a 4-item *different* array; we did not combine arrays containing different numbers of items.

On half of the testing trials, the colored background cued the pigeons to make a *same* response, whereas on the other half of the testing trials, the colored background cued the pigeons to make a *different* response. Because mixture testing trials involved arrays containing some *same* and some *different* arrays (Entropy 1, 2, and 3 arrays), a *same* response was considered to be the choice of the array with the *lower* entropy level, whereas a *different* response was considered to be the choice of the array with the *higher* entropy level. On training trials, only the correct response was reinforced; incorrect responses were followed by correction trials (*differential* reinforcement). On testing trials, *any* choice response was reinforced (*nondifferential* reinforcement); *no* correction trials were given. Nondifferential reinforcement is unlikely to result in the birds' learning anything systematic about the novel entropy combinations that were presented in testing.

In all of the reported tests of statistical significance, an alpha level of .05 was adopted.

## Results

Training to criterion took a mean of 12.25 sessions. Mean discrimination performance started near the chance level of 50% and rapidly reached the 85% level of accuracy on both same-cued trials and different-cued trials. Two birds reached the 85% accuracy level in 10 days; the other 2 birds did so in 13 and 16 days.

Results for mixture and number testing are shown in Fig. 3; as was predicted, when either the item mixtures or the numbers of items in the arrays were varied, accuracy increased as the disparity between the entropy levels of the two simultaneously presented arrays increased. This increase in accuracy was similar for same-cued trials and different-cued trials. Because number testing and mixture testing included different entropy disparity combinations (number testing, 1.0, 2.0, 3.0, 3.6, 4.0, 4.32, and 4.58; mixture testing, 1.0, 2.0, 3.0, 4.0), we first separately analyzed the effect of entropy disparity in number testing sessions and in mixture testing.

A repeated measures 4 (entropy disparity) × 2 (trial type: same-cued vs. different-cued) × 12 (sessions) analysis of variance (ANOVA) on choice accuracy during mixture testing revealed a significant main effect of entropy disparity, $F(3, 9) = 129.41$, $MSE = 0.13$, $p < .001$. Accuracy was 64%, 74%, 84%, and 89% for entropy disparities of 1, 2, 3, and 4, respectively; as was expected, the larger the entropy disparity between the arrays, the better was discrimination performance. There was no main effect of type of trial, nor did type of trial interact with any other factor, all $F$s < 1, indicating that the birds' overall accuracy was similar for same-cued trials (78%) and different-cued (77%) trials.

Finally, there was no significant effect of session, nor did session interact with any other factor, all $F$s < 1. Thus, there were no reliable changes in accuracy across the 12 testing sessions; the birds responded similarly to the new mixture testing combinations regardless of the amount of experience that they had with them.

A repeated measures 7 (entropy disparity) × 2 (trial type: same-cued vs. different-cued) × 12 (sessions) ANOVA on choice accuracy during number testing revealed a significant main effect of entropy disparity, $F(6, 18) = 60.88$, $MSE = 0.82$, $p < .001$. Accuracy was 59%, 66%, 79%, 86%, 90%, 92%, and 90% for entropy disparities of 1, 2, 3, 3.6, 4, 4.32, and 4.58, respectively. Just as with mixture testing, the larger the entropy disparity between the arrays, the better discrimination performance was. Nonetheless, there was a limit to the birds' improvement in performance: Beyond an entropy disparity of 3.6, planned comparisons disclosed that there was no significant increase in the pigeons' accuracy, $F$s < 1.

We also found a significant effect of type of trial: Accuracy was slightly higher for different-cued trials (81%) than for same-cued trials (77%), $F(1, 3) = 18.35$, $MSE = 0.01$, $p = .02$. However, type of trial did not interact with any other factor, all $F$s < 1; critically, it did not interact with entropy disparity, $F(6, 18) = 0.66$, $MSE = 0.09$, so that the difference in accuracy between same-cued trials and different-cued trials was not due to lower accuracy to arrays containing any specific number of items. Finally, there was no significant effect of session, nor did session interact with any other factor, all $F$s < 1. Thus, there were no reliable changes in accuracy across the 12 testing sessions; just as in mixture testing, the amount of experience with the new number combinations did not affect the birds' accuracy.

