

The repetition discrimination task: An objective method for studying perceptual grouping

STEPHEN E. PALMER

University of California, Berkeley, California

AND

DIANE M. BECK

University of Illinois, Urbana-Champaign, Illinois

Five experiments are reported that demonstrate the use of the repetition discrimination task (RDT) to study perceptual grouping effects objectively and quantitatively. Experiments 1 and 3 validate the method by measuring grouping based on proximity, color similarity, common region, and element connectedness. Experiment 2 compares the RDT effects for proximity grouping to explicit subjective ratings of grouping strength in identical displays. Experiments 4 and 5 investigate the effects of size and orientation of surrounding ovals in displays in which competing organizations are present. In each case, the RDT produces clear, consistent patterns of response times that are consistent with predictions based on grouping. It thus represents an objective method for studying the full range of grouping phenomena originally described by Wertheimer.

In 1923, Max Wertheimer called attention to one of the fundamental problems of visual perception: How does the visual system organize the complex array of light striking the retinal mosaic into the coherent spatial array of objects that we experience when we view the world? Although Wertheimer did not by any means solve this complex and difficult problem, he did make substantial inroads by identifying a number of stimulus-driven principles of grouping. These include the well-known factors of proximity, similarity, good continuation, and common fate that are discussed in virtually all textbooks on vision (see, e.g., Palmer, 1999). Other principles of grouping have since been added to this list, including common region (Palmer, 1992), element connectedness (Palmer & Rock, 1994a, 1994b), and synchrony (Blake & Yang, 1997), but Wertheimer's original article is still the definitive work.

The method Wertheimer (1923) used to investigate perceptual grouping was phenomenological demonstration. He published displays that illustrated the principles he proposed and left it up to readers to decide whether or not their phenomenology on viewing the displays agreed with the descriptions he gave in the accompanying text. Obviously, most people agreed, for Wertheimer's principles of grouping are among the oldest, best-known, and most widely cited findings in perceptual psychology. The implicit agreement among his readers was sufficient for his purpose, which was simply to demonstrate the effectiveness of isolated stimulus factors in organizing individuated elements into groups.

There are two aspects of Wertheimer's method that limit its usefulness for more sophisticated scientific purposes,

however. One limitation is that it is qualitative rather than quantitative. That is, it generates no numbers that can be used to measure the strength of a single grouping factor or of the combined effect of multiple factors. This limitation is important because Wertheimer's principles are essentially *ceteris paribus* rules—rules that hold with certainty only when all else is equal in the sense that no other grouping factor differentially influences the outcome—and *ceteris paribus* rules provide no way to predict the combined influence of multiple conflicting factors. Quantitative theories of multifactor integration are needed to overcome this problem, and such theories can be tested only with methods that allow grouping strengths to be quantitatively measured in both single-factor and multifactor designs.

A different limitation is that Wertheimer's demonstration method is phenomenally based, or subjective, rather than physically based, or objective. The distinction we intend to make between objective and subjective methods is simply that the former requires that there be observer-independent (i.e., physically defined) correct responses to the stated task, whereas the latter requires only that there be observer-dependent (i.e., phenomenally defined) correct responses.¹ At least for certain purposes, such as determining the accuracy of observers' performance and any response biases that may exist, it is preferable to have a method that measures performance against a physically well-defined (objective) criterion rather than a purely phenomenal (subjective) one. Objective methods, for example, would enable effective study of the extent to which

grouping processes are subject to strategic influences of various sorts (see, e.g., Beck & Palmer, 2002).

The first problem (i.e., the qualitative nature of standard grouping demonstrations) can be overcome by performing an experiment in which one actually measures the proportion of trials on which different observers report perceiving each possible grouping outcome in an ambiguous display. This approach has recently been employed with elegance and success by Kubovy and his associates (e.g., Kubovy, Holcombe, & Wagemans, 1998; Kubovy & Wagemans, 1995), by Kurylo (1997), and by our own laboratory (e.g., Palmer, Neff, & Beck, 1996; Palmer & Nelson, 2000; Rock, Nijhawan, Palmer, & Tudor, 1992). Such quantitative experimental methods do not alter the subjective nature of the task, however. Observers in such experiments are asked to report the grouping that they perceive on each trial, for which there is no objective, physically defined correct response. The probabilities of the various reports are simply computed within a well-defined framework of phenomenal report.

One can introduce objectively correct responses into a grouping paradigm by requiring participants to report a particular physically defined kind of grouping and then measuring how long the responses take and/or how accurate they are. For example, using complex displays of illusory figures induced by smaller elements, Palmer and Nelson (2000) required observers to discriminate grouping based on the orientation of the illusory figures on some blocks of trials and on the inducing elements on other blocks of trials. Thus they effectively forced participants to perceive the displays as grouped in a particular way before responding and then measured how long it took participants to make the required response. Another example is the detection paradigm developed by Field, Hayes, and Heiss (1993), which required participants to detect the presence or absence of strings of elements (as objectively defined by an algorithm devised by the experimenters) in a noisy display. The researchers then measured the accuracy with which such detections could be made

under different conditions. The problem in both cases is that such tasks are necessarily at least somewhat removed from the basic phenomenon that the paradigm is aimed at understanding (i.e., spontaneous grouping in the absence of instructions about what should be perceived).

In the hope of fostering multiple converging methods for the study of grouping effects, the present article reports a new objective method, which we call the *repetition discrimination task* (RDT). This method provides a physically defined correct response on a task that appears, superficially at least, to be unrelated to grouping and that can be studied under speeded performance conditions to generate quantitative measures of grouping strength. Participants are presented with displays like the ones shown in Figure 1, each consisting of a single row of squares and circles that alternate except for a pair in which the same shape is repeated. (These displays are actually quite similar to those originally employed by Wertheimer in demonstrating his principles of grouping subjectively.) The participants' task on each trial is to determine whether the repeated shapes are squares or circles. This is an objective task because a physically defined fact allows observers' responses to be classified as correct or incorrect. Responses are made by pressing a button—one for squares and another for circles—as quickly as possible.

