Journal of Experimental Psychology:
Human Perception and Performance
1993, Vol. 19, No. 3, 564–579

Copyright 1993 by the American Psychological Association, Inc.
0096-1523/93/$3.00

Feature Distribution and Biased Estimation of Visual Displays

Robert L. Goldstone

Perceptual equivalents of confirmation biases and framing effects are observed in subjects' estimates of feature numerosity. Subjects are asked to estimate the percentage of display items that have a particular feature. Features either are randomly distributed or are spatially clustered such that features of the same type tend to be close. Subjects systematically overestimate the prevalence of features in clustered displays. The pattern of results is best explained by a regional salience bias: Features tend to be more salient if they belong to regions that have a high concentration of instruction-mentioned features. The regional salience bias is contrasted with a feature salience bias: Features tend to be more salient if they are mentioned in the instructions. The relations among the observed perceptual bias and traditional confirmation biases, numeric estimation, and attention are discussed.

Much of the recent work in reasoning and judgment concerns the use of heuristics to make decisions. Heuristics that usually yield reasonably accurate judgments also result in systematic biases under some circumstances (for reviews, see Fischhoff & Beyth-Marom, 1983; Hogarth, 1981; Kahneman, Slovic, & Tversky, 1982). Although much of the work suggesting biased judgment has been conducted on fairly "high-level" cognitive tasks, the burden of the experiments reported here is to show that versions of these biases also exist in more perceptual domains. When subjects are asked to judge the number or percentage of figures in a display, effects are apparent that are analogous to findings that have been explained by "confirmation biases" and "framing effects."

Confirmation Biases and Alternative Accounts

A "confirmation bias" is a tendency for people to seek or use confirmatory rather than disconfirmatory evidence (Devine, Hirt, & Gehrke, 1990; Mynatt, Doherty, & Tweney, 1977; Wason, 1960, 1966; Wason & Johnson-Laird, 1972). In an early study suggesting confirmation bias, Wason (1960) asked subjects to discover the rule used to generate a set of number triplets. Subjects were initially told that the triplet "2, 4, 6" belonged to the set. Subsequently, subjects generated triplets to test their hypotheses and were told whether the triplet obeyed the rule. Subjects would typically hypothesize a rule such as "numbers ascending by twos." They would often test the hypothesis by asking whether "6, 8, 10" belonged to the set, expecting the answer to be "yes." Subjects rarely asked whether series such as "6, 7, 8" belonged to the set. In general, subjects tended to generate triplets that were consistent with their hypothesis and rarely attempted to falsify their hypothesis.

It is not clear whether there is a tendency for people to ignore falsifying evidence when it is presented or simply a tendency not to search for falsifying evidence. Mynatt et al.'s (1977) subjects, when confronted with unambiguous falsifying evidence, used the evidence to reject their incorrect hypothesis. Beattie and Baron (1988) found that some subjects continued to retain hypotheses that were falsified by immediately available evidence. In summary, much evidence exists that subjects are biased to seek out evidence that confirms their hypothesis and also perhaps to dismiss or distort evidence that would oppose their hypothesis (see also Lord, Ross, & Lepper, 1979; Shaklee & Mims, 1982).

Framing Effects

Framing effects refer to the influence of how a question is framed on subject's judgments (Levin, Schnittjer, & Thee, 1988; Tversky & Kahneman, 1981). Normatively identical questions receive different responses, depending on the wording of the question. In one experiment (Tversky & Kahneman, 1981), subjects are asked to imagine a disease that would kill 600 people if nothing were done. One group is asked to choose between Programs A and B, and another group is asked to choose between Programs C and D:

> If Program A is adopted, 200 people will be saved. If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved. (Seventy-two percent choose A; 28% choose B.) If Program C is adopted, 400 people will die. If Program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die. (Twenty-two percent choose C; 78% choose D.)

The choice between Programs A and B is equivalent to the choice between Programs C and D. The only difference concerns whether the problem is framed in terms of lives saved or lost. Inconsistent responses to the two problems arise from

Robert L. Goldstone, Department of Psychology, Indiana University.

I thank Asher Cohen, Igor Gavanski, Jamin Halberstadt, Edward Hirt, Douglas Medin, and Steven Sherman for useful comments and suggestions. Useful reviews were provided by Lana Trick, Ryan Tweney, Jesse Whitlow, and an anonymous reviewer.

Correspondence concerning this article should be addressed to Robert L. Goldstone, Psychology Building, Indiana University, Bloomington, Indiana 47405.

the conjunction of a framing effect with a tendency for responses to losses to be more extreme than responses to gains. Subjects are biased to interpret a problem in terms of the vocabulary directly mentioned in the instructions or description for the problem.

Although not explicitly linked in the literature, there is a clear relation between framing effects and results that have been taken to support confirmation biases. One alternative to a "confirmation bias" account of experimental findings is that there is a tendency for subjects to select stimuli with properties that match the properties that are explicitly mentioned in the rule (Beattie & Baron, 1988; Wason & Evans, 1975). This bias is called a "matching bias." Snyder and Swann (1978) and Devine *et al.* (1990) found that, when asked to generate questions to determine a person's extraversion, subjects are more likely to generate questions that probe for extraverted behavior ("What would you do if you wanted to liven things up at a party?"). When asked to determine a person's introversion, subjects are more likely to ask questions that probe for introversion ("In what situations do you wish you could be more outgoing?"; see also Slowiaczek, Klayman, Sherman, & Skov, 1992). Likewise, framing effects occur when subjects selectively weight properties that are alluded to in the instructions or cover story. For both framing effects and confirmation biases, the particular properties that are mentioned by the experimenter influence behavior.

Biased Perception

As the previous research indicates, tests of confirmation biases and framing effects have predominantly used tasks that involve logical reasoning, high-level inference, or explicit hypothesis formation. Although the biases are found in several domains (person perception, evaluation of gambles, financial decisions, and verification of logical rules), it is not clear whether they are also found in more perceptual domains. The objective of the reported experiments is to extend the work on cognitive biases to more perceptual judgments and to analyze the biases' cause. The tasks to be reported do involve cognitive processes of estimation and judgment; however, the tasks are based on perceptually available properties of scenes and do not involve "high-level" processes such as explicit hypothesis formation. There is little reason for expecting to find a clear distinction between perceptual and conceptual processes. Still, along a graded continuum, previous tests of confirmation biases and framing effects have fallen close to the conceptual pole. As such, theories and results from research on attention have rarely been applied to these biases. By extending previous work to somewhat simpler and more perceptual tasks, we can possibly provide a grounding of the some aspects of the biases in perceptual and attentional mechanisms (see General Discussion).

As with Tversky and Kahneman's (1981) demonstrations of framing effects, evidence in favor of perceptual framing effects is based on between-subjects inconsistency. Evidence for systematic biases is uncovered if different groups of subjects tend to give judgments that are mutually inconsistent. We use "bias" to refer to decisions that systematically depart from ideal solutions. Researchers have argued that biases and heuristics play an adaptive and salutary role (Cohen, 1981; Klayman & Ha, 1987; Tweney & Doherty, 1983) for an organism. In the tasks to be described, information-processing limitations preclude optimal or perfect performance. The biases discussed here are probably the result of reasonable strategies for approaching very difficult tasks.

Subjects are asked to make assessments of the percentage or numerosity of certain features in a display. Subjects are asked to select the display with the greater number (percentage) of black/white squares (or digits/letters) or are asked to estimate the percentage of squares in a display that are black/white. Evidence for two interpretations of confirmation bias may be found. First, subjects may be biased to attend selectively to the feature mentioned in the instructions, as suggested by analogy to framing effects and the matching bias. This is called a "feature salience bias." If subjects are told to estimate the percentage of black squares, for example, they may attend to black squares in the display more than the white squares. There is evidence for this bias if the sum of the estimate of black squares and the estimate of white squares exceeds 100%. The claim is that whatever features are mentioned in the instructions are more likely to be noticed, or receive a greater weight, than nonmentioned features.

Second, subjects may be biased to attend selectively to stimulus regions that have a high concentration of instruction-mentioned features. This is called a "region salience bias." For example, if one region of the display has many black squares and the instructions explicitly mention black squares, then subjects would selectively attend to squares within this region more than to other squares.

This region salience bias makes several different experimental predictions from the feature salience bias. First, it does not necessarily predict that the sum of the estimate of black squares and the estimate of white squares will exceed 100% if the squares are well distributed. If all regions have approximately the same concentration of the instruction-mentioned feature, then all regions should be approximately equally salient. Second, the region salience bias predicts that the distribution of features in a display should influence estimates. Specifically, displays that have same-valued features clustered close to each other should receive higher estimates than displays that have randomly distributed features. If a stimulus contains clusters of instruction-mentioned features and if a region salience bias exists, then the features within those regions will be salient. Because there are (for the most part) instruction-mentioned features in these regions, percentage estimates of instruction-mentioned features will be high. Thus, an interesting prediction of the region salience bias is that displays with clusters of black and white squares will be judged to have greater percentages of both black and white squares than a display with randomly distributed squares. A third corollary prediction of the region salience bias is that features that are not mentioned in the instructions may be more salient than features that are mentioned in the instructions. Consider the case in which subjects are asked to estimate the percentage of black squares in a display. According to feature salience bias, all black squares will be

more salient than all white squares. According to the region salience bias, white squares that are in predominantly black regions will be more salient than black squares that are in predominantly white regions.