Because we were primarily interested in whether different ways of creating the same levels of array variability measurably affected the birds' performance, our next



Fig. 3 Mean percentage of correct responses in mixture testing (left) and number testing (right) according to entropy disparity. For presentation purposes, we collapsed the data across all 12 number testing sessions and all 12 mixture testing sessions. The entropy disparity of 4.0 is the training value; it is included in the graphs for comparison purposes
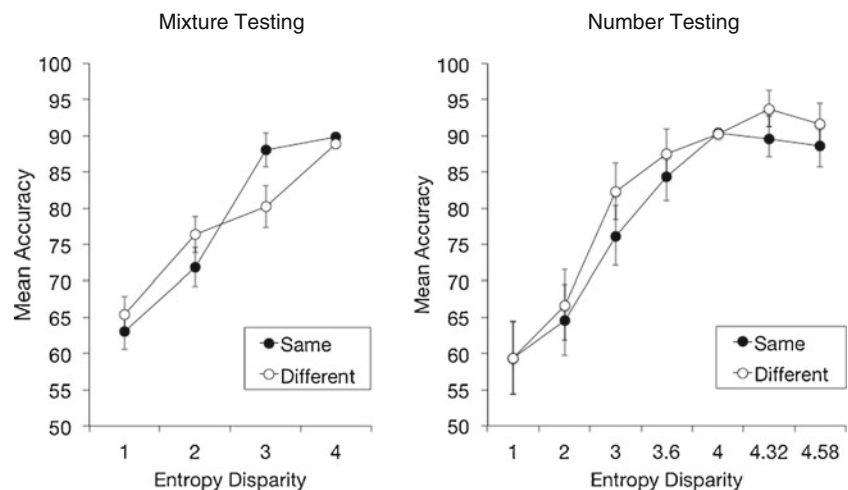
analysis focused on mixture and number combinations that portrayed the same entropy disparities: 1.0, 2.0, and 3.0. So, number trials displaying entropy disparity levels of 3.6, 4.32, and 4.58 were not included in the analysis, because there were no corresponding mixture trials with the same entropy disparities. In addition, we dropped trials involving an entropy disparity of 4.0 (the training value); because these trials were identical in the number and mixture conditions (in each case, the arrays comprised 16 items: all same and all different), they could not contribute to any disparities in mixture and number testing performance.

A repeated measures 3 (entropy disparity) × 2 (trial type: same-cued vs. different-cued) × 2 (testing type: mixture vs. number) × 12 (sessions) ANOVA on choice accuracy revealed a significant main effect of entropy disparity, $F(2, 6) = 43.32$, $MSE = 0.13$, $p < .001$, again showing that overall accuracy in the testing phase was determined by the entropy disparity between the simultaneously presented arrays. Accuracy gradually rose as the disparity between the entropy levels of the arrays increased; the rise was highly similar for both same-cued trials and different-cued trials (there was no statistical difference between same-cued and different-cued trials; $F < 1$). Mean accuracy percentages for the entropy disparity levels of 1.0, 2.0, and 3.0, were 61%, 68%, and 82% for same-cued trials and 62%, 71%, and 81% for different-cued trials.

There was also a main effect of testing type; overall accuracy in mixture testing (74%) was higher than overall accuracy in number testing (68%), $F(1, 3) = 11.10$, $MSE = 0.14$, $p < .05$. So, the particular means of changing variability—by varying the type and frequency of same and different items in the array (mixture testing) or by changing the number of items within the *same* and *different* arrays (number testing)—did have an effect on same–different discrimination

performance, contrary to the predictions of a straightforward entropy account. No other effects or interactions were statistically significant.