Response time (RT) and accuracy in the RDT are typically measured in three different conditions. In within-group trials, a grouping factor (such as common region, shown in Figure 1A) biases the target pair to be perceived as part of the same group. In between-group trials, the same factor biases the target pair to be perceived as part of two different groups (Figure 1B). In neutral trials, the factor is equalized and thus should not bias the pair one way or the other (Figure 1C). The primary expectation is that the target pair will be detected more quickly when its elements are in the same group than when they are in different groups. The difference in RTs between the within-group and the between-group trials can then be taken as a measure of the grouping effect of the manipu-

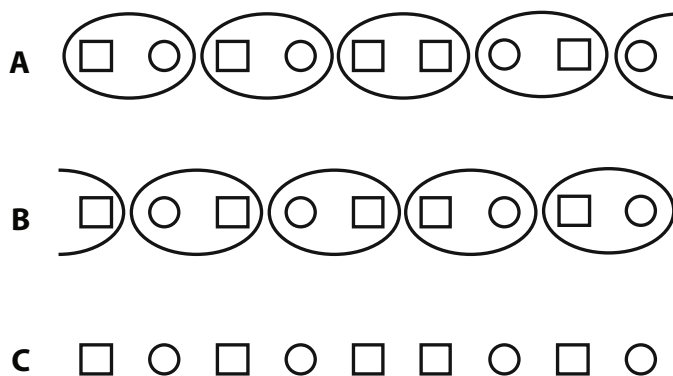


Figure 1. The logic of the repetition discrimination task (RDT). Participants should be faster at determining the shape of the repeated pair of elements when they are within the same group (A) than when they are within different groups (B), as defined by some grouping factor (here, common region). Performance on each can be compared to a neutral display (C) in which the factor is either absent or unbiased.

lated factor. Performance on neutral trials is expected to be intermediate in RT, in such a way that consistent grouping produces faster responses and inconsistent grouping produces slower responses than do the neutral grouping trials. The latter prediction can be problematic in displays like those shown in Figure 1, because other differences may also influence RTs, such as the presence of the ovals in the within-group condition versus their absence in the neutral condition.

Notice that the response of the participant in the RDT is not superficially related to grouping at all, because the RDT directly measures how long it takes the participant to discriminate the shape of a repeated pair of elements. The RDT is therefore an indirect (or implicit) measure of grouping rather than a direct (or explicit) one, a distinction explored in some detail by Kubovy and Gepshtein (2003). The fact that the RDT is indirect is simultaneously a strength and a weakness. It is a strength because it may allow grouping effects to be studied without the participant's being aware that grouping is the object of study, thus circumventing possible biases, demand characteristics, and/or strategic effects that may be involved (see, e.g., Beck & Palmer, 2002). It is a weakness because it leaves the RDT open to the objection that it may not actually measure grouping at all. We will return to this issue in discussing Experiment 2, when we compare the results of the RDT directly with those of subjective ratings of grouping. We also hasten to add that we claim no special privilege for the RDT relative to other kinds of objective tasks one might employ. One could ask participants to determine other properties of a given target pair—e.g., their color similarity or their left–right order rather than their shape similarity—and still have a paradigm that is sensitive to grouping effects. We have found the RDT to be relatively simple to use and effective in producing robust effects that we believe are due to grouping.

The prediction that grouped pairs of repeated elements in the RDT will be discriminated faster than ungrouped pairs is related to the results of several other experiments that used quite different objective tasks. One task required participants to verify the presence of parts within whole figures (Palmer, 1977; Reed & Johnsen, 1975). For example, Palmer (1977) found that three-line probes within connected six-segment figures were detected more quickly and accurately when the three-line probes constituted a natural part of the test figure than when they were spread over two different parts. He interpreted the results as indicating that people could search for multiple elements more quickly when they were contained within the same perceptual part than when they were in different parts. An analogous difference in RTs is expected in the RDT (i.e., faster responses when the target pair is located within a single group than when it is between groups), except that the targets in the RDT are separate objects within unconnected groups rather than parts within a single connected object.

Other findings relevant to the present paradigm come from experiments on object-based attention. In one of the earliest such studies, Duncan (1984) found that participants could better detect two different properties when they

were part of the same perceptual object than when they were part of two different objects in the same general location of the display. His findings indicate that participants can process information more efficiently when it comes from a single object than when it comes from multiple objects, consistent with the predictions of object-based theories of attention. Similar results supporting object-based attention have since been reported by others using both discrimination tasks (e.g., Vecera & Farah, 1994) and detection tasks (e.g., Egly, Driver, & Rafal, 1994). If such effects generalize from objects to groups of separate objects, then within-group performance will be superior to between-group performance in the RDT. It is by no means clear, however, that the results from experiments either on object-part hierarchies (e.g., Palmer, 1977) or on object-based attention (e.g., Duncan, 1984) guarantee similar effects for the looser perceptual groupings of spatially disconnected objects that are studied in the RDT experiments reported below. As Palmer and Rock (1994b) suggested, the grouping of multiple separate objects into superordinate perceptual units in the present displays may unfold quite differently in time from the perception of uniformly connected objects in the displays employed by Duncan (1984), Egly et al. (1994), Palmer (1977), and Reed and Johnsen (1975). We therefore cannot simply assume that results analogous to those found in these other paradigms will also be found in the RDT.

In this article, we present five experiments that demonstrate the usefulness of the RDT for studying the grouping of multiple disconnected objects. In Experiments 1 and 2, we validate the method using perhaps the best-known and most robust factor of grouping: proximity. In Experiment 3, we extend the method to grouping by color similarity as well as to the more recently reported factors of common region and element connectedness. In Experiment 4, we use the RDT to test a claim made by Palmer (1992) about the relative effectiveness of large versus small regions in grouping by common region. In Experiment 5, we use the RDT to tease apart different stimulus factors (size vs. orientation) that were confounded in Experiment 4. The RDT proves to be a useful method capable of answering subtle questions about grouping that are difficult or impossible to answer using classical demonstration methods.

EXPERIMENT 1 Grouping by Proximity

In the first experiment we decided to investigate participants' performance in the RDT when proximity was the factor used to induce the relevant grouping conditions: within-group versus between-group versus neutral. Proximity was chosen because it is perhaps the most basic, robust, and well-studied of all grouping factors.