The following experiments address these differential predictions of the feature and region salience biases. In addition, the experiments bear on the question of how best to characterize a perceptual confirmation bias. The perceptual biases introduced here are most safely viewed as analogous to, rather than instances of, the confirmation biases discussed in hypothesis testing and reasoning domains (see General Discussion). In the current experiments, subjects entertain hypotheses only in the loose sense that they devise a plausible estimate. However, the proposed perceptual biases are similar to traditionally discussed confirmation biases and framing effects in that there is a tendency to base judgments on the particular properties that are mentioned in the instructions or cover story. Given this similarity, two competing accounts of confirmation biases can be tested for the perceptual stimuli used. Under the first account, confirmation biases are due to an insufficient search for disconfirming evidence. By the second account, confirmation biases occur because disconfirming evidence is ignored even when presented. This second account, like the feature salience bias, predicts that, if given black-focusing instructions, subjects will neglect white squares in predominantly black regions. The first account would be consistent with the finding that these same white squares are more salient than are black squares from white regions. By the first account of confirmation bias and the region salience bias, subjects are biased in where they look for evidence, but they do not discount evidence that is obtained once they attend to it.

The difference between a feature salience bias and a region salience bias was anticipated by Tweney, Doherty, and Mynatt (1981), who distinguished between the failure to seek disconfirmatory evidence and the failure to use disconfirmatory evidence when it is available. A region salience bias implies a failure to seek evidence about features not mentioned in the instructions. It does not imply a failure to use such evidence once it has been discovered. A pure region salience bias suggests that "disconfirming" (i.e., information not mentioned in instructions) evidence is used once it is attended but is systematically neglected by search strategies that focus on "confirming" (i.e., instruction-mentioned) evidence. A feature salience bias implies a general failure to attend to evidence from features not mentioned in the instructions.

In summary, the current estimation experiments serve four purposes. First, they extend the range of framing effect and confirmation bias theories to tasks that are more perceptually based than previous tests of the effects. This serves to generalize previous results and to place constraints on a general theory of biases that can account for the extended results. Second, the experiments test different interpretations of the confirmation bias. In particular, the experiments allow us to distinguish between the failure to search for disconfirming evidence and the failure to use disconfirming evidence when it is available. Third, the experiments provide a link between traditional explanations of high-level biases and work in attention and perceptual organization. Many of the hypotheses put forth to explain confirmation bias effects are only applicable in cases in which explicit hypothesis formation occurs. The current experiments suggest an attentional underpinning for at least some cases of confirmation bias.

Experiment 1

The first experiments attempt to obtain evidence for the unique hypothesis of the region salience bias: that clustered displays will be judged to have greater percentages of instruction-probed features than randomly distributed displays. The experiment uses a forced-choice method. Two displays, each with two feature types, are presented simultaneously, and subjects are asked to choose the display with a greater percentage of one of the features. One of the displays has randomly distributed features, and the other display is designed to have clusters of features of the same type. If a feature is more salient, it belongs to a region with a high concentration of instruction-queried features; then the clustered displays will tend to be selected over the distributed displays regardless of which feature is mentioned. Although subjects know that all displays contain the same number of parts and that all parts are of Type 1 or Type 2, according to the region salience bias, they will inconsistently judge the clustered display to contain a greater percentage of both Type 1 and Type 2 features.

Method

Subjects. Forty-eight undergraduate students from the University of Michigan served as subjects to fulfill a course requirement.

Materials. Figure 1 shows examples of the two possible display types. Displays were composed of 60 black and white squares framed by a 7.6 cm × 11.4 cm rectangle. Each square had an area of 0.25 cm^2. Displays were presented on a Macintosh SE computer with a white background. Subjects sat approximately 55 cm from the screen. A program created the displays during the experiment according to two constraints. First, the centers of any two squares were always separated by at least 0.40 cm. Second, the number of black (or white) squares varied from 27 to 33 according to a uniform random distribution. Within these two constraints, the location and shade of squares were randomized.

To obtain a clustered display, displays as described previously were transformed to increase the probability of identically shaded squares being close to each other. A target square was selected at random. The square's nearest neighbor was found. If the square and its nearest neighbor possessed the same shade, no transformation was applied, and another target square was selected. Otherwise, a square with the same shade as the target square was selected at random from the rest of display. The shades of this square and the nearest neighbor were swapped. A square's shade could be changed more than once if it was the nearest neighbor of two squares with different shades. This procedure was repeated 15 times. No transformations were required to produce distributed displays.

Procedure. Subjects were given the following instructions: "You will see two displays on the computer screen. Each display contains 60 boxes in all. Some of the boxes are white and some are black. Your task is to decide which display has a higher percentage of black [or white] boxes in it."

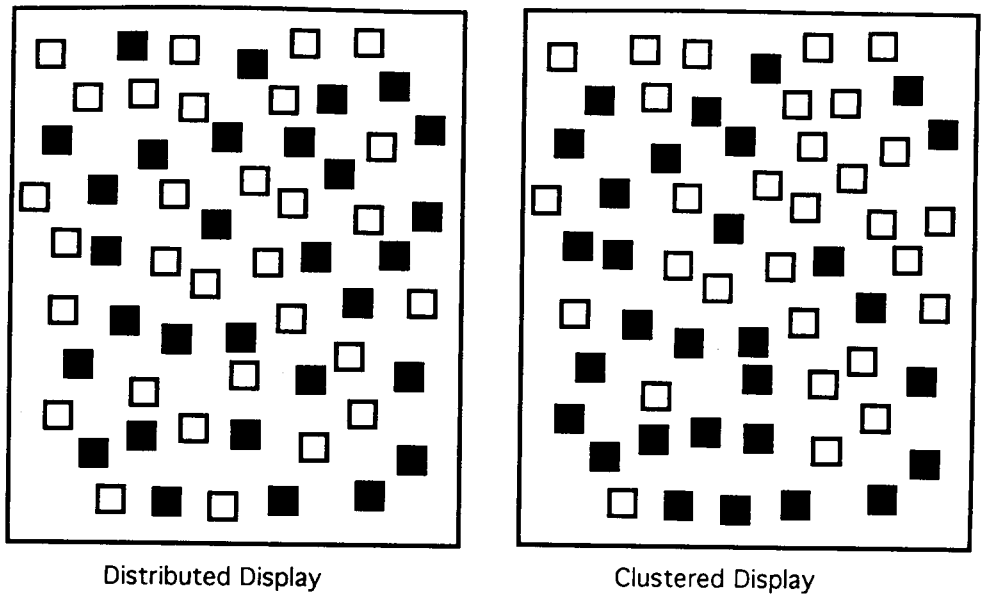

Figure 1. Sample distributed and clustered displays from Experiment 1. (In the clustered display, there is a tendency for squares with the same shade to be close. Subjects who are asked to decide "Which display has a greater percentage of —— boxes?" tend to select the clustered display when the blank is filled in with either "black" or "white.")

Half of the subjects were told to choose the display with a greater percentage of black boxes, and the other half were told to choose the display with a greater percentage of white boxes. Subjects were told to press the "Q" key if they believed the left display had the higher percentage of black (or white) boxes and "P" if they believed the right display had the higher percentage of black (or white) boxes. Subjects were reminded that all of the displays contained the same number of boxes. Consequently, if a display had a greater percentage of black (or white) boxes than another display, then it also would have a lower percentage of white (or black) boxes. Subjects were also instructed not to spend more than 10 s on a trial. It was stressed that subjects were not being asked to count the boxes but only to make a rough assessment.

Subjects saw 80 trials. On each trial, one distributed display and one clustered display were shown side by side. Below the displays, the question appeared: "Which display of 60 boxes contains a greater percentage of white [or black] boxes?" Subjects were always queried with the same feature on every trial; subjects were randomly assigned to either a black- or white-instructions group. The left or right position of the two displays was randomized. Because each display randomly varied in the number of black boxes from 27 to 33, on approximately six sevenths of the trials there was a correct answer. On the remaining trials, both displays contained the same percentage of black boxes.

After the 80 trials were completed, subjects were asked, "Did you notice anything about the way the boxes were organized in the displays?" If the experimenter received a negative reply, a figure similar to Figure 1 was shown to subjects and they were asked, "Did you notice that some displays had many boxes with the same color clumped together like the display on the right? If so, did you tend to choose these displays or not?"

Results

The principal question tested by Experiment 1 is: "Do subjects tend to choose the clustered display as containing a greater percentage of black/white boxes than the distributed display?" The results confirm this hypothesis. When subjects are told to select the display with a greater percentage of white boxes, the clustered display is selected 55.9% of the time. Similarly, when told to select the display with a greater percentage of black boxes, the clustered display is chosen 56.2% of the time. These percentages do not significantly differ, and in further analyses, data from the two instruction conditions are combined. The overall percentage of clustered display choices (56.0%) is significantly greater than 50%, two-tailed $t(47) = 3.78$, $p < .05$. For 41 of the 48 subjects, clustered displays are chosen more often than distributed displays (binomial $p < .05$).

Because the computer generates random displays on-line, subjects will not receive the same number of trials that have a greater percentage of instruction-mentioned squares in the clustered display. We can compare the number of trials in which a subject should choose the clustered display with the number of trials in which the subject does choose the clustered display. Forty of 48 subjects choose the clustered display more frequently than they should (binomial $p < .05$).

Figure 2 shows that subjects are sensitive to small differences in number of squares between two displays. The horizontal axis represents the relative numerical superiority of the clustered display along the probed dimension. Because the distributed and clustered displays both contain between 27 and 33 black (or white) squares, the two display types never differ by more than 6 squares of a particular shading. The general upward trend of the line shows that as the number of black (or white) boxes in the clustered display increases relative to the distributed display, subjects choose the clustered display as containing a larger percentage of black (or white) boxes with increasing likelihood.