We then proceeded to further investigate the discrepancy between mixture and number trials. We noticed that, although number testing trials always involved arrays containing all identical items (*all same*) and all nonidentical items (*all different*), this was not the case for mixture testing trials, which could involve either all-same (0 vs. 1, 0 vs. 2, 0 vs. 3) or all-different (4 vs. 3, 4 vs. 2, 4 vs. 1) arrays. For this reason, we conducted two different number–mixture comparisons: On the one hand, we compared number testing trials with all-same–mixture-testing trials, and on the other hand, we compared number-testing trials with all-different–mixture-testing trials.

Even when the entropy disparity was the same, accuracy was higher for all-same–mixture trials than for number trials (Fig. 4, left), but accuracy was very similar for all-different–mixture trials and number trials (Fig. 4, center). Indeed, when we compared all-same–mixture trials and all-different–mixture trials, we found that accuracy for these two types of trials was different as well, with the former higher than the latter (Fig. 4, right).

In order to confirm these observations, we performed a repeated measures 3 (testing combination: all-same–mixture vs. all-different–mixture vs. number) × 3 (entropy disparity) × 2 (trial type: same-cued vs. different-cued) × 12 (sessions) ANOVA on choice accuracy. This analysis yielded a main effect of entropy disparity, $F(2, 6) = 34.57$, $MSE = 0.18$, $p < .001$, again showing that, the larger the entropy disparity, the better the accuracy. More interesting at this point, the disparity between testing combinations was significant as well, $F(2, 6) = 36.73$, $MSE = 0.06$, $p < .001$. Planned comparisons showed that accuracy was higher for all-same–mixture trials
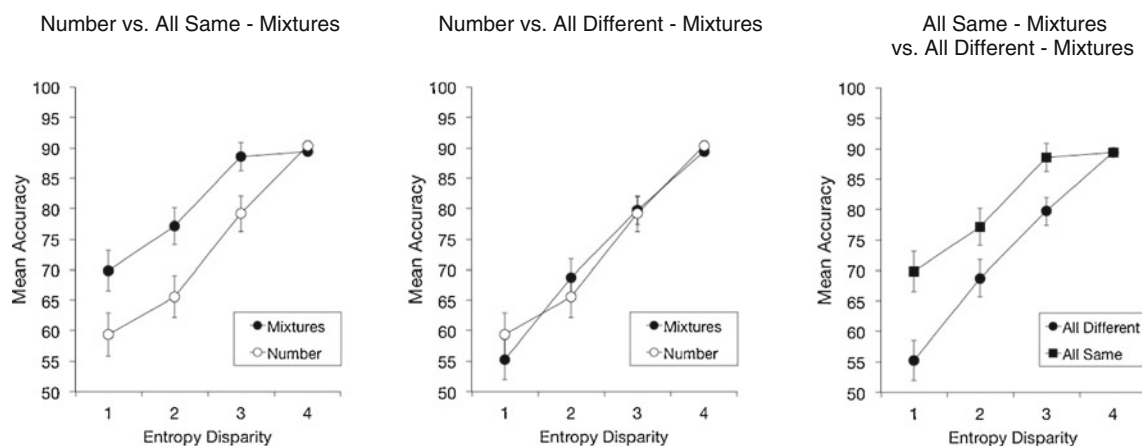


**Fig. 4** Mean percentage of correct responses for the critical testing combinations according to entropy disparity. For presentation purposes, we collapsed the data across all 12 number testing sessions and all 12 mixture testing sessions. Left: Comparison between number trials and all-same–mixture trials. Center: Comparison between number trials and all-different–mixture trials. Right: Comparison between all-same–mixture trials and all-different–mixture trials. The entropy disparity of 4.0 is the training value; it is included in the graphs for comparison purposes

than for number trials, $F(1, 6) = 54.18$, $MSE = 0.52$, $p < .001$. Accuracy was very similar for all-different–mixture trials and number trials, $F(1, 6) = 0.01$, $MSE = 0.01$, $p > .1$. And accuracy was higher for all-same–mixture trials than for all-different–mixture trials, $F(1, 6) = 56.00$, $MSE = 0.54$, $p < .001$.