The stimulus set is illustrated in Figure 2. Each display consists of a row of nine elements (to save space, only seven are shown in Figure 2) that alternate between squares and circles except for a single pair of adjacent same shapes (always squares in Figure 2) somewhere within the middle five elements (always elements 5 and 6 in Figure 2). The design is defined by the orthogonal combination of the

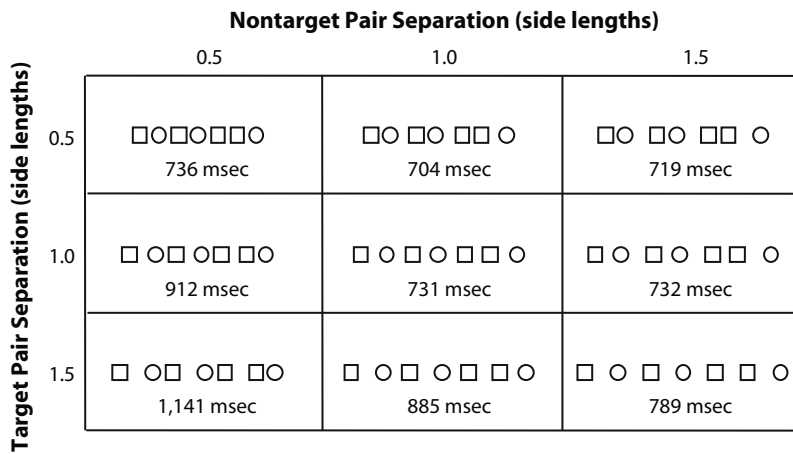


Figure 2. Stimuli for Experiment 1 on proximity grouping. The separation between elements in the target pair and between adjacent nontarget elements was varied orthogonally to produce neutral conditions (along the positive diagonal), within-group conditions (above the positive diagonal), and between-group conditions (below the positive diagonal). (Actual displays contained nine elements rather than seven.)

target pair separation and the nontarget pair separation, as measured by the distance between the target elements (target pair separation) and the distance between one target element and its adjacent nontarget element (nontarget pair separation) in terms of the length of one side of a square. The stimuli along the positive diagonal are thus examples of equally spaced (neutral) arrays with different absolute spacing. The stimuli above this diagonal are examples of within-group arrays (in which the target pair separation is smaller than the nontarget pair separation) of different grouping strengths, and the stimuli below this diagonal are examples of between-group arrays (in which the target pair separation is larger than the nontarget pair separation) of different grouping strengths. The upper right display is thus the strongest within-group stimulus and should yield the shortest discrimination times, whereas the lower left display is the strongest between-group stimulus and should yield the longest discrimination times.

Method

Participants. Eleven naive participants took part in the experiment for partial credit in an introductory psychology course.

Apparatus. Stimuli were displayed on an Apple Color Plus 14-in. color monitor via a Macintosh IIsi using PsyScope, a presentation programming language (Cohen, MacWhinney, Flatt, & Provost, 1993). Participants sat approximately 2 ft (60 cm) from the monitor and responded by pressing one of two keys on a button box with their left or right index finger. The keys were labeled with a square or a circle and their positions (left vs. right) were counterbalanced over participants.

Design. The experiment was a five-way factorial, within-participants design. The factors included: target pair shape (square vs. circle), target pair separation (0.5, 1.0, or 1.5 side lengths), nontarget pair separation (0.5, 1.0, or 1.5 side lengths), target pair position (elements 3–4, 4–5, 5–6, or 6–7, counting from the left), and the number of repetitions of the above conditions (1–8). The sides of the squares and diameter of the circles subtended approximately 0.8° of visual angle, and the entire display subtended between approximately 10° and 16° of visual angle, depending on the separations used.

Procedure. The participants were read instructions that explained the task and encouraged them to respond as quickly as possible while keeping their error rate to no more than about 5%. Participants were shown examples of equal spacing, close spacing, and far spacing of the target pair as well as examples in which the position of the target pair varied. This acquainted the participants with these factors and the fact that they were irrelevant for the task to be performed. Participants were then given 36 practice trials on the computer before the experimental trials began. The 576 experimental trials were then presented in 8 blocks between which participants were allowed to take a break if they wished.

The stimulus array was presented and left on until the participant responded. The array disappeared immediately after the response was made, and a 500-msec delay preceded the start of the next trial. If the response was incorrect, the computer beeped.

Results and Discussion

Each participant's RTs were averaged over replications as a function of target pair separation, nontarget pair separation, and target pair position, and they were submitted to an overall ANOVA. The results indicate that all main effects and interactions were highly significant: target pair separation [$F(2,20) = 50.54, p < .001$], nontarget pair separation [$F(2,20) = 36.78, p < .001$], target pair position [$F(3,30) = 59.03, p < .001$], and interactions between target pair separation and nontarget pair separation [$F(4,40) = 23.42, p < .001$], target pair separation and target position [$F(6,60) = 4.18, p < .01$], and target pair separation, nontarget pair separation, and target position [$F(12,120) = 1.95, p < .05$]. The overall error rate was 3%.

The main results of interest are shown below the stimuli in Figure 2 and are plotted in Figure 3. Several findings are noteworthy. First, notice that the within-group conditions (the three open squares in Figure 3) yielded, on average, faster RTs than the neutral conditions (the three filled circles) [$F(1,10) = 11.89, p < .01$] and that the neutral conditions yielded faster RTs than the between-group conditions (the three open diamonds) [$F(1,10) = 45.79$,

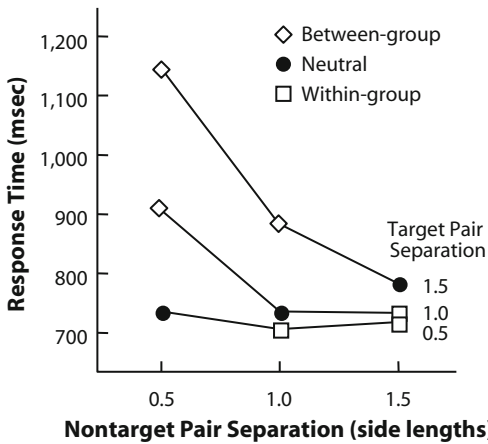


Figure 3. Results of Experiment 1 for separation conditions. Mean response times are plotted as a function of the nontarget pair separation (*x*-axis) and target pair separation (curves).

$p < .001$]. This is the predicted effect of proximity grouping on performance in the RDT, and it indicates that RTs in the RDT are a plausible measure of grouping effects. Second, notice that the responses in the neutral conditions are almost as fast as those in the within-group conditions. This probably occurred because the shape similarity of the target elements themselves produced some grouping in the neutral conditions. Therefore, shape similarity would tend to make the target pair group together and be perceived quickly even when proximity is neutral. Third, notice that there is a relatively clean division of the conditions into those with fast RTs, consisting of the within-group and neutral conditions (roughly 700–800 msec), and those with slow RTs, consisting of just the between-group conditions (roughly 900–1,200 msec). This partition of the conditions seems to be a reasonable portrayal of the subjective organization of the stimuli, but we will defer further discussion of this issue until Experiment 2, in which we examine direct ratings of grouping for these stimuli.

There is an important and unavoidable confound in the present design that requires comment: The faster RTs in the within-group conditions necessarily correspond to shorter distances between the elements of the target pair, and the slower RTs in the between-group conditions necessarily correspond to longer distances. If RT were solely a function of absolute distance between elements of the target pair, with no grouping influences at all, one would still expect there to be differences in the predicted direction: within-group fastest, neutral intermediate, and between-group slowest. Indeed, the data for the three neutral conditions contain evidence that absolute distance matters, because performance was slower for the 1.5 side-length spacing (788 msec) than for the 0.5 and 1.0 spacing (736 and 731 msec, respectively) [$F(1,10) = 29.72, p < .001$].