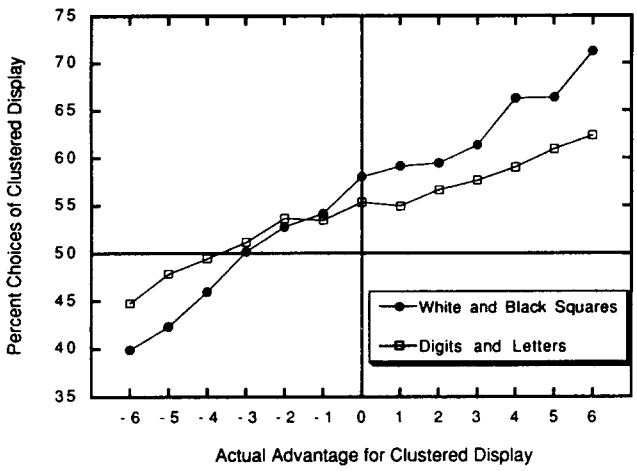

Figure 2. Results from Experiment 1. (The probability of subjects selecting the clustered display increases as the actual numeric advantage of the clustered display over the distributed display increases. More important, the clustered display is chosen more often than the distributed display at all levels of numeric advantage.)

Figure 2 also shows that subjects are biased to select the clustered display over the distributed display. The best fitting regression line is:

Percentage of choices of clustered display = 2.4(actual advantage of clustered display) + 56.0.

If there were no bias for subjects to respond by selecting the clustered display, the line would be 6.0 percentage points lower on the vertical axis and would pass through the point (0 advantage for clustered display, 50% clustered display choice). Even when there is a small numerical advantage for the distributed display, choices for the clustered display still exceed 50%.

The tendency to select the clustered display over the distributed display decreases with practice. For the first 40 trials, the clustered display is chosen 57.9% of the time; this decreases to 54.1% for the second 40 trials, paired two-tailed $t(46) = 2.58, p < .05$. Response accuracy also increases with practice. For the first 40 trials, the correct display is chosen 53.9% of the time; for the second 40 trials, the correct display is chosen 56.5% of the time, paired two-tailed $t(46) = 2.15$, $p < .05$. This trend for decreased susceptibility to illusions caused by the distribution of items in an estimation task was also found by Ginsburg (1976).

A fair number of subjects did notice that some displays were clustered and others were more distributed. Seven of 48 subjects noticed the different distribution of shades when prompted with the open-ended question, and 10 more subjects said that they noticed the difference when explicitly prompted with "Did you notice that some displays had many boxes with the same color clumped together?" Interestingly, the bias to choose clustered displays was less pronounced for these subjects than for those who did not verbally show awareness of the clustered–distributed difference (54.5% and 56.8% choices of the clustered displays, respectively), two-tailed $t(46) = 1.68, p < .05$.

Subjects spent an average of 6.5 s on each display. Subjects were split into two groups on the basis of whether their average response time was faster (mean = 5.1 s) or slower (mean = 8.3 s) than the median response time. The groups did not significantly differ on the magnitude of their preference for clustered displays (slow group = 56.4% clustered choices; fast group = 56.0% clustered choices).

Discussion

The results are consistent with the hypothesis that subjects have a tendency to selectively attend to stimulus areas that have a high concentration of instruction-mentioned features. A between-subjects inconsistency arises. Subjects tend to select the clustered display as having both a greater percentage of black squares and a greater percentage of white squares. Clustered displays are judged to have a higher percentage of white squares about 56% of time and are judged to have a higher percentage of black squares as often. This yields a between-subjects consistency violation. The same display cannot have both a greater percentage of black and white squares, given that it is quite clear to subjects that displays only contain these two squares. It appears that when subjects are instructed to judge two displays with respect to "percentage of *x*," they selectively look at, or give greater weight to, regions with a relatively large amount of direct evidence of *x*. This occurs even though the prevalence of non-*x* stimulus parts is normatively just as important for the judgment as the prevalence of *x* parts are.

A number of objections must be addressed. First, it might be expected that percentages are difficult for subjects to calculate; consequently, subjects may resort to the cognitively easier process of estimating numerosity. As such, subjects may interpret "Which display has a higher percentage of black boxes" as "Which display has more black boxes?" This possibility is less likely than it might otherwise be because it was twice stressed to subjects before the experiment that they were to make percentage judgments. In addition, on every trial, the question "Which display has a higher percentage of black boxes" appeared. Furthermore, even if subjects were making sheer numerosity judgments, the between-subjects responses are still inconsistent as long as subjects understood that the number of squares was the same for all displays. The fact that the two displays each contained 60 squares was also shown on every trial. Thus, it is still inconsistent for subjects to select the clustered display as containing both more black squares and more white squares, given that the total number of squares is fixed.[1]

[1] Replications of Experiments 1, 2, and 3 were conducted with numerosity, rather than percentage, instructions. For example, in the replication of Experiment 1, subjects were asked to "choose the display that has a greater number of black [or white] squares." Again, subjects were told that the total number of squares in a display was held constant at 60. In all three replications, numerosity estimates were greater for clustered displays than for distributed displays. Thus, the results from the reported experiments do not appear to be an artifact of the percentage estimate task.

Second, it is not sufficient to explain the effect of clustering by claiming that both white and black squares are more salient when clustered than when distributed. One version of this claim might be that a square with Shade x is more influential if surrounded by other squares with Shade x because together they form a salient "region of Shade x" feature. This explanation fails to take into account the basic symmetry of black and white square distribution for clustered displays. When displays are clustered, then both white and black squares are clustered to the same extent. If both shades are more salient when clustered, then it is unclear why the clustered display would be chosen as having a higher percentage of one of the shades than the distributed display. In contrast to a general increase in square salience with clustering, the results suggest that stimulus areas with a high concentration of the shade mentioned in the instructions are selectively attended.

Third, it is not sufficient to claim that the shade mentioned in the instructions is more salient than the unmentioned shade, as claimed by the feature salience bias. This bias, by itself, does not explain the effect of shade distribution on choices. The effect of shade distribution indicates that same-valued squares are not equally salient in all spatial configurations. The proposed account is that squares are more salient if they belong to areas that have a high concentration of instruction-mentioned shades. As such, if the instructions mention black squares, then both black and white squares that are located in a region with many black squares will be particularly salient. This hypothesis is directly tested in Experiment 4.

Fourth, the claim that the influence of shade distribution is strategic and based on a task demand can be rejected. A postexperimental interview served to divide subjects into two groups: those who showed verbal awareness of the clustered–distributed distinction and those who did not. Those subjects who did not show verbal awareness showed more of a bias to select clustered displays than the distinction-aware subjects. A possible reason for this difference comes from an observation cited by a few of the distinction-aware subjects. This observation was expressed by one subject as follows: "I figured that you were trying to get me to choose [the clustered display] by having the black boxes close to each other." Thus, the distinction-aware subjects may have been less likely to choose the clustered displays because they realized that there would be an incorrect tendency to see the clustered displays as having a greater percentage of the instruction-mentioned shade. A similar phenomenon was noted by Jacoby and Whitehouse (1989), who found that subjects who consciously remember seeing a word gave qualitatively different responses (ratings of fame) than unaware subjects. At any rate, the evidence suggests that clustering shades has a stronger effect when subjects are unaware of the clustering.

Experiment 2

Experiment 2 is an attempt to replicate the results from Experiment 1 using targets with more abstract definitions. In Experiment 1, subjects were told to make judgments based on the relative percentages of black and white squares. The shade difference is highly salient, and one might well expect black and white squares to be discriminated in parallel and with great speed (Treisman & Gelade, 1980). Perhaps the observed clustering effect is obtained only when the target feature (the feature mentioned in the instructions) is distinguishable on the basis of low-level, physical cues. On the other hand, Duncan (1983) argued that attention can be directed selectively to letters and digits even though these two groups cannot be distinguished by low-level features. Experiment 2 tests whether the bias obtained in Experiment 1 is evident when digits and letters are used as the two feature types. One group of subjects is told to "select the display that has a greater percentage of digits," whereas a second group is told "select the display that has a greater percentage of letters." There are no necessary and sufficient primitive physical features that serve to discriminate the two groups. If a tendency to select clustered over distributed displays is still found, then the results seem to apply to both physically and categorically defined targets.

Method

Subjects. Forty-four undergraduate students from the University of Michigan served as subjects to fulfill a course requirement.

Materials. Displays were composed of 60 letters and digits framed by a 7.6 cm × 11.4 cm rectangle. The eight digits "2," "3," "4," "5," "6," "7," "8," and "9" were used with equal frequency, as were the letters "A," "B," "C," "D," "E," "G," and "K." The letters and digits were approximately 0.5 cm tall. The basic method for creating displays was identical to Experiment 1.

To obtain a clustered display, a target symbol was selected at random from the entire set of displayed symbols. The symbol's nearest neighbor was found. If the symbol and its nearest neighbor were both digits or both letters, no change occurred and another target symbol was selected. Otherwise, a symbol of the same type (digit or letter) as the target symbol was randomly located. The locations of this symbol and the nearest neighbor were swapped. This procedure was repeated 15 times.

Procedure. Subjects were given the following instructions: "You will see two displays on the computer screen. Each display contains 60 letters and digits. Your task is to decide which display has a higher percentage of letters [or digits] in it." No postexperimental interview was conducted. In all other ways, the procedure was identical to Experiment 1.

Results

When subjects were told to select the display with the greater percentage of digits, the clustered display was selected 54.7% of the time. With letter instructions, the clustered display was chosen 54.4% of the time. The overall percentage of clustered display choices (54.5%) was significantly greater than 50%, two-tailed $t(43) = 3.4$, $p < .05$. Thirty-three of 44 subjects selected clustered displays more often than distributed displays (binomial $p < .05$). Similarly, 34 of 44 subjects selected clustered displays more often than they should have, given the actual randomly generated displays shown.

Figure 2 shows the percentage of clustered display choices as a function of the actual numerical advantage of the clustered display along the instruction-mentioned feature. As

with Experiment 1, there is evidence that subjects are influenced both by the actual numerical advantage of a display and by the distribution of stimulus parts in the displays. When there is no actual advantage to either display, the clustered display is chosen on 55.4% of the trials.