How might we explain this pattern of responding? The difference in accuracy between all-same–mixture and all-different–mixture trials can be accommodated by an entropy account. We previously found that same–different accuracy in this conditional task depends not only on the entropy disparity between the arrays involved in the discrimination, but also on the location of each of the two arrays along the entropy scale (Castro & Wasserman, 2011; see also Young & Wasserman, 2002). That is, within each entropy disparity level, discrimination is easier when the two entropy values are at the lower end of the scale than when the two entropy values are at higher points along the scale; for example, when the entropy disparity is 1.0, choice accuracy is higher when Entropy 0 and Entropy 1 arrays are presented than when Entropy 3 and Entropy 4 arrays are presented (in Castro & Wasserman, $M = 74\%$ and $M = 55\%$, respectively; in the present experiment, $M = 72\%$ and $M = 55\%$, respectively). This pattern of performance is consistent with a *logarithmic* relationship between entropy and discriminative behavior; it can be effectively captured by an entropy account if the entropy values are transformed into a logarithmic scale.

We thus recalculated the entropy disparity by using the logarithmic transformation of the specific entropy values involved in the discrimination, and we fit a logistic function to the data from all-same–mixture trials and all-different–mixture trials (see Castro & Wasserman, 2011, for more details about this method). This nonlinear fitting technique included one free parameter: gradient (the steepness of the slope of the logistic function, corresponding to the increment in discrimination accuracy as the disparity in entropy increases). Using nonlinear regression, we fit the data using this equation:

$$\frac{1}{1 + e^{-gradient \times (LogEntropyDisparity)}} \quad (2)$$

The estimate of gradient that produced the best goodness of fit was 5.78. The regression accounted for 98% of the variance, and it was statistically significant, $F(1, 6) = 353.58$, $MSE = 0.141$, $p < .001$. The fitted curve is depicted in Fig. 5. (The values for the corresponding number trials are also included as individual unfitted points, for comparison purposes.) Thus, the pigeons' pattern of responding on trials involving all-same–mixture arrays and all-different–mixture arrays can be nicely described by entropy disparity if the specific values that are used to create the disparity are spaced logarithmically rather than linearly.
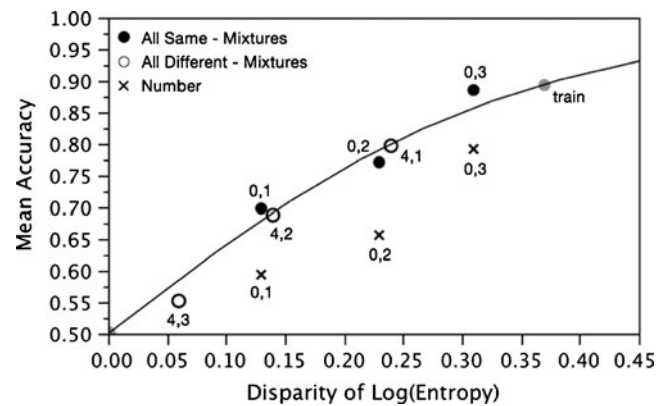


**Fig. 5** Fit of choice accuracy (probability of correct response) for all-same–mixture trials (entropy disparities: 0 vs. 1, 0 vs. 2, and 0 vs. 3), all-different–mixture trials (entropy disparities: 4 vs. 3, 4 vs. 2, and 4 vs. 1), and number trials (entropy disparities: 0 vs. 1, 0 vs. 2, and 0 vs. 3), according to the logarithmic transformation of the entropy disparity between the simultaneously presented arrays. All mixture arrays contained 16 items; the number arrays contained 2 items (0 vs. 1), 4 items (0 vs. 2), or 8 items (0 vs. 3). See description in the text