The overall pattern of results is not consistent with predictions based strictly on the distance between the target pairs, however. If it were, RTs should be affected only by

the target pair separation and not by the distance between adjacent nontarget elements (nontarget pair separation). The problem with this alternative hypothesis in terms of absolute distance effects is perhaps most obvious for the 1.0 condition of nontarget pair separation (i.e., the middle row of Figure 2). Because the distance between the target elements is 1.0 side lengths in each case, RTs should be the same, but they clearly were not. When the adjacent nontarget elements were closer (0.5 side lengths), RTs were about 180 msec slower than when the adjacent nontarget elements were the same distance (1.0 side lengths) or farther (1.5 side lengths).

Such interactions between target pair and nontarget pair separation show that it is the distance between the elements of the target pairs relative to the distance between adjacent nontarget elements that governs performance. Of course, one could maintain that it is not grouping per se that matters, but simply relative distances. There is no evidence against this interpretation in the present experiment, but later we will present evidence of similar effects due to grouping by color, common region, and element connectedness in which relative distances are constant and equal (see Experiments 3–5). At this point, however, we acknowledge the ambiguity in interpretation and maintain that the results are consistent with either a grouping account or a relative distance account.

Figure 4 shows the effect of target position for within-group, neutral, and between-group conditions. There is a very pronounced U-shape to these functions in all three grouping conditions, as indicated by the large effect of target position [$F(3,30) = 59.28, p < .001$]. There was also a smaller interaction between grouping condition and target position [$F(6,60) = 4.55, p < .001$], because the effect of target position was greatest for the between-group condition and smallest for the within-group condition. However, for all three conditions, when the target pair was in either of the central two positions, it was perceived significantly more quickly than when it was in either of the outer two positions [$F_s(1,10) = 157.61, 70.32, 32.35, p <$

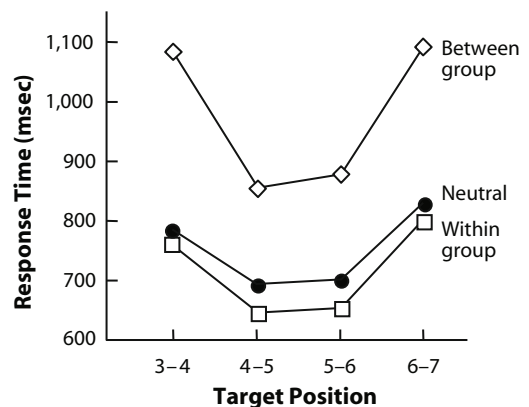


Figure 4. Results of Experiment 1 for target positions and grouping conditions. Mean response times are plotted as a function of the position of the target pair (*x*-axis) and the grouping condition (curves).

.001]. These differences presumably reflect either higher spatial resolution in central than in peripheral vision or a scanning strategy that begins in the center of the displays. Eye movements were not monitored, so we have no basis for deciding between these possibilities.

Another alternative to the grouping explanation that has been suggested is that participants were performing the task by detecting the symmetry axis, which always fell between the two identical elements, and then classifying the shapes of these two elements. Symmetry alone, however, would not produce the differences obtained between the within-group and between-group displays, because both displays were symmetrical about the same axis. Symmetry is detected more rapidly and easily when elements are closer to the axis of reflection (Julesz, 1971), as they were in the within-group displays, and when the axis is close to the center of the display, where acuity is greatest. Again, the effect of distance would have to be based on relative rather than absolute displacement between elements to account for the obtained effects, as noted above, but an account based on distance effects in symmetry detection cannot be ruled out from the present data. We will return to this possibility in Experiment 3, in which several other grouping factors are studied, none of which involve different distances between elements.

EXPERIMENT 2

Subjective Ratings of Proximity Effects

Experiment 1 showed that grouping affects performance in the RDT as predicted and that the RDT may therefore provide a reasonable, objective, quantitative measure of grouping. To compare these results with a corresponding subjective grouping task, we used the same stimulus arrays in Experiment 2 and asked a different set of participants to rate the strength or degree of the grouping between the elements of the target pair in each display.

Method

Participants. Ten naive participants took part in the experiment for partial credit in an introductory psychology course.

Apparatus. The apparatus was the same as that used in Experiment 1.

Design. The experimental design was the same as that employed in Experiment 1.

Procedure. The participants were read instructions that asked them to rate each stimulus display in terms of "how easy or difficult it is to see that the row contains a pair of adjacent circles or squares" on a scale from 1 (*very weak*) to 9 (*very strong*). (For a full description of the rating instructions, see the Appendix.) Participants recorded their ratings using the numerical keys on the computer keyboard. Displays remained visible until the participants made their responses.

Results and Discussion

The main results of interest are plotted in Figure 5. Significant differences in mean ratings are evident for the effects of target pair separation [$F(2,18) = 74.98, p < .001$], nontarget pair separation [$F(2,18) = 32.75, p < .001$], and their interaction [$F(2,18) = 3.72, p < .05$]. There is a strong negative correlation with the corresponding data

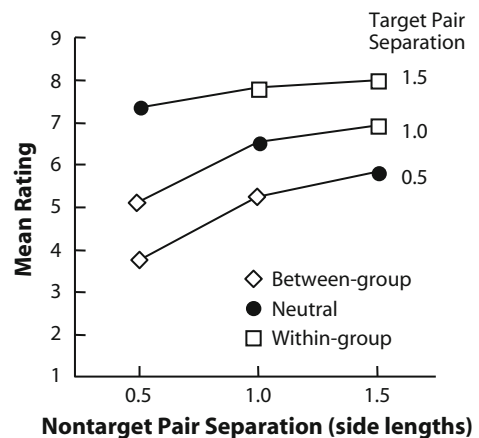


Figure 5. Results of Experiment 2 for separation conditions. Mean ratings of grouping strength are plotted as a function of the nontarget pair separation (*x*-axis) and target pair separation (curves).

from Experiment 1 ($r = -.92, p < .001$], indicating that high ratings in this experiment corresponded closely with short RTs in Experiment 1. All of this is as it should be.