As with Experiment 1, the clustered display is chosen less frequently with practice. The percentage of clustered display choices decreases from 55.7% to 53.3% from the first 40 to the second 40 trials, paired two-tailed $t(42) = 2.46, p < .05$. The percentage of correct choices increases from 52.8% to 55.6%, paired two-tailed $t(42) = 2.83, p < .05$. The average response time in Experiment 2 is 7.6 s. Faster- and slower-than-average subjects did not significantly differ in their preference for clustered displays (slow subjects = 54.8% clustered choices; fast subjects = 54.2% clustered choices).

Discussion

Experiment 2 extends the results of Experiment 1 to materials that have relatively abstract types of parts. When subjects decide which of two displays has a greater percentage of a certain type of element, subjects tend to select the display that has elements of the queried type concentrated in a few regions. This result holds when the different parts are distinguishable on the basis of unidimensional physical differences and also when they are distinguished by their conceptual grouping. The hypothesized explanation is that regions with a high concentration of instruction-mentioned parts are salient to subjects, and the parts within these regions have a disproportionate influence on subjects' choice of displays. Again, it is not sufficient to claim that all instruction-mentioned parts are made salient. Rather, the salience of a region is based on the type of part mentioned in the instructions.

Experiment 3

Although the forced-choice paradigm of Experiments 1 and 2 offers striking evidence of inconsistent between-subjects responses, it also suffers from shortcomings. Forced-choice judgments may be based on factors unrelated to the explicit instructions given to subjects. For example, subjects may have a tendency to choose the display that is most aesthetically pleasing to them. If it is further assumed that clustered displays are generally more appealing than distributed displays, then the bias to select the clustered display does not require the postulation of disproportionate attention to particular regions.

Because of the alternative explanations possible for clustered display choices within a forced-choice paradigm, converging evidence is needed. In Experiment 3, subjects are shown single displays and are asked to estimate the percentage of black or white squares in the display. The displays again vary in the degree to which squares of the same shade are close to each other. The prediction, according to the region salience bias, is that displays with many close, like-shaded squares (clustered displays) will be judged to have a greater percentage of the probed shade than distributed displays. Even though the displays are composed of only black and white squares, and therefore the percentage of black and white squares must total 100%, the combined percentage estimates of subjects responding to clustered displays is expected to exceed 100%. The alternative explanations given to Experiments 1 and 2 are not tenable if the expected results from Experiment 3 are obtained. Subjects are not asked to select a display. Instead, they provide a percentage estimate for each display. As such, dispositions to choose the clustered display over the distributed display for aesthetic or other reasons should not bias the current experiment.

Method

Subjects. Thirty-four undergraduate students from the University of Michigan served as subjects to fulfill a course requirement.

Materials. The same materials as in Experiment 1 were used, with two exceptions. First, three different levels of feature distribution were used: distributed, moderately clustered, highly clustered. Moderately clustered displays were obtained by swapping 12 squares as described in Experiment 1. Highly clustered displays were obtained by swapping 24 squares. In general, highly clustered displays are more clearly differentiated into black and white areas than are moderately clustered displays.

Second, displays had greater variation in number of black and white squares than did displays in Experiment 1. The number of black boxes in a scene varied according to a uniform random distribution from 15 to 45 (from 25% to 75%).

Procedure. Subjects were given the following instructions: "You will see a scene appear on the screen for a short period of time (1 s).[2] The scene is composed of 60 boxes all in all. Some of these boxes are black and others are white. Your task is to estimate the percentage of boxes in the scene that are black [or white]."

Two examples were presented to clarify the meaning of "percentage." Subjects were told to imagine scenes with 30 and 40 black boxes and were told that the best estimates for these scenes would be 50% and 66%, respectively.

Subjects were shown 80 scenes. A scene appeared on the screen for 1 s. Immediately after the scene was erased, the question "What percentage of the scene's boxes were black [or white]?" appeared. The subject then typed a number between 0 and 100. After a pause of 4 s, the next scene was presented. Subjects were evenly split into white and black instruction groups. On any given trial, distributed, moderately clustered, and highly clustered displays were each shown with a probability of 1 of 3.

Results

The results indicate that higher percentage estimates are given as the displays become more clustered. Overall, average estimates are: distributed displays = 48.6%, moderately clustered displays = 52.7%, and highly clustered displays = 55.0%. Each of these estimates is significantly different from the true average percentage estimates, which were all approximately 50%, two-tailed $t(33) > 3.2, p < .05$. Twenty-seven of the 34 subjects gave estimates for the highly

[2] A replication of Experiment 2 was run with 5 s viewing time. Estimates were somewhat more accurate, but there was no interaction between viewing time and display type on percentage estimates.

clustered display than were too high (binomial $p < .05$), and 22 subjects gave estimates that were too low for the distributed displays. Thirty subjects gave higher estimates for clustered displays than for distributed displays; an analysis of the displays shows that only 16 subjects should have given higher estimates for clustered displays. These estimates include both black- and white-instructions groups because these groups did not differ significantly in their estimates for each display type. These results indicate that subjects shown distributed displays significantly underestimate the percentage of both black and white squares that are present. Conversely, subjects shown clustered displays significantly overestimate the percentage of both black and white squares.

Unlike Experiments 1 and 2, there is no interaction between practice and display type. Instead, there is a main effect caused by practice such that all percentage estimates decreased with increasing trials (average estimate for the first 40 trials = 52.9%, for the last 40 trials = 51.4%), paired $t(32) = 2.4$, $p < .05$. An interpretation of the results that reconciles those from Experiments 1 and 3 is that, in both cases, practice increases accuracy. The average correct percentage of black [or white] squares is 50%. Subjects' estimates come closer to this percentage as they make more estimates, even though no feedback is given concerning the correctness of the estimates.

Figure 3 shows average estimates plotted as a function of the actual percentage of a probed shade (i.e., the shade that was mentioned in the instructions). The graph illustrates the general trend for more clustered displays to yield higher estimates than less clustered displays. In addition, the graph shows that this trend depends on the actual percentage of the shade in the scene. That is, in predicting subjects' estimated percentages, there is an interaction between display type and the actual percentage of the probed shade, $F(3, 32) = 4.2$, $MS_e = 120$, $p < .05$. For example, if there are few probed squares in a scene, then clustered and distributed scenes yield similar estimates for the percentage of black squares in the scene. The increased estimates for clustered scenes come only when the scenes have many squares with the probed shade. This is shown in Figure 3 by lines that diverge with increasing probed shade percentages.

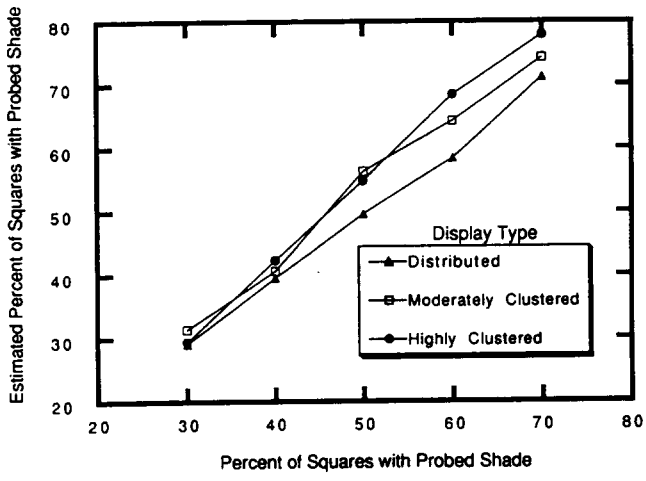

Figure 3. Actual and estimated percentages of probed features (the features mentioned in the instructions) from Experiment 3. (Although the percentage of squares with a probed shade can vary freely from 25% to 75%, the data are grouped into five categories, for example, 25–35 → 30, 35.01–45 → 40, and so on. The graph show that clustered displays receive greater estimates than do distributed displays.)

Discussion

Experiment 3 supports the general conclusion from Experiments 1 and 2. Subjects tend to overestimate the percentage of black and white squares in clustered displays. Subjects also underestimate the percentage of black and white squares from distributed displays, but to a lesser degree. These systematic estimation errors have been interpreted in terms of a tendency for selective attention to be given to regions that have a high proportion of instruction-mentioned items. This tendency results in a between-subjects inconsistency. For clustered scenes, subjects' average estimate of the percentage of black squares is about 55%. A different group of subjects gives approximately the same average estimate for the percentage of white squares in the scene. Both of these figures cannot be correct on average or the scene would be 110% filled with squares. Most likely, if the same subject were asked, "What percentage of the boxes are black boxes?" and "What percentage of the boxes are white?" for the same scene, then no inconsistency would arise. If there were a tendency to overestimate percentages, the tendency would be suppressed by the knowledge that percentages must sum to 100%. However, the between-trials/between-subjects experiment design permits observation of inconsistencies that are not suppressed because the inconsistencies are not evident to the subject on any given trial.

Although the subjects displayed systematic biases in their estimates, the results also point to a generally high degree of accuracy in estimates. The accuracy of the estimates is particularly surprising given that subjects are shown a scene for only 1 s; there is no opportunity for subjects to count the squares of a particular shade. Collapsing over distributed, moderately clustered, and highly clustered display types, subjects' estimates of a shade's percentage closely follow the actual percentage of the shade. If we had observed a tendency to overestimate percentages for all three display types, we might have concluded that all objects with the shade value mentioned in the instructions are made selectively salient (the feature salience bias). This claim can be rejected, as it was in Experiments 1 and 2. Percentage estimates for distributed scenes are actually slightly underestimated. This underestimation implies that if the instructions require subjects to judge the percentage of black squares, then black squares are not necessarily more salient than white squares. Instead, the instructions make squares from regions with many black squares more salient.