We should note that Young and Wasserman (1997) reported that the relationship between their birds' same–different discrimination behavior and entropy could be largely explained by a linear fit. But a later reanalysis of those data (reported by Young & Wasserman, 2002) suggested that a logarithmic function provided an even better fit. Thus, as Young and Wasserman (2002) suggested, a logarithmic relationship between entropy and same–different discrimination was actually present even in the early experiments. Later studies (Castro & Wasserman, 2011; Young & Wasserman, 2002), as well as the present experiment, confirm the greater explanatory value of a logarithmic relationship between entropy and same–different discriminative behavior.

However reassuring that close fit may be for entropy explaining key aspects of the pigeons' same–different discrimination behavior (cf. Fig. 4, right, and Fig. 5), it turns out that this logarithmic transformation does not help to explain the discrepancy between number and all-same–mixture-testing arrays (see Fig. 5). In this case, within each of the entropy disparity levels, the arrays involved in the discrimination portrayed identical values (0 vs.1, 0 vs. 2, 0 vs. 3, and 0 vs.4), so that they are placed at identical locations along the logarithmic scale. Nonetheless, accuracy is clearly not equivalent. We address possible explanations for this striking discrepancy in the upcoming Discussion section.

## Discussion

After initial training with multiitem arrays located at the extreme values along our entropy continuum (Entropy 0 vs. Entropy 4), pigeons readily responded to new testing

combinations containing intermediate entropy levels. In testing with these intermediate entropy combinations, we found that, as the disparity in entropy between the arrays increased, the pigeons' choice accuracy progressively increased (Fig. 3). This relation held true for two distinctly different ways of creating those entropy disparities: by varying the number of items on *same* and *different* testing trials and by varying the combination of same and different items on mixture testing trials.

In addition, at each entropy disparity level created by various same–different mixtures, the birds found display pairs involving lower entropy values (all-same–mixture trials) to be more discriminable than display pairs involving higher entropy values (all-different–mixture trials) (Fig. 4, right). These results closely replicate the recent findings of Castro and Wasserman (2011).

An entropy account can explain the increase in accuracy as the result of the increase in the disparity in entropy between the choice arrays. Also, the representation of entropy along a logarithmic scale can explain why an entropy disparity that is created from values nearer the low end of the entropy scale (all-same–mixture trials) yields better discriminative performance than does the same entropy disparity that is created from values nearer the high end of the entropy scale (all-different–mixture trials) (cf. Fig. 4, right, and Fig. 5).

As we argued in Castro and Wasserman (2011; see also Young & Wasserman, 2002), a logarithmic relationship between entropy and same–different discrimination behavior is consistent with the idea that distinguishing *same* from *different* collections of items follows Weber's law. According to Weber (1834/1996), for the disparity in the intensities of two stimuli to be noticeable, this disparity should be proportional to the physical intensities of the stimuli; the more intense two stimuli, the greater the disparity between them ought to be in order to be distinguishable. In our case, the greater the entropy of two same–different arrays, the greater the disparity between their individual entropy values should be in order to be discriminable. Thus, equivalent entropy disparities should be more discriminable at the lower end of the entropy scale (where there is less variability) and less discriminable at the upper end of the entropy scale (where there is more variability); our results strongly support this notion. Thus, the pigeon's abstract same–different concept can better be deemed to be *dimensional* than *categorical*, much like such physical dimensions as weight, size, or brightness and such other abstract dimensions as number (see Cantlon & Brannon, 2006, for the application of Weber's law to numerical cognition).

However, this entropy account cannot explain the disparity in accuracy that arose when the entropy values of the arrays involved in the discrimination were the same (1 vs. 0, 2 vs. 0, and 3 vs. 0) but were created either by manipulating the number of icons (number trials) or by manipulating the combination of same and different icons within the array (all-same–mixture trials) (Fig. 4, left).