Even so, there are some noteworthy differences between the rating data (Figure 5) and the corresponding RT data in Experiment 1 (Figure 3). First, the interaction between target pair separation and nontarget pair separation is much less pronounced here than in Experiment 1. Indeed, the rating data look very nearly additive with respect to these two factors, and the interaction between target and nontarget pair separation is only marginally significant. Second, the rating data are less categorical with respect to grouping than the RDT RTs. By this we mean that the rating data provide less evidence of a clear division into those arrays in which the target pair elements were grouped together versus those in which they were not. There is a cut at a rating of about 5.4 that separates the within-group and neutral conditions from the between-group conditions, but the separation is quite narrow and there is a lot of variation within the conditions that are separated by this cut. It is not clear how to interpret these differences, however. They may be substantive, but they might also arise simply from a nonlinear relation between ratings and RTs. Because we have no reason to believe that this relation should be linear, we will tentatively assume that nonlinearities are the cause.

Figure 6 shows the effects of target position on the rating data. Although the U-shape of these functions is not as pronounced as the corresponding positional effects in the RT data (Figure 4), these functions are nevertheless reliable, both as a main effect [$F(3,27) = 5.87, p < .01$] and in their interaction with grouping condition [within-group, neutral, and between-group conditions, $F(6,54) = 2.55, p < .05$]. The difference in the size of these effects may be attributable to a nonlinear relation between ratings and RTs; it may also plausibly reflect the fact that participants were able to look at the displays in Experiment 2 for an extended period of time and make eye movements

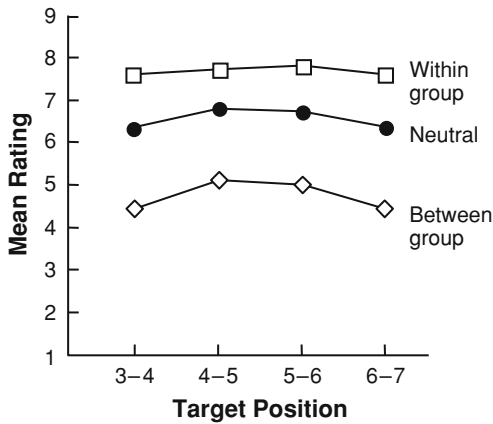


Figure 6. Results of Experiment 2 for target positions and grouping conditions. Mean ratings of grouping strength are plotted as a function of the position of the target pair (x-axis) and the grouping condition (curves).

to the target pairs, thus attenuating positional effects in Experiment 1 caused by differential visibility at different retinal positions.

The agreement between the results of Experiments 1 and 2 is perhaps not as strong as one would have expected, but it is nevertheless substantial enough to support the conclusion that the RDT is measuring something closely akin to grouping strength. Indeed, there are certain respects in which the data from the RDT appear to conform more closely to the categorical grouping one expects a priori than to the more graded structure of the explicit rating data.

We should emphasize that the subjective rating data are not necessarily the more accurate reflection of participants' perceived grouping. In rating the displays, participants may have adopted an evaluation strategy that was more cognitive than perceptual. The fact that the ratings followed the experimental factors very closely may be due to participants' realizing what factors were being manipulated and making their ratings accordingly. For example, even if they perceived the within-group and neutral conditions as more similar, participants may have used the precise distance relationships in the displays to spread their ratings more evenly over the 9-point scale.

EXPERIMENT 3 Color Similarity, Common Region, and Element Connectedness

The third experiment uses the RDT to examine grouping based on three other factors: color similarity, common region, and element connectedness. Figure 7 shows a small sample of the stimuli, including a single neutral condition (Figure 7A) and one within-group and one between-group condition for each factor we studied (Figures 7B, 7C, and 7D). One reason for studying these factors was to demonstrate that the results of Experiment 1 are not restricted to grouping by proximity. This is particularly important because of the confound between relative distance and

grouping that was mentioned in the discussion of Experiment 1, both as an explanatory factor by itself and in conjunction with the symmetry-based alternative explanation to grouping. All three of the factors in Experiment 3 produce grouping with equal spacing between the elements. Indeed, common region actually induces a slight distance illusion, as readers may be able to verify by inspecting Figure 7C, in which the grouped elements within the same oval appear to be slightly farther apart than the adjacent, nongrouped elements in different ovals. Common region therefore provides a particularly strong test of the hypothesis that the results of Experiment 1 can best be explained by general grouping effects or by relative distance effects, either alone or in conjunction with an account in terms of symmetry.

Color similarity (Figure 7B) was studied because it is another well-established classical grouping factor that can produce very robust grouping effects. Common region (Figure 7C) and element connectedness (Figure 7D) are more recently described grouping variables (Palmer, 1992; Palmer & Rock, 1994b) that have not yet received much study. Another reason for choosing these particular factors is that color similarity, like proximity, is an example of what Palmer (1992) called *intrinsic* grouping variables, whereas common region and element connectedness are

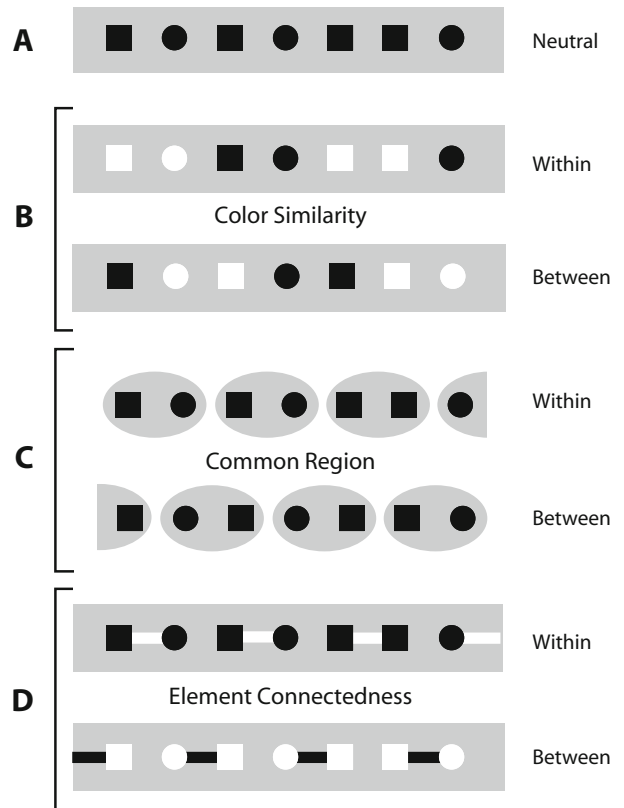


Figure 7. Example stimuli for Experiment 3. Relative to neutral control conditions (A), within-group and between-group conditions were created for the factors of color similarity (B), element connectedness (C), and common region (D). (Actual displays contained nine elements rather than seven.)

examples of *extrinsic* or *induced* grouping variables. The distinction is that intrinsic grouping factors are based on the properties of the grouped elements themselves (e.g., their positions and their colors), whereas extrinsic grouping factors are based on relations between the target elements and other contextual elements that induce them. Grouping by common region, for example, is induced by the presence of the ovals that surround certain pairs of elements in the displays, and grouping by element connectedness is induced by the bars that connect certain pairs of elements. The results of the present experiment thus will allow us to examine whether systematic differences emerge between intrinsic and extrinsic grouping factors within the RDT paradigm.