This region salience bias can also account for the interaction between display type and actual percentage of the probed shade. If there are few objects with the probed shade, then it makes little difference whether the display is clustered or distributed. In either case, objects with the nonprobed

shade will often be close to a probed object that is the focus of attention at one time. If there are many objects with the probed shade, then the display type will have a larger influence. If the probed shades are clustered within a particular region, and if subjects selectively attend to that region, then percentage estimates will be high because very few nonprobed shades will encroach into the region. If the probed shades are distributed across many regions, then nonprobed shades will always be close to a probed shade, decreasing estimates of the probed shade's prevalence. In short, the region salience bias, combined with statistically verified properties of clustered and distributed scenes, predicts that shade distribution influences percentage estimates more when displays have many objects with the probed shade.

The underestimation of distributed displays was not anticipated. One possible explanation is that subjects assume that, on average, displays will contain 50% black [or white] squares. Because subjects overestimate the percentage of probed features for clustered displays, they must give systematically underestimated percentages for distributed displays to arrive at an overall average of 50%.

Experiment 4

The preceding experiments have suggested that subjects are biased to attend selectively to spatially defined regions that have many objects with the feature mentioned in the task instructions. This hypothesis is distinguished from the claim that subjects are biased to attend to objects with features that are mentioned in the instructions. The influence of feature distribution implies that subjects' attention to regions of a stimulus depends on the regions' featural composition. Experiment 4 is designed to be a more direct test of this hypothesis. In Experiment 4, squares of different shades are individually removed from clustered scenes. Squares can be removed from regions that have many or few squares with the probed feature. Orthogonally, squares can be removed that have the probed value or not. The current hypothesis is that removing squares from regions with many instruction-mentioned values will have a larger effect on estimates than removing squares from regions with few instruction-mentioned values. The design of the experiment explicitly allows this hypothesis to be distinguished from the hypothesis that all instruction-mentioned values have a larger effect on estimates.

In addition, Experiment 4 can discriminate between two versions of the region salience bias. According to both versions, squares with the probed value are more salient if they are clustered in a particular region. According to one version, the higher estimates in the clustered condition over the distributed condition occur because it is difficult to attend to many noncontiguous regions. Because clustered displays have fewer noncontiguous regions with black squares, they receive higher estimates of black squares. Thus, this version's claim is that distributing attention across many regions is difficult. According to the other version, the higher estimates in the clustered condition are due to a tendency to process everything within an attended region. That is, ignoring a value that is near to an attended value is difficult or unnatural. Loosely speaking, the first claim states that "clustered scenes get high estimates because people cannot find or notice all of the black squares in a distributed display," whereas the second claims that "clustered scenes get high estimates because people cannot help noticing the white squares that are close to black squares in distributed displays." The two versions may seem logically equivalent, but Experiment 4 gives grounds for differentiating the claims by observing the effects of nonprobed values in regions with many probed values and probed values in regions with many nonprobed values. If distributing attention across many probed values in noncontiguous regions is difficult, then probed values in regions with many nonprobed values should have little effect on estimates. If ignoring values that are near probed values is difficult, then nonprobed values in regions with many probed values should be highly salient. Both of the claims may be true.

Experiment 4 provides a more direct measure of the salience of individual squares than earlier experiments. By comparing scenes with no squares removed to scenes that have a few squares removed, the relative influence of specific types of squares can be determined. Experiment 4 quantifies the influence of particular sources of information on percentage estimates.

Method

Subjects. Twenty-seven undergraduate students from Indiana University served as subjects to fulfill a course requirement.

Materials. The control condition display, the display in which no squares are removed, was generated in the same manner as in Experiment 1. Displays consisted of 60 black and white squares. The number of black squares varied uniformly from 27 to 33. The clustering of the control display was obtained by swapping 15 squares in the manner previously described.

Six other types of display were generated from the control display. Each of these display types eliminated 2 of the original 60 squares. The six display types were as follows: two white squares removed from black regions, two squares removed from white regions, two white squares removed from neutral regions, two black squares removed from black regions, two black squares removed from white regions, and two black squares removed from neutral regions. Examples of squares falling into these six types are shown in Figure 4. In determining whether a square fell in a black, white, or neutral region, the proximity of the square to black and white squares was computed using the following equation:

$$P_{s,c} = \sum_{i \in \{c\}} e^{-((x_i - x_s)^2 + (y_i - y_s)^2)/2}$$

where $P_{s,c}$ is the proximity of square s to the set of squares c (either the set of black squares or the set of white squares), and x_i and y_i are the horizontal and vertical coordinates of square i (for further explanation and justification of this formula, see Nosofsky, 1986). The Euclidean distance between squares is converted to an exponential function in determining proximity to capture the intuition that close neighbors will have substantially greater influence in determining a square's region than will distant neighbors. A square falls in a black region if $(P_{s,\text{black}} - P_{s,\text{white}}) > 25$. A square falls in a white region if $(P_{s,\text{white}} - P_{s,\text{black}}) > 25$. If neither of these conditions applied, the square falls in a neutral region.

The six display types were obtained by randomly selecting a square and calculating $P_{s,c}$ values to determine the square's region.

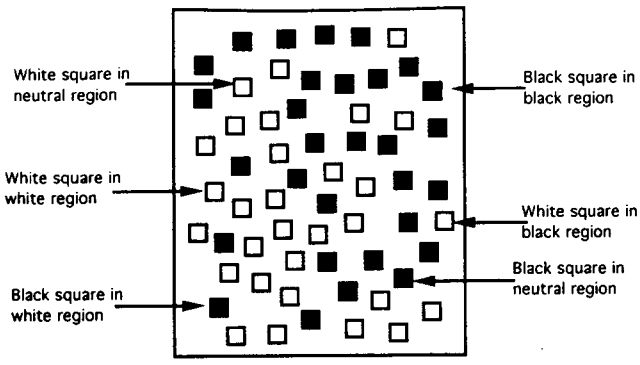

Figure 4. Examples of the six types of squares that are removed from control displays in Experiment 4. (Either black or white squares can be removed, and the squares can be removed from black, white, or neutral regions.)

If the square's region and shade corresponded to the desired display type, the square was deleted from the display before the display was shown. This procedure was repeated two times.

Procedure. The procedure closely followed that of Experiment 3. Subjects were instructed to "estimate the percentage of boxes in the scene that are black." Unlike the previous experiments, no white-instructions group was run because of the strong mirroring between white- and black-instructions groups previously observed.

Subjects were shown 80 scenes. A scene appeared on the screen for 5 s. Immediately after the scene was erased, the following question was presented: "What percentage of the scene's boxes were black?" The subject then typed a number between 0 and 100. The six display types and the control display were each presented with a probability of one of seven. The particular display type shown on any given trial was randomized.

Results

Figure 5 shows the estimates given for each of the six display types and the control display in which no squares were removed. Subjects were always asked to judge the percentage of black squares in a display. Thus, as anticipated, when white squares are removed from the control display, estimates increase; when black squares are removed, estimates decrease. Although the average actual percentage of black squares in the displays was approximately 50%, subjects' estimated percentage is 52.5%, two-tailed $t(28) = 6.1$, $p < .05$. Overall, removing black squares decreases estimates as much as removing white squares increases estimates: no squares removed = 52.6%, black squares removed = 49.9%, and white squares removed = 55.0%, two-tailed paired t test on mean difference from control condition, $t(26) = 1.9$, $p > .1$.

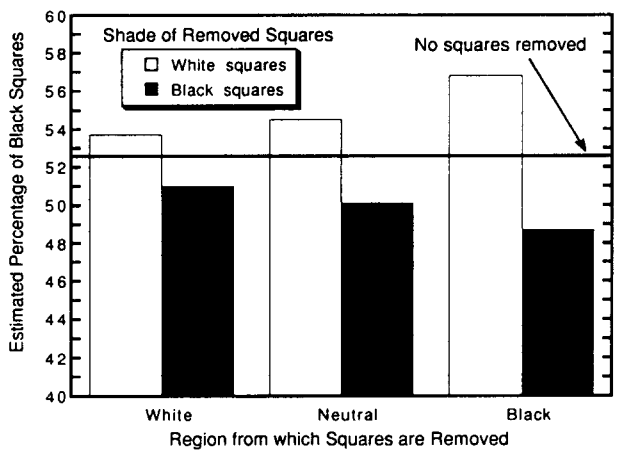

Figure 5. Results from Experiment 4. (When instructions direct subjects to estimate the percentage of black squares, black and white squares influence percentage estimates more when they are removed from black compared with white or neutral regions. White squares removed from black regions influence estimates more than black squares removed from white regions do.)

The influence of two white or black squares on estimates is not constant. Instead, the influence depends on the region from which the squares are removed. As Figure 5 shows, both black and white squares influence estimates most when they are removed from regions with many black squares, have intermediate influence in the neutral regions, and have the least influence in regions with clustered white squares. Removing black squares has significantly more effect when they are removed from black regions than other regions, overall $F(2, 27) = 4.2$, $MS_e = 2.4$, $p < .05$. Likewise, removing white squares has significantly more effect when they are removed from black regions than neutral regions and when removed from neutral regions than white regions, overall $F(2, 27) = 5.6$, $MS_e = 2.8$, $p < .05$. There is also a marginally significant two-way interaction between shade of a removed square and region of the removed square, $F(1, 54) = 3.2$, $MS_e = 3.4$, $p < .10$. Essentially, the region of the removed square makes more of a difference when white squares are removed than when black squares are removed.

An analysis of subject responses yields a similar, although weaker, pattern of results. Each subject's estimate for display type was determined and compared with the actual percentage of black squares for the display type. Nineteen of 27 subjects were more influenced by the removal of squares from black regions than from white regions (binomial $p < .05$). Fifteen subjects were more influenced by the removal of black squares than white squares ($p > .05$).