The role of generalization decrement

Because our birds had originally been trained with 16-item *same* and *different* arrays, it could be argued that their poorer performance with 2-, 4-, and 8-item number testing arrays than with 16-item mixture testing arrays resulted from generalization decrement based on the smaller quantity of items in the number testing arrays. However, we have found in previously published research that poor performance with few-item arrays is not due solely to stimulus generalization decrement. For instance, when Young et al. (1997) trained pigeons with arrays containing different numbers of icons from the outset of the experiment, the birds still failed to learn to discriminate *same* from *different* arrays that involved only two or four icons; indeed, the pigeons even more poorly discriminated arrays with a larger number of icons than if those few-item arrays had not been concurrently presented.

In Young et al. (1997), a two-key forced choice procedure was used, in which the birds were presented with either a *same* array or a *different* array on the screen and they had to select one of two arbitrary report buttons. In Castro et al. (2010), we used the same conditional discrimination procedure as in the present experiment; after training the birds with 16-item *same* and *different* arrays, we proceeded to train them in a stepwise fashion with arrays containing fewer icons (Experiment 3). First, we included arrays containing 2, 4, 6, 8, 10, and 12 icons (20 daily sessions); next, we presented arrays containing 2, 4, and 6 icons (20 daily sessions); then, we presented arrays containing 2 and 4 icons (40 daily sessions); and, finally, we presented arrays with only 2 icons (40 daily sessions). After such extensive training, the birds' highest accuracy for 2-, 4-, and 8-item arrays was 64%, 83%, and 95%, respectively. When these same birds were later trained with 16-item mixture arrays, the combinations portraying the same entropy values (1 vs. 0, 2 vs. 0, and 3 vs. 0) yielded even higher accuracy levels: 75%, 93%, and 97%, respectively (Castro & Wasserman, 2011). Again, accuracy was lower for number trials than for mixture trials, exactly the same pattern that we obtained in the present study. Thus, it does not appear that the mere novelty of the arrays containing a smaller number of items is the reason for the discrepancy in accuracy between number and mixture arrays involving identical levels of entropy.

The role of redundant information

We next tried to identify the characteristics of the number and mixture trials that might be influencing the pigeons' performance. All of the number testing trials contain all-same arrays; but, in the case of the mixture testing trials, the all-same arrays always contain 16 items, whereas in the case of the number testing trials, the all-same arrays contain 2, 4, or 8 items. Because entropy is simply a measure of variability, it does not predict any change in performance for all-same arrays containing different number of items; in these cases, entropy is always 0, regardless of the array containing 2, 4, 8, or 16 items.

In point of fact, if accuracy were to fall when the number of items is decreased, this drop ought to be due to the failure to classify the *different* arrays as *different.* The entropy of a 2-item *different* array is 1.0; so, this array should be more similar to the 16-item *same* training array, 0.0, than it is to the 16-item *different* training array, 4.0. Our prior empirical findings confirmed this asymmetry. When the number of items was reduced from the training value of 16, the decrement in accuracy was observed on *different* trials only; accuracy on *same* trials involving a small number of icons remained high (Young & Wasserman, 2001; Young et al., 1997; Wasserman et al., 2001a).[1]

Intuitively, however, sameness, as well as differentness, ought to be easier to perceive when many items instantiate that relation than when few items do so. Indeed, considerable research with the redundant-target paradigm has revealed a performance advantage when information is provided by a larger number of items (e.g., Miller, 1982; Mordkoff & Yantis, 1991; Van der Heijden, La Heij, & Boer, 1983).

In a redundant-target task, performance on trials on which only one target is given is compared with performance on trials on which two or more identical targets are given. For example, subjects may have to detect the presence of a target letter, X. A reaction time advantage is normally observed when a trial contains two Xs, as compared with when a trial contains only a single X (cf. Kinchla, 1974). According to *coactivation* models (e.g., Miller, 1982), responses to redundant signals are especially fast, because two or more sources provide additional activation

to trigger the appropriate response. Activation naturally grows faster when it is provided by several sources rather than by only one.