Method

Participants. Ten naive participants took part in the experiment for partial credit in an introductory psychology course.

Apparatus. The apparatus was the same as that used in Experiment 1.

Design. The experimental design was the same as that employed in Experiment 1 except that color similarity, common region, and element connectedness were the grouping factors employed.

Procedure. The procedure was the same as that followed in Experiment 1 except for the difference in the stimulus materials.

Results and Discussion

The primary data of interest are shown in the histograms in Figure 8. In all three cases, the within-group conditions yielded significantly faster RTs than the between-group conditions [$F(1,9) = 52.9, p < .001$ for color similarity; $F(1,9) = 18.0, p < .01$ for common region; and $F(1,9) = 6.65, p < .05$ for element connectedness]. These differences demonstrate that these factors, in addition to proximity, produce RTD effects consistent with predictions based on subjective grouping. None of the factors in the present experiment produced differential spacing, further demonstrating that RDT effects can be obtained in cases in which they cannot be explained by relative distance. The

fact that common region and element connectedness also produced differences in the predicted direction for within-group and between-group conditions also argues against an explanation in terms of symmetry. The within-group and between-group displays in these two conditions are equally symmetric, with no differences in distance from the central axis, so it is not clear how symmetry could explain the effects obtained by these conditions.

The pattern of results for the neutral conditions versus the within-group and between-group conditions differs among the three factors. Color similarity produced the same basic result as proximity in Experiment 1: within-group conditions yielded RTs that were slightly but not significantly faster than those obtained from the neutral conditions [$F(1,9) = 0.64, p = .55$], whereas between-group conditions yielded RTs that were much slower than those obtained from the neutral conditions [$F(1,9) = 77.37, p < .001$]. Common region produced a markedly different pattern: neutral conditions produced RTs that were faster than those obtained from both the within-group [$F(1,9) = 20.78, p < .01$] and the between-group [$F(1,9) = 45.79, p < .001$] conditions. The same was true for element connectedness [$F(1,9) = 50.64, 76.71, ps < .001$]. The faster responses to neutral conditions in the latter two cases may be due to the fact that the inducing elements—i.e., the ovals in common region and the bars in element connectedness—were absent in the neutral conditions, making the neutral conditions simpler and less cluttered than the within-group and between-group conditions. Thus in neutral conditions, ignoring the inducing elements may not have required any attentional selection, and these conditions may therefore have been processed more quickly.

These differences in relation to the neutral trials are consistent with Palmer's (1992) distinction between intrinsic and extrinsic grouping. The two intrinsic principles we have studied (proximity and color similarity) showed the predicted pattern: Performance on the neutral trials was intermediate between the within-group and the between-group trials. The two extrinsic principles (common region and element connectedness) showed a different pattern: Performance on neutral trials was faster than that on either within-group or between-group trials. With only two examples of each type in hand, drawing any firm conclusion would be premature, but the differences suggest significantly different processing.

EXPERIMENT 4

Size Effects in Grouping by Common Region

The fourth experiment employs the RDT to settle an empirical question that has been raised but is difficult to answer using phenomenological demonstrations: What happens when two different organizations compete in the same display? In the original article on grouping by common region, Palmer (1992) included figures in which two versions of common region conflicted with each other (see Figures 9C and 9D). Based on his subjective experience, Palmer suggested that the small ovals provided

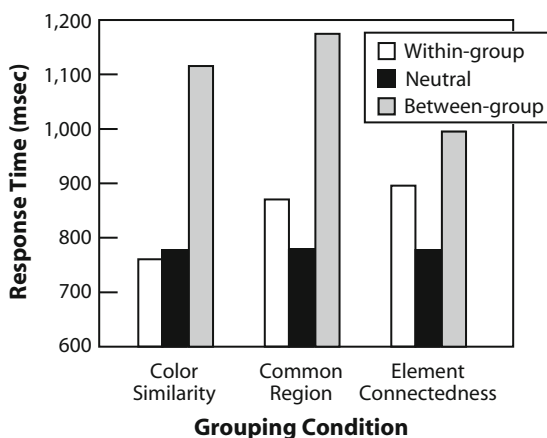


Figure 8. Results of Experiment 3. Mean response times are plotted for neutral, within-group, and between-group conditions based on color similarity, common region, and element connectedness.

stronger grouping, but admitted that such phenomenological evidence was insufficient to decide the issue. The RDT provides a more precise method for addressing this question. In the present experiment, the small and large ovals were arranged so that they competed with each other in influencing the perceived organization of the squares and circles. The empirical issue of whether grouping effects are stronger within the large ovals or the small ones amounts to asking whether observers are faster at detecting the adjacent squares (or circles) when they are within the large ovals or the small ones. To determine whether any advantage for small or large ovals extended to conditions without competition, the same (within-group) displays were presented with just the small ovals (Figure 9A) or just the large ovals (Figure 9B) as control conditions.

Method

Participants. Thirteen naive participants took part in the experiment for partial credit in an introductory psychology course.

Apparatus. The apparatus was the same as that used in Experiment 1.

Design. Three types of common region displays were used: small ovals only (Figure 9A), large ovals only (Figure 9B), and competing small and large ovals (Figure 9C). These three common region types were combined with the following factors: (1) common region of the target pair [for competing oval displays, small oval (e.g., Figure 9C) vs. large oval (e.g., Figure 9D)]; for small and large oval displays, within-oval (e.g., Figure 9A) vs. between-ovals (e.g., Figure 9B)]; (2) shape of the target pair (squares or circles); (3) position of the

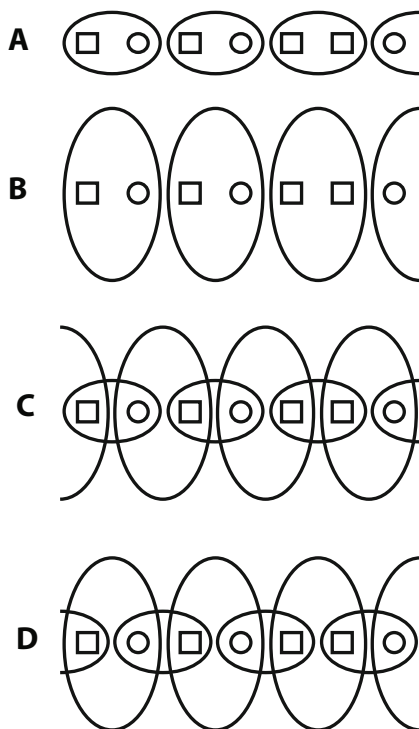


Figure 9. Example stimuli for Experiment 4. Repeated pairs were presented within small ovals alone (A), within large ovals alone (B), within small ovals and between large ovals (C), or within large ovals and between small ovals (D). (Actual displays contained nine elements rather than seven.)