Discussion

As with previous experiments, Experiment 4 gives evidence of a systematic overestimation of the prevalence of a probed features in a clustered display. All of the displays shown were clustered, and subjects gave estimates averaging greater than 50% for the percentage of black squares in the displays.

Relative to the control condition in which no squares are removed, removing black squares has no more of an effect than does removing white squares. Removing two white squares increases percentage estimates of black squares as much as removing two black squares decreases estimates. The effect is inconsistent with the claim that objects become more salient if they have a feature value mentioned in the task instructions. This null result should not be overinterpreted. In determining a square's influence, the region to which the square belongs is more important than the square's featural

identity, but the current experiment may not have sufficient power to uncover a small influence of featural identity.

There are occasions when a white square (the feature not explicitly mentioned in the instructions) is more influential than a black square. A white square that is located in a predominantly black region is more influential than a black square in a white region. This evidence distinguishes between the feature and region salience biases. The feature salience bias predicts all black squares to be more salient than white squares. The region salience bias correctly predicts that white squares may be more salient than black squares if the white squares occur in predominantly black regions.

Both of the formulations of the region salience bias are supported. According to the first formulation, the higher estimates in the clustered condition over the distributed condition occur because it is difficult to attend to many noncontiguous regions. This hypothesis is supported by the finding that black squares are less influential when they come from white, as opposed to black, regions. According to the second formulation, the higher estimates in the clustered condition are due to a tendency to process everything within an attended region. This hypothesis is supported by the finding that white squares are more influential when they occur in black, as opposed to white, regions. In short, clusters of instruction-mentioned features function both to draw attention to the region and draw attention away from other regions.

Experiment 5

Experiment 5 is conducted to eliminate one final criticism of the earlier experiments. In the previous experiments, the extent of clustering of the instruction-mentioned feature is confounded with the extent of clustering of the feature not mentioned in the instructions. For example, in Experiments 1 and 3, as the white boxes become more clustered, so do the black squares. The design of Experiment 5 separates the overall level of clustering from the level of clustering of instruction-mentioned features. Displays consist of three types of features, and subjects are instructed to estimate the prevalence of one of the features. Items with the instruction-mentioned feature may be distributed or relatively clustered. Independently, items with features not mentioned in the instructions may be clustered. By introducing a third feature, it is possible to have displays with clustered features, even though the instruction-mentioned feature is not clustered.

We predict that overestimation will occur only when the instruction-mentioned feature is clustered. If instruction-mentioned features are clustered, then subjects should selectively base their estimates on regions that have many instruction-mentioned features. If features not mentioned in the instructions are clustered, then overestimates should not result because these clustered regions should not attract subjects' attention.

On the other hand, one can argue that the percentage of all items in all clustered displays is overestimated. This hypothesis predicts that displays that have clusters of features not mentioned in the instructions will still be overestimated. Thus, overestimation should be significantly higher for both clustered displays than for the fully distributed display.

Method

Subjects. Thirty-six undergraduate students from Indiana University served as subjects to fulfill a course requirement.

Materials. Displays were similar to those used in Experiment 1, with a few exceptions. First, three shades of squares were used: white, gray, and black. Each display contained between 17 and 23 squares of each shade. The particular number of squares with each shade was determined by independent uniform random distributions. The total number of squares varied between 51 (17×3) and 69 (23×3).

Subjects were shown the three types of displays illustrated in Figure 6 with equal frequency. When all shades were distributed, no swapping of shades occurred. For the other two display types, there were clusters of two shades. In one display type, clusters of two shades were created that were not probed by the instructions. In the other display type, clusters of two shades were created, and one of the shades was probed by the instructions. For example, if subjects were asked to estimate the percentage of black squares in a display, then Figure 6b is an example of a display that has nonprobed shades clustered and Figure 6c is a display that has the probed shade clustered.

Clusters of two of the three shades were created by a small modification of the clustering algorithm in Experiment 1. If shades x and y were to be clustered, then squares with shade x were selected at random. The square's nearest neighbor with shade y was identified, and the two squares' shades were exchanged. This procedure was repeated eight times.

Procedure. Subjects were given the following instructions: "You will see a scene appear on the screen for a short period of time (1 s). The boxes will either be gray, black, or white. Your task is to estimate the percentage of boxes in the scene that are black [or white or gray]." Twelve subjects were assigned to the black-, white-, and gray-instructions conditions each.

Three examples were presented to clarify the meaning of "percentage." Subjects were shown scenes with 3 black squares of 6 squares, 3 black squares of 10 squares, and 5 black squares of 10 squares and were told what the proper percentage estimate would be in each case.

Subjects were shown 80 scenes. The same procedure and randomizations in Experiment 3 were used.

Results and Discussion

Figure 7 presents the results from Experiment 5. There was a significant effect of cluster method on estimated percentages, $F(2, 70) = 7.4$, $MS_e = 0.24$, $p < .05$. Higher percentage estimates were given for displays that have the probed shade clustered than for the other two cluster methods, two-tailed $t(35) > 4.5$, $p < .05$, Bonferroni adjusted. Overall, the percentage estimates were as follows: probed shade clustered = 34.9%, all shades distributed = 33.1, and nonprobed shades clustered = 32.9. A subject analysis was also conducted. For each display type, the true percentages (approximately 33%) and the subject's degree of underestimation (or overestimation) for each cluster method were determined. Twenty-five of the 36 subjects overestimated displays with the probed shade clustered more often than they overestimated fully distributed displays (binomial Z, $p < .05$). Twenty-nine subjects

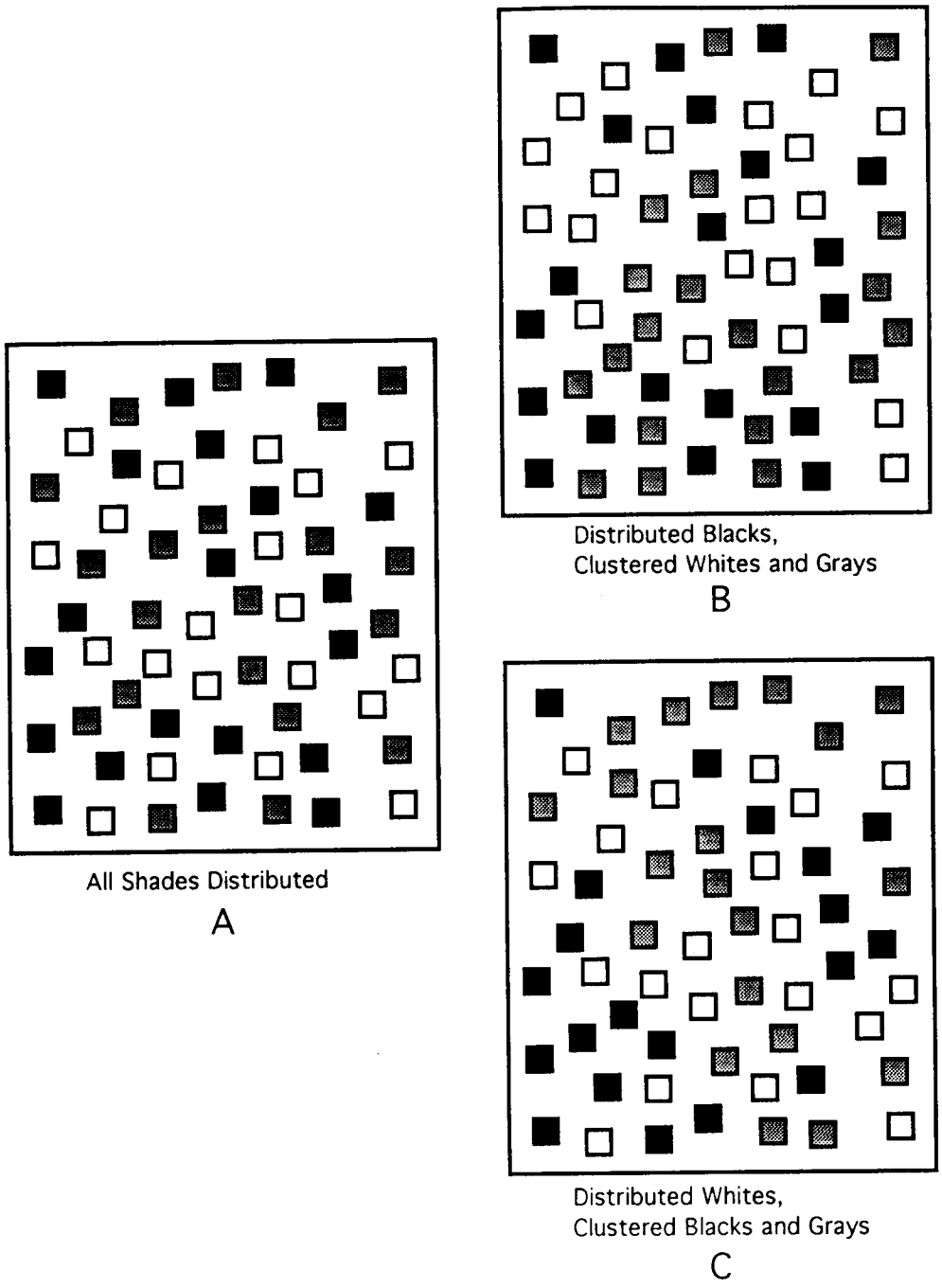

Figure 6. Samples of the three types of displays used in Experiment 5. (All features are distributed in panel a. If subjects are instructed to estimate the percentage of black squares, then panel b has both nonprobed shades clustered and the probed shade distributed. Panel c has the probed shade and one nonprobed shade clustered.)

overestimated displays with the probed shade clustered more often than they overestimated displays with both nonprobed shades clustered (binomial Z, $p < .05$).