In the case of our *same* and *different* arrays—where all of the relationships among the icons in a given array are of sameness or differentness—we might say that, as the number of items increases, the number of relations of sameness or differentness also increases. In a 2-item array, there is only one relationship involving those 2 individual items; but, in a 16-item array, multiple relationships are present, for each of the items with all of the other icons. The larger the number of relationships, the greater the sources of activation for the appropriate response, yielding better performance as the number of items increases.

Although we noted above that our prior results have shown that accuracy for *same* arrays did not suffer due to the reduction in the number of items, there was one case in which performance on *same* arrays was impaired when fewer items were presented. This impairment occurred when Young, Wasserman, Hilfers and Dalrymple (1999) trained pigeons to discriminate *successive lists* of same and different items, in which only one item at a time was given. There, reducing the list length had an equivalent detrimental effect on pigeons' accuracy on *both same* trials and *different* trials. With successively presented items, more information was better for both *same* and *different* lists.

Young et al. (1999) concluded that, in addition to entropy, what they called an "evidence accumulation process" also influenced the birds' discrimination performance. Although the accumulation of evidence was supposed to occur because the information was presented in a sequential way, it is conceivable that this process of accumulation occurs even with the simultaneous presentation of the information, in an automatic and parallel way, as the coactivation model argues. If so, then even when the entropy values of the arrays involved in the combination is the same, accuracy for all-same–mixture arrays may be higher because sameness in the 16-item *same* array is easier to perceive than sameness in the arrays containing fewer numbers of items.

Perhaps there is something about the simultaneous presentation of *two* multiitem arrays in the conditional same–different task that makes performance more susceptible to an "evidence accumulation process" than in the more familiar same–different task, in which a single multiitem array is presented and reports of *same* and *different* are required to arbitrary response options. If that were the case, it might help us understand why identical disparities in entropy lead to discrepancies in choosing the more or less variable array conditional on a superordinate cue.

The obvious candidates are that (1) the pigeons on the conditional same–different task can directly compare one array with the other (something that is not possible with the single array procedure) and (2) the pigeons on the

---

[1] The conditional same–different discrimination task that we used in the present experiment does not permit one to observe the asymmetry between two-item *same* and *different* arrays that has often been reported. On both same-cued trials and different-cued trials, one *same* array and one *different* array are presented simultaneously on the screen, and the pigeons have to select the *same* or the *different* array depending on the color of the background. Thus, pigeons should have difficulty choosing correctly on *both* same-cued and different-cued trials, because on both types of trials they have to choose between an array of Entropy 0.0 (*same* array) and an array of Entropy 1.0 (*different* array).

conditional same–different task must directly respond to the more or less variable array conditional on a superordinate cue (rather than responding to one of two spatially and visually distinctive "more" or "less" variable report keys with the single array procedure). What is not yet clear is why or how those task differences leverage disparate discrimination processes.

Final considerations

The fact that entropy—a simple mathematical measure of visual display variability—cannot explain all of our present results does not invalidate the idea that variability lies at the root of same–different discrimination of our multiitem arrays. Prior findings such as the graded change in responding as the displays progressively change from being all same to all different (observed in animals and some humans; Wasserman et al., 2001b; Young & Wasserman, 1997, 2001) and the keen sensitivity to the number of same and different items (observed in animals and most humans; Castro et al., 2006; Flemming et al., 2007; Young et al., 1997) strongly suggest that learning of the same–different concept may be based on the discrimination of perceptual variability.