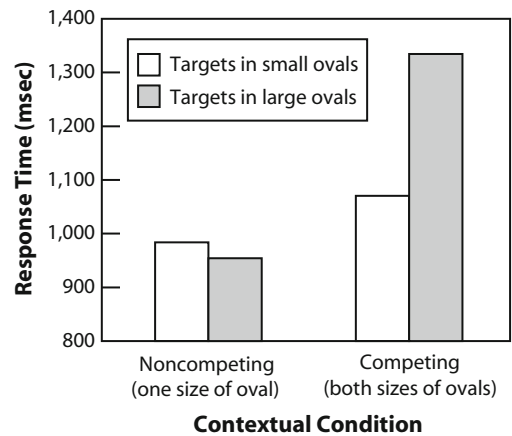


Figure 10. Results of Experiment 4. Mean response times are plotted for discriminating the shape of the target elements within small versus large ovals when only one size of oval is presented and when both sizes are presented together in conflicting displays.

target pair (elements 3–4, 4–5, 5–6, or 6–7, counting from the left); and (4) repetitions (1–4).

Procedure. The procedure was basically the same as that followed in Experiment 1 except for the difference in the stimulus materials. All three common region types were randomly intermixed within a block. Participants were given instructions analogous to those in Experiment 1 and received practice on 48 trials.

Results and Discussion

The main result of interest is whether, for the competing displays, the participants were faster in determining the shape of the target pair when both target elements were within the large oval or when both were within the small oval. The data (see Figure 10) unequivocally show faster responses to target pairs within the small ovals (1,068 msec) than to those within the large ovals (1,326 msec) [$F(1,12) = 33.51, p < .001$]. If we assume that these measures of performance in the RDT reflect the relative strength of grouping, as argued from the results of the first three experiments, then the RDT gives a clear and decisive answer to the question of whether large or small ovals are more effective in grouping by common region: The small ovals dominate perceived grouping.

Interestingly, the opposite effect is present for the noncompeting, single-oval displays. When the target pair was located within the same large oval, participants were faster (928 msec) at determining the shapes than when the target pair was located within the smaller ovals (975 msec) [$F(1,12) = 9.61, p < .01$]. This small advantage may be related to the advantage for the neutral condition over the common region conditions in Experiment 3; that is, the large ovals produced less crowded displays than the small ovals and may therefore have required less attentional selection. The increase in RTs to the small oval displays thus may reflect an overall cost of attentional filtering.

Regardless of the reason for the large oval advantage in noncompeting displays, it is clear that placing the ovals in competition is essential to obtaining the small oval advantage. Alone, both oval sizes effectively group the stimulus,

but when the ovals are placed in direct conflict with each other, smaller ovals produce much stronger grouping.

EXPERIMENT 5 Size Versus Orientation Effects in Grouping by Common Region

Unfortunately, there is a confound in the stimuli used in Experiment 4 that clouds the conclusion that size matters in grouping by common region: The smaller ovals were oriented horizontally, along the configural line of elements, whereas the larger ovals were oriented vertically, perpendicular to the configural line of elements. It is therefore possible that the crucial factor in the results of Experiment 4 was the orientation of the ovals rather than their size.

Experiment 5 uses the RDT to resolve this issue. We designed a set of stimuli that orthogonally varied the size and orientation of the inducing ovals, as shown in Figure 11. The contours of the enclosing regions were either ovals or circles and the relative sizes of the regions were either the same as or different from each other.

Method

The methods were the same as those employed in Experiment 4 except for the differences in stimulus displays. Thirteen naive participants took part in the experiment for partial credit in an introductory psychology course.

Results and Discussion

The primary data of interest are shown under the stimuli in Figure 11. The upper number beneath each display indicates the mean RT required to discriminate the shape of the target pair when it was located between groups defined by the small and/or horizontal ovals, and the lower number indicates the RT when the pair was located within

groups defined by the small horizontal ovals. The RT difference between these two measures, indicating the extent to which the smaller and/or horizontal ovals dominated the larger and/or vertical ones, is indicated to the right of the raw data. The pattern of results indicates that both size and orientation appeared to play a role. When the difference scores were subjected to an ANOVA, collapsed over all other factors, there were main effects of both size [$F(1,12) = 34.53, p < .001$] and orientation [$F(1,12) = 18.58, p < .01$], as well as an interaction between them [$F(1,12) = 9.71, p < .01$] such that the combination of small and horizontal ovals produced an even greater RT difference than either small or horizontal ovals presented alone. Clearly, both size and orientation matter.

Summary and Conclusion

We have presented the results of five experiments that employ the RDT to study grouping phenomena in an objective, quantifiable way. Experiment 1 showed that proximity grouping produced the predicted differences in RTs for within-group versus between-group conditions, although the results were open to an alternative interpretation in terms of relative distance effects. Experiment 2 showed that the data from Experiment 1 were at least qualitatively similar to subjective ratings of grouping strength with the same displays. Experiment 3 further demonstrated RDT effects with three other grouping factors—color similarity, common region, and element connectedness—that did not have the relative distance confound present in Experiment 1. Further, the pattern of results for the neutral versus within-group/between-group conditions in Experiments 1 and 3 suggests a measurable difference between intrinsic grouping factors (e.g., proximity and color similarity) and extrinsic grouping factors that is unlikely to be noticed in standard phenomenologi-

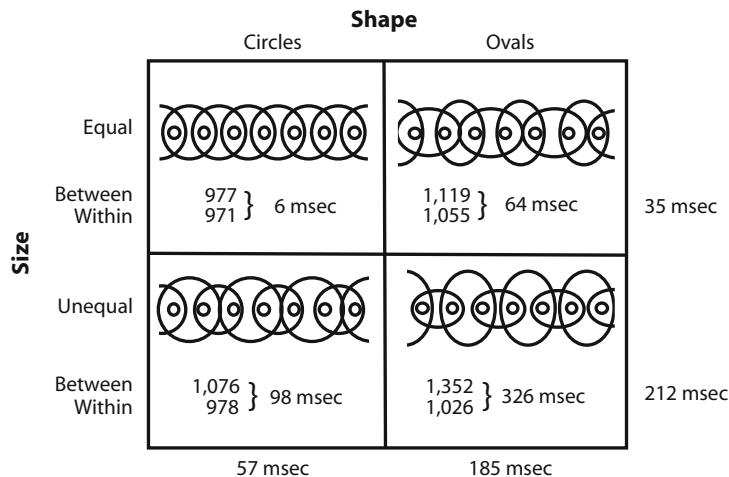


Figure 11. Stimuli and results for Experiment 5. The effects of the shape and size of surrounding boundaries were assessed by orthogonally combining circles versus ovals with equal areas versus unequal areas. Response times for between-group (top) and within-group (bottom) conditions are given below the stimuli, together with the response time difference that measures the size of the grouping effect in each case.

cal demonstrations. Experiments 4 and 5 showed how the method can be used to get clear answers to two empirical questions about grouping by common region that are not easily decided by phenomenology. The analysis of RDT RTs provides clear evidence that both the size and the orientation of an enclosing region are important variables in grouping by common region and that their combined effect is superadditive.