There was a marginally significant influence of probed shade on estimation, $F(2, 70) = 2.7$, $MS_e = 0.36$, $p < .1$, with black probe = 33.9%, white probe = 33.7, and gray probe = 33.3. This may be attributable to the somewhat lower discriminability of the gray shades. Black and white squares may "pop out" slightly more than do gray squares. There was no interaction between probed shade and cluster method.

The results of Experiment 5 are consistent with the claim that regions that contain a high proportion of instruction-mentioned features are particularly salient when a subject makes a percentage estimate. Conversely, the claim that any

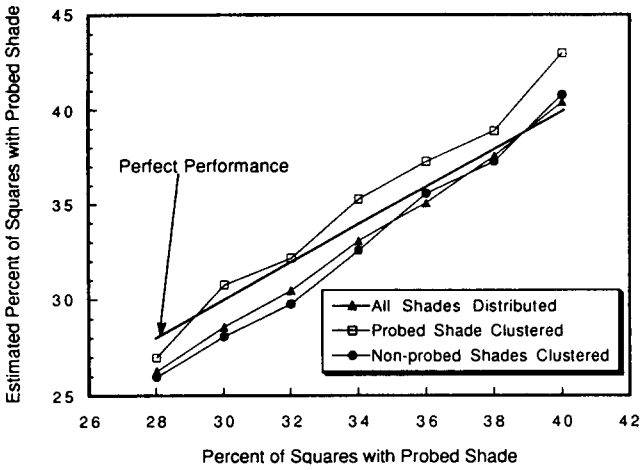

Figure 7. Results from Experiment 5. (Subjects overestimate displays that have the probed shade clustered but do not overestimate displays that have both nonprobed shades clustered.)

clustering produces overestimates of probed features was not supported. Displays that have clustered features that are not probed are not systematically overestimated. In fact, there was some tendency for subjects to give lower estimates for these displays than for the fully distributed displays. The presently important claim is that clustering produces systematic overestimation only when the probed feature is one of the features that is clustered.

General Discussion

The reported experiments provide evidence that subjects demonstrate confirmation biases and framing effects in proportion estimation tasks. Forced-choice judgments and estimations of proportions reveal a tendency for subjects to be influenced more by display components that occur in regions that have a high rather than low concentration of instruction-mentioned features. Evidence is found for the region salience bias from all five experiments. Evidence for the feature salience bias was not found. In Experiment 4, there was no tendency for subjects to overestimate the percentage of instruction-mentioned features in randomly distributed displays. If instruction-mentioned features are generally more salient than are other features, then subjects should be biased to overestimate percentages for all display types. In Experiment 5, removing black squares did not generally decrease estimates of black squares any more than removing white squares increased estimates.

Relation to Confirmation Bias

The results are more compatible with a "failure to search for negative evidence" than with a "discounting of available negative evidence" interpretation of confirmation biases. If negative evidence is discounted in general, then we would expect features not mentioned in instructions to be generally less influential than instruction-mentioned features. This appears not to be the case. If instructed to estimate the percentage of squares that are black, white squares are sometimes more influential than black squares depending on where the squares are positioned. White squares have a large impact on estimates if they are located in regions with many black squares. As such, evidence against the hypothesis that there is a high percentage of black squares is not ignored or discounted when it is available. Instead, the current confirmation bias appears to be best described as a tendency to attend selectively to stimulus areas that have a relatively large amount of confirming evidence.

The current experiments cannot distinguish this confirmation bias from a matching bias; both biases could account for our results. A matching bias occurs if there is a tendency to select only those stimuli with properties that match the properties mentioned in the rule. An amended version of this bias ("tendency to select only those stimuli that occur in regions with many properties that match the properties mentioned in the rule") accounts for our results. To distinguish the matching bias from the confirmation bias, it is necessary to run a version of the experiment in which subjects are probed, for example, to "estimate the percentage of squares that are not black." According to the matching bias, subjects would attend selectively to regions with many black squares, whereas according to a strategy of hypothesis confirmation, subjects would attend selectively to regions with many white squares. In pilot experiments with these instructions, subjects most often reported "flipping the instructions around to remove the negative" during the experiment, yielding results that are consistent with the hypothesis confirmation strategy. However, this tendency may be an artifact of the fact that subjects are required to make many estimations, and the form of the question remains constant across trials.

The results obtained in the current experiments are analogous to those found by Mynatt et al. (1977): "When confronted with unambiguous falsifying evidence, [subjects] utilized it in precisely the correct way—by rejecting their incorrect hypotheses... Our subjects did not, on the other hand, appear to look for and test alternative hypotheses" (p. 324). Similarly, our subjects did not attend to regions that contained high concentrations of negative evidence, but they were strongly influenced by negative evidence that was located in regions with a large amount of positive evidence.

Experiment 4 allows further refinement of the theory that subjects search selectively for confirmatory data. According to one theory of the differential search for confirmatory over disconfirmatory data, it is relatively easy to selectively sample instruction-mentioned features in clustered displays because they are separated from the other features. By this theory, the greater overestimation of clustered than distributed displays is due to an increased accessibility of probed features. By the second theory, the greater overestimation is caused by a failure to ignore nonprobed features that are in the same region as probed features. The two theories respectively propose that (a) confirming evidence is highlighted when clustered and is hard to uncover when distributed, and (b) disconfirming evidence is highlighted when close to confirming evidence and is ignored when far from confirming evidence. Both processes are implicated in the observed confirmation bias. When subjects are instructed to

estimate black squares, black squares (as expected by the first theory just posited) and white squares (as expected by the second theory) influence estimates more when they come from black rather than white regions.

It might seem bizarre that a system would be biased not to look for certain kinds of information but would appropriately use the information once it has been detected. However, this scheme has the advantage of allocating attention to those regions that are explicitly mentioned in the instructions and yet assimilating information that is not explicitly mentioned. Allocating attention to instruction-mentioned regions is generally a reasonable heuristic when explicitly mentioned items are likely to be important for the task at hand. In the current research, all of the elements are equally important for a percentage estimation, but across different estimation tasks, instruction-mentioned features are more likely to be important. In particular, if subjects are asked to estimate the number rather than the percentage of a type of element, then only items with the instruction-mentioned feature need to be attended. Other researchers made similar claims as to the adaptive value of the confirmation bias (Klayman & Ha, 1987). This heuristic runs afoul only when the probability of feature x being in a region given that feature y is in the region is not equal to the unconditional probability of feature x being in the region. When feature locations are independent (nonclustered), the heuristic of locating instruction-mentioned features and observing the featural state of neighboring objects provides estimates with very little bias.[3] The problem of biased estimates for displays with nonindependent feature locations may be more than compensated by the advantage gained by only requiring a subset of a display's information to be registered. In particular, in cases in which there are many relevant features and relatively few instruction-mentioned features, the strategy of focusing on areas with instruction-mentioned features may greatly facilitate a difficult judgment task. Widespread departures of subjective estimates from true percentages would occur if negative evidence were completely dismissed. Much greater subjective or objective agreement on percentages can be obtained, without unreasonable processing requirements, by assuming that regions are selectively attended to on the basis of their immediate pertinence to the task.

Attention and Framing

The current explanation for the systematic overestimation of clustered displays has implicitly used a notion of attention. In Experiment 4, individual white and black squares were selectively removed from white, black, and neutral regions. Estimates from these displays were compared with those obtained when no squares were removed. By this comparison, quantitative estimates are obtained about how influential individual squares are on subjects' estimates. It is assumed that the influence of a particular square on estimates is directly related to the attention given to it.

To explain the obtained results, both a framing effect and a regional attention process are invoked. The framing effect is needed because of the influence of instructions on feature salience. The regional attention process is needed because it is not always instruction-mentioned features that are the most salient. Rather, features that occur in regions that have a high concentration of features that are mentioned in the instructions are the most salient. Much of the research in attention is consistent with the constraints of regional attention. "Spotlight" (Eriksen & Murphy, 1987; LaBerge, 1983) and zoom lens (Eriksen & Yeh, 1985) metaphors of attention are similar in hypothesizing a "beam" of attention that is spread across a contiguous spatial region. LaBerge and Brown's (1989) gradient model also assumes that attentional resources are allocated to spatially contiguous regions. That is, these models and their supporting empirical data suggest that it difficult or impossible to allocate attention to two disconnected spatial regions without attending to the intervening region (but also see Egly & Homa, 1984). If attention must be directed to a display in the same manner that a spotlight or zoom lens is directed, then it would be difficult to attend to all randomly distributed features of a certain type and only to those features. However, if a cluster of like-valued features occupies a particular spatial region, then a "beam" of attention could selectively pick out those features. In short, research on the constraints of spatially directed attention provides a converging basis for the proposed regional attention claim.

There is also some evidence for spatially directed attention playing a role in the numeric estimation of display elements. Van Oeffelen & Vos (1982, 1984) recorded eye fixations from subjects who were asked to count the number of dots in a display. They observed a tendency for subjects to divide the display into clusters of dots that are defined by spatial proximity. Eye fixations are generally directed toward clusters rather than individual dots. A single eye fixation typically suffices to process up to five dots if they are in a spatially contiguous cluster.

The link between traditional research on visual attention and the importance of an item during estimation is suggestive but needs further exploration. The current experiments do not measure eye fixations, and attention is not explicitly manipulated by precuing subjects to certain areas of a display. Still, a functional notion of attention is required to accommodate the finding that different areas of a display have different influences on estimates and that the areas are defined by the instructions and by the configuration of elements within a display.