Nevertheless, other factors, which an entropy explanation does not take into account, may also affect the learning of a same–different discrimination. Such factors include the degree of similarity among the items involved in the discrimination (Smith, Redford, Haas, Coutinho, & Couchman, 2008; Young, Wasserman, & Ellefson, 2007) and the spatial distance among the visually presented items (Young et al., 2007). Indeed, Young et al. (2007) developed the finding differences model expressly to accommodate the roles that are played by similarity and spatial distance in same–different discrimination performance.

Thus, entropy might not be the whole story, but it is certainly a substantial part of it. Same–different discrimination behavior is far from being a simple cognitive feat; all of the complexities and intricacies underlying this fundamental feat of cognition are still to be ascertained and fully explained.

## References

Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science, 17,* 401–406.

Castro, L., Kennedy, P. L., & Wasserman, E. A. (2010). Conditional same–different discrimination by pigeons: Acquisition and generalization to novel and few-item displays. *Journal of Experimental Psychology. Animal Behavior Processes, 36,* 23–38.

Castro, L., & Wasserman, E. A. (2011). The dimensional nature of same–different discrimination behavior in pigeons. *Journal of Experimental Psychology. Animal Behavior Processes, 37,* 361–367.

Castro, L., Young, M. E., & Wasserman, E. A. (2006). Effects of number of items and visual display variability on same–different discrimination behavior. *Memory & Cognition, 34,* 1689–1703.

Flemming, T. M., Beran, M. J., & Washburn, D. A. (2007). Disconnect in concept learning by rhesus monkeys (*Macaca mulatta*): Judgment of relations and relations-between-relations. *Journal of Experimental Psychology. Animal Behavior Processes, 33,* 55–63.

Gibson, B. M., Wasserman, E. A., Frei, L., & Miller, K. (2004). Recent advances in operant conditioning technology: A versatile and affordable computerized touch screen system. *Behavior Research Methods, Instruments, & Computers, 36,* 355–362.

Kinchla, R. (1974). Detecting target elements in multielement arrays: A confusability model. *Perception & Psychophysics, 15,* 149–158.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14,* 247–279.

Mordkoff, J. T., & Yantis, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology. Human Perception and Performance, 17,* 520–538.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Smith, J. D., Redford, J. S., Haas, S. M., Coutinho, M. V. C., & Couchman, J. J. (2008). The comparative psychology of same–different judgments by humans (*Homo sapiens*) and monkeys (*Macaca mulatta*). *Journal of Experimental Psychology. Animal Behavior Processes, 34,* 361–374.

Van der Heijden, A. H. C., La Heij, W., & Boer, J. P. A. (1983). Parallel processing of redundant targets in simple search tasks. *Psychological Research, 45,* 235–254.

Wasserman, E. A., Fagot, J., & Young, M. E. (2001a). Same–different conceptualization by baboons (*Papio papio*): The role of entropy. *Journal of Comparative Psychology, 115,* 42–52.

Wasserman, E. A., & Young, M. E. (2010). Same–different discrimination: The keel and backbone of thought and reasoning. *Journal of Experimental Psychology. Animal Behavior Processes, 36,* 3–22.

Wasserman, E. A., Young, M. E., & Fagot, J. (2001b). Effects of number of items on the baboon's discrimination of same from different visual displays. *Animal Cognition, 4,* 163–170.

Weber, E. H. (1996). *On the tactile senses*. Hove: Erlbaum. Original work published 1834.

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. (2001). Entropy and variability discrimination. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 2,* 278–293.

Young, M. E., & Wasserman, E. A. (2002). The pigeon's discrimination of visual entropy: A logarithmic function. *Animal Learning & Behavior, 30,* 306–314.

Young, M. E., Wasserman, E. A., & Ellefson, M. R. (2007). A theory of variability discrimination: Finding differences. *Psychonomic Bulletin & Review, 14,* 805–822.

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.

Young, M. E., Wasserman, E. A., Hilfers, M. A., & Dalrymple, R. (1999). The pigeon's variability discrimination with lists of successively presented visual stimuli. *Journal of Experimental Psychology. Animal Behavior Processes, 25,* 475–490.