We believe that the RDT will prove to be an important tool in the experimental analysis of perceptual grouping. The RDT is likely to be useful in solving grouping problems that involve discovering the integration rule when multiple factors are simultaneously present. Experiments 4 and 5 addressed a very simple example of this sort. Such problems require quantitative data so that mathematical models of integration may be tested in rigorous ways. The RDT has already been used to address the class of problems that involve determining whether grouping is a truly obligatory, bottom-up process, or whether it is "cognitively penetrable" in the sense of being subject to strategic effects under voluntary control (see Beck & Palmer, 2002). Perhaps the RDT method, in conjunction with other quantitative objective methods, will enable the understanding of grouping to move beyond the purely qualitative limits of standard phenomenological demonstrations.

AUTHOR NOTE

This research was supported in part by Grant 1-R01-MH46141 from the National Institute of Mental Health to S.E.P. The authors thank Michael Kubovy for his comments on two earlier drafts of this manuscript and several anonymous reviewers for their useful critiques. Correspondence concerning this article should be addressed to S. E. Palmer, Psychology Department, University of California, Berkeley, CA 94720-1650 (e-mail: palmer@cogsci.berkeley.edu).

REFERENCES

- BECK, D. M., & PALMER, S. E. (2002). Top-down influences on perceptual grouping. *Journal of Experimental Psychology: Human Perception & Performance*, *28*, 1071-1084.
- BLAKE, R., & YANG, Y. (1997). Spatial and temporal coherence in perceptual binding. *Proceedings of the National Academy of Sciences*, *94*, 7115-7119.
- COHEN, J., MACWHINNEY, B., FLATT, M., & PROVOST, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, *25*, 257-271.
- DUNCAN, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*, 501-517.
- EGLY, R., DRIVER, J., & RAFAL, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*, 161-177.
- FIELD, D. J., HAYES, A., & HEISS, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field." *Vision Research*, *33*, 173-193.
- JULESZ, B. (1971). *Foundations of cyclopean perception*. Chicago: University of Chicago Press.
- KUBOVY, M., & GEPSHTEIN, S. (2003). Perceptual grouping in space and space-time: An exercise in phenomenological psychophysics. In R. Kimchi, M. Behrmann, & C. R. Olson (Eds.), *Perceptual organization in vision: Behavioral and neural perspectives* (pp. 45-86). Mahwah, NJ: Erlbaum.
- KUBOVY, M., HOLCOMBE, A. O., & WAGEMANS, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology*, *35*, 71-98.
- KUBOVY, M., & WAGEMANS, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative Gestalt theory. *Psychological Science*, *6*, 225-234.
- KURYLO, D. D. (1997). Time course of perceptual grouping. *Perception & Psychophysics*, *59*, 142-147.
- PALMER, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, *9*, 441-474.
- PALMER, S. E. (1992). Common region: A new principle of perceptual grouping. *Cognitive Psychology*, *24*, 436-447.
- PALMER, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.
- PALMER, S. E., NEFF, J., & BECK, D. (1996). Late influences on perceptual grouping: Amodal completion. *Psychonomic Bulletin & Review*, *3*, 75-80.
- PALMER, S. E., & NELSON, R. (2000). Late influences on perceptual grouping: Illusory figures. *Perception & Psychophysics*, *62*, 1321-1331.
- PALMER, S. [E.], & ROCK, I. (1994a). On the nature and order of organizational processing: A reply to Peterson. *Psychonomic Bulletin & Review*, *1*, 515-519.
- PALMER, S. [E.], & ROCK, I. (1994b). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review*, *1*, 29-55.
- REED, S. K., & JOHNSEN, J. A. (1975). Detection of parts in patterns and images. *Memory & Cognition*, *3*, 569-575.
- ROCK, I., NIJHAWAN, R., PALMER, S. [E.], & TUDOR, L. (1992). Grouping based on phenomenal similarity of achromatic color. *Perception*, *21*, 779-789.
- VECERA, S. P., & FARAH, M. J. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: General*, *123*, 146-160.
- WERTHEIMER, M. (1923). Untersuchungen zur Lehre von der Gestalt: II. *Psychologische Forschung*, *4*, 301-350. [Partial translation in W. D. Ellis (Ed.) (1950). *A sourcebook of Gestalt psychology* (pp. 71-81). New York: Humanities Press.]

NOTE

1. We note that the difference is not, as is often supposed, that objective tasks have correct answers and subjective ones do not. Both have correct answers. The difference is that the correct answer in an objective task can be defined independently of the observer, based on some physical fact (e.g., which of two displays is larger or brighter), whereas the correct answer in a subjective task can be defined only for a particular observer at a particular moment, based on a phenomenal or experiential fact (e.g., which of two displays appears more beautiful to the observer). We note further that in both cases, observers must make their responses based on phenomenal impressions (e.g., which display *appears* larger or brighter or more beautiful). The only difference between objective and subjective tasks lies in whether there is some other, physically well-defined fact of the matter that allows observers' responses to be classified by the experimenter as correct or incorrect. This classification is possible for 2AFC judgments of size or brightness but not for 2AFC judgments of beauty.

APPENDIX

Rating Instructions for Experiment 2

In this experiment we will want you to rate the degree to which it is easy or difficult to see that the row contains a pair of circles or a pair of squares. You will rate the degree of ease on a 9-point scale. If the pair stands out very well it should get the maximal rating of 9. If, on the other hand, the same shape pair is difficult to see, it should get the minimal rating of 1. All intermediate ratings will get numbers between 1 and 9. Thus, a rating of 5 would be given to a row which seems to you to be about halfway between a very easy row and a very hard row. You will see a wide range of examples, so please try to use all of the numbers between 1 and 9.

(Manuscript received December 8, 2004;
revision accepted for publication March 8, 2006.)