The current results are also consistent with a body of evidence suggesting that concentrated information influences judgments more than distributed information. The current claim has been made that attention is selectively drawn to regions with many instruction-mentioned features. Goldstone, Medin, and Gentner (1991) found that features of one type influence the similarity of stimuli more when many other relevant features of the same type are present (see also

[3] However, even with this heuristic, there will be a slight bias to overestimate the percentage of instruction-mentioned displays. If feature locations are independent, displays are composed of two equally probable features, and attention is given to an instruction-mentioned feature and its N neighbors, then the estimate will be based on $(N/2) + 1$ instruction-mentioned features and $N/2$ features not mentioned in the instructions.

Shepard, 1964; Tversky & Gati, 1982). Likewise, Whitlow and Skaar (1979) found that letters that occur frequently in a few strings tend to be judged as more frequent than those that occur more total times but with fewer concentrated appearances. More generally, Einhorn (1970) argued for subjects' use of several decision strategies, one being the "disjunctive model [by which] a person is judged on his best ability regardless of his other attributes" (p. 223). All of these examples instantiate the general trend for information sources that point in the same direction to have a greater impact if the information is clustered (spatially, temporally, or dimensionally) than if the information is distributed.

If the reasonable hypothesis is made that the influence of a feature on estimates is directly related to the attention given to the feature, then the current experiments provide a methodology for investigating the allocation of attention. One potential advantage of the method is that the displays are presented under more natural conditions than in many attention experiments. Many attention experiments present stimuli for a very brief time, under degraded conditions, or with precuing of certain spatial locations. These valuable techniques address many questions that could not be answered otherwise. However, there also is a demand for techniques that study the allocation of attention in judgments under clear, extended viewing conditions. The methods introduced here add to the existing list of such techniques.

Relation to Numerical Estimation Phenomena

Researchers (Alam, Luccio, & Vardabasso, 1986; Ginsburg, 1976; Ginsburg & Goldstein, 1987) have found that regular, evenly spaced distributions of black dots are perceived as more numerous than displays with nonhomogeneous distributions (clusters) of black dots. In a similar phenomenon, Frith and Frith (1972) found that one large cluster of black dots appears to contain more elements than several small clusters. These results might seem to contradict our results. In our experiments, the clustered displays are overestimated relative to the distributed displays.

Probably the most important difference between the experiments is that our subjects are given instructions to estimate the percentage of only one type of element in displays with multiple element types. This method leads subjects to sample disproportionately from regions with many elements of the type mentioned in the instructions. In the previous work suggesting greater overestimation of distributed relative to clustered displays, the instructions do not selectively heighten the importance of one type of display element relative to another.

Although previous research has investigated the influence of element distribution on estimation tasks, the current research indicates an interaction between element distribution and task instructions. As Experiment 5 reveals, clustering items in a display is not sufficient for overestimation. Overestimation occurs only if the items that are clustered are the same items that the subject is biased to sample. Subjects have an instructional predisposition to look for a particular type of item, and if the stimulus configuration facilitates the biased search, then overestimation occurs. Both the instructional and environmental forces must be present to obtain the observed overestimation.

Perceptual and Conceptual Biases: Analogs or Instantiations?

Because of the many differences between traditionally discussed confirmation biases and the currently reported effects, only the conservative claim that the effects are analogous to one another has been proposed. However, an interesting possibility is that both biases are instantiations of a general type of bias. Under this second interpretation, a common underlying account could be given to both biases. Some of the explanations proposed for conceptual confirmation biases can be rejected as too restricted to fit the role of a general perceptual or conceptual bias. Explanations that involve explicit hypothesis-testing strategies, high-level reasoning concerning possible causes of an event, motivation to have a rule verified or to find the correct rule quickly, and imposed causal schemas are of severely limited applicability to the percentage estimation task. In the tasks currently described, there are no explicit hypotheses to be confirmed or rejected, and little reason to expect that high-percentage estimates are more motivating than lower estimates. Instead, explanations of conceptual confirmation biases that refer to a biased search of confirmatory information (Baron, 1985) are general enough to encompass the current perceptual biases.

Recent experiments by Gary Levine, Steven Sherman, and myself provide some favorable evidence for the thesis that the biases described here generalize beyond the visual estimation task. Similar biases exist when subjects make memory-based estimations. For example, four different groups of subjects are asked the following: "How many states are landlocked?", "How many states have a coast?", "How many states have seven or more letters in their name?", and "How many states have fewer than seven letters in their name?" We find that the sum of estimates for the first two questions averages to more than 50, the correct sum. Abstractly, this corresponds to a clustered display. In one's mental image of the United States, coastal states occur in clusters, as do landlocked states. This overestimation is not found when the last two questions are summed and compared with 50. This corresponds to a distributed display. For most people, countries are not ordered by the number of letters in their name. It is too early to claim that traditional confirmation bias effects and the current perceptual estimation bias have the same cause, but at least there is some evidence that the reported confirmation bias or framing effect extends to tasks that do not involve the estimation of perceptually presented materials.

The simplicity of the percentage estimation tasks is useful for constraining general accounts of biased judgment. The tasks suggest that a biased search or weighting of evidence occurs even when the relevant information is perceptually available to subjects (Experiment 1), when subjects have little motivation to produce biased estimates, and when explicit hypothesis formation is not required. Different cases of confirmation bias may have different causes (Fischhoff & Beyth-Marom, 1983); still, there are grounds for optimism that per-

ceptual and attentional factors may underlie a potentially large set of biases.

References

Alam, S., Luccio, R., & Vardabasso, F. (1986). Regularity, exposure time and perception of numerosity. *Perceptual & Motor Skills, 63,* 883–888.

Baron, J. (1985). *Rationality and intelligence.* Cambridge, England: Cambridge University Press.

Baron, J., Beattie, J., & Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning: II. Congruence, information, and certainty. *Organizational Behavior and Human Decision Processes, 42,* 88–110.

Beattie, J., & Baron, J. (1988). Confirmation and matching biases in hypothesis testing. *Quarterly Journal of Experimental Psychology, 43,* 269–297.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral & Brain Science, 4,* 317–370.

Devine, P. G., Hirt, E. R., & Gehrke, E. M. (1990). Diagnostic and confirmation strategies in trait hypothesis testing. *Journal of Personality and Social Psychology, 58,* 952–963.

Duncan, J. (1983). Perceptual selection based on alphanumeric class: Evidence from partial reports. *Perception & Psychophysics, 33,* 533–547.

Egly, R., & Homa, D. (1984). Sensitization of the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 778–793.

Einhorn, H. J. (1970). The use of nonlinear, noncompensatory models in decision making. *Psychological Review, 3,* 221–230.

Eriksen, C. W., & Murphy, T. D. (1987). Movement of attentional focus across the visual field: A critical look at the evidence. *Perception & Psychophysics, 14,* 255–160.

Eriksen, C. W., & Yeh, Y.-Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 583–597.

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review, 90,* 239–260.

Frith, C. D., & Frith, U. (1972). The solitaire illusion: An illusion of numerosity. *Perception & Psychophysics, 11,* 409–410.

Ginsburg, N. (1976). Effect of item arrangement on perceived numerosity: Randomness vs regularity. *Perceptual & Motor Skill, 43,* 663–668.

Ginsburg, N., & Goldstein, S. R. (1987). Measurement of visual cluster. *American Journal of Psychology, 100,* 193–203.

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the nonindependence of features in similarity judgments. *Cognitive Psychology, 23,* 222–262.

Hogarth, R. M. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin, 90,* 197–217.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General, 118,* 126–135.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis-testing. *Psychological Review, 94,* 211–228.

LaBerge, D. (1983). Spatial extent of attention to letters in words. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 371–379.

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review, 96,* 101–124.

Levin, I. P., Schnittjer, S. K., & Thee, S. L. (1988). Information framing effects in social and personal decisions. *Journal of Experimental Social Psychology, 24,* 520–529.

Lord, C., Ross, L., & Lepper, M. (1979). Biased assimilation and attitude polarization: The effect of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37,* 2098–2109.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. In P. N. Johnson-Laird & P. C. Wason (Eds.), *Thinking* (pp. 315–325). Cambridge, England: Cambridge University Press.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Shaklee, H., & Mims, M. (1982). Sources of error in judging event covariations: Effects of memory demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 208–224.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology, 1,* 54–87.

Slowiaczek, L. M., Klayman, J., Sherman, S. J., & Skov, R. B. (1992). Information selection and use in hypothesis testing: What is a good question, and what is a good answer? *Memory & Cognition, 20,* 392–405.

Snyder, M., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology, 36,* 1202–1212.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Tversky, A., & Gati, I. (1982). Similarity, separability, and the triangle inequality. *Psychological Review, 89,* 123–154.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211,* 453–458.

Tweney, R. D., & Doherty, M. E. (1983). Rationality and the psychology of inference. *Synthese, 57,* 139–161.

Tweney, R. D., Doherty, M. E., & Mynatt, C. R. (1981). Introduction. In R. D. Tweney, M. E. Doherty, & C. R. Mynatt (Eds.), *On scientific thinking* (pp. 115–128). New York: Columbia University Press.

Van Oeffelen, M. P., & Vos, P. G. (1982). Configurational effects on the enumeration of dots: Counting by groups. *Memory & Cognition, 10,* 396–404.

Van Oeffelen, M. P., & Vos, P. G. (1984). Enumeration of dots: An eye movement analysis. *Memory & Cognition, 12,* 607–612.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12,* 129–140.

Wason, P. C. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology.* Harmonsworth, Middlesex, England: Penguin Books.

Wason, P. C., & Evans, J. St. B. T. (1975). Dual processes in reasoning? *Cognition, 3,* 141–154.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content.* London: Batsford.

Whitlow, J. W., Jr., & Skaar, E. (1979). The role of numerosity in judgments of overall frequency. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 409–421.

Received December 23, 1991
Revision received September 16, 1992
Accepted September 17, 1992 ■