

## RESEARCH REPORTS

# Who Should Stand Next to the Suspect? Problems in the Assessment of Lineup Fairness

Richard Gonzalez and John Davis  
University of Washington

Phoebe C. Ellsworth  
University of Michigan

A common procedure for assessing the fairness of a lineup is to give a verbal description of the perpetrator to people who did not witness the incident and ask them to select the likely perpetrator from the lineup. If people who never saw the perpetrator nonetheless make the "right choice" significantly more often than chance, the implication is that the lineup is unfairly suggestive. Little is known, however, about the factors that might bias this mock witness procedure. Two such biasing factors were examined in this study: the arrangement of photos in the lineup and the diagnosticity of the description. The results suggest that placing the target between 2 low-similarity foils increased the likelihood that he would be chosen, but only when the verbal description contained few diagnostic features. Implications for applied researchers and the construction of lineups are discussed.

The composition of the line-up—how many persons are in it, what they look like, what they wear, who they are—is a matter of great importance, for unless it is completely free from suggestive influences, its value will be greatly diminished, if not destroyed. (Wall, 1965, p. 52)

A suggestive lineup occurred in the case of a man accused of robbing a Western Union office in California (*Foster v. California*, 1968). Foster, who was almost 6 ft (1.83 m) tall, was placed in a three-man lineup with two men who were 6 in. (15.24 cm) shorter. What is it about this three-man lineup that makes it seem unfair relative to the commonsense view of what a lineup should be? The two shorter men created a background that made Foster "pop out" in comparison to the two foils because of the height discrepancy. If the witness described the perpetrator as "tall," then he or she can simply decide that Foster is the best match to the feature tall. Thus, the ar-

range of the lineup and the verbal description of the perpetrator are important to consider when constructing lineups.

The Foster lineup is not an isolated incident. Wall (1965) provided numerous examples of egregiously biased lineups: a defendant being the only Asian in a lineup, a dark-haired person surrounded by light-haired foils, and a suspect in his twenties surrounded by men over 40 years old.

In this article we address two questions related to the construction of lineups and the assessment of lineup fairness: (a) Does the arrangement of the lineup members influence lineup fairness? and (b) Does the nature of the verbal description influence the fairness of the lineup? These two questions have not systematically been examined in the literature, and their answers may have implications for standard procedures used by researchers such as the mock witness procedure.

These questions are also relevant to the construction of actual police lineups. We recently conducted a pilot survey of 44 detectives in Washington and Michigan (Davis, Gonzalez, & Ellsworth, 1993). The survey contained several questions about lineup procedures and construction. Detectives indicated that when conducting photo lineups they match the foils to the suspect on characteristics such as hair color, facial hair, skin tone, and eye color. Moreover, when constructing photo lineups, some detectives (13 out of 44) indicated that they check for contextual fairness (e.g., background, color vs. black and white, distance from subject, age of the photograph, and

---

Richard Gonzalez and John Davis, Department of Psychology, University of Washington; Phoebe C. Ellsworth, Department of Psychology, University of Michigan.

This research was supported by an intramural grant from the Graduate School Research Fund at the University of Washington. We thank Stephanie Bennett and Erin Moline for their help in conducting the experiment. Elizabeth Loftus and Maryanne Garry provided helpful comments and suggestions on a previous version of this article.

Correspondence concerning this article should be addressed to Richard Gonzalez, Department of Psychology, NI-25, University of Washington, Seattle, Washington 98195. Electronic mail may be sent via Internet to gonzo@u.washington.edu.

so forth) when selecting photographs to fill out the lineup. One detective even posed the proposition that we examined in the present study: "I try to get them as close as I can. I also place the filler photos that look most like the suspect around the suspect." In general, however, the responses indicated that while the detectives were obviously very concerned with making lineups as fair as possible, they used no formal methods for assuring fairness. The officers sometimes relied on the intuitions of other detectives and defense attorneys to ensure that a lineup was fair.

### Assessment of Lineup Fairness

One measure of lineup fairness was proposed by Doob and Kirshenbaum (1973). They defined a fair lineup as one in which all lineup members match the description of the suspect. They argued that the fairness of a lineup can be tested by presenting the lineup to mock witnesses—people who did not observe the crime and so never saw the actual perpetrator. The mock witnesses were given a verbal description of the suspect and asked to select the lineup member who best matched the description. A fair lineup should yield a uniform distribution of choices across the lineup; that is, each lineup member should be selected by an equal proportion of mock witnesses. The statistical test used by Doob and Kirshenbaum (a chi-square test) involves comparing the observed proportion of mock witnesses who select each lineup member with the theoretical distribution of  $1/k$ , where  $k$  is the number of people in the lineup. We call this the *uniformity test*.

Wells, Leippe, and Ostrom (1979) proposed the *functional size* of a lineup as a measure of fairness. The functional size is the reciprocal of the proportion of mock witnesses who pick the suspect. For example, if 10 out of 50 mock witnesses pick the suspect, then the functional size of the lineup is  $1/0.2 = 5$ . The main difference between the functional size measure and the uniformity test proposed by Doob and Kirshenbaum (1973) is that the latter compares the observed proportion to a theoretical lineup in which every member has an equal likelihood of being selected. The functional size measure, on the other hand, is simply a descriptive statistic focusing on the target (i.e., the reciprocal of the proportion of mock witnesses who select the target).

One drawback of the functional size measure is that it is not sensitive to the distribution of responses. Two six-person lineups may have the same functional size yet produce different distributions of choices by mock witnesses. Compare one lineup in which one sixth of the mock witnesses select each lineup member to another lineup in which one sixth of the mock witnesses select the suspect, and the remaining mock witnesses (five sixths)

all select the same foil. Both yield the identical functional size of 6, but clearly the first is a "fairer" lineup than the second because the responses are more evenly distributed.

An attempt to improve on the functional size measure was made by Malpass (1981) in his *effective size* measure. Effective size is related to the uniformity test used by Doob and Kirshenbaum (1973). Two major differences are that the effective size measure (a) ignores the foils who are never chosen and (b) takes the absolute difference between observed and expected proportions rather than the squared difference. (For a more detailed discussion of effective size and a computational formula see Malpass, 1981). The two hypothetical lineups mentioned above have different effective sizes: When the mock witnesses are equally distributed across the lineup, the effective size equals 6, but when one sixth of the mock witnesses select the suspect and the remaining mock witnesses select the same foil, the effective size is 1.33. Thus, at least on the criterion of sensitivity to the distribution of mock witness responses, effective size is preferred to functional size.

For different definitions of lineup fairness and the distinction between lineup size and lineup bias see Malpass and Devine (1983). Brigham, Ready, and Spier (1990) provided empirical tests of various measures, comparing the measures' reliability, sensitivity, and discriminability across several lineups. Our concern here, however, is not with analytic techniques but with the stimulus presented to the witness.

### Arrangement of the Foils

The research on the assessment of lineup fairness contains an implicit assumption that the similarity of the foils to the suspect is what defines the fairness of the lineup (Luus & Wells, 1991). However, the perception of similarity, like many kinds of psychological phenomena, is influenced by context. Processes such as contrast and assimilation may occur in a lineup—the foils placed next to the suspect could create a local context that influences the perception of the suspect. Thus, the similarity between Person A and Person B may differ according to the relative similarity and positioning of the other people in the lineup.

Consider 6-ft (1.83-m) Foster placed in a lineup with four foils: two foils are 6 ft tall and two foils are much shorter. A witness who encoded the perpetrator as tall might choose differently when Foster is placed between the two tall foils than when Foster is placed between the two short foils. Even though the members of the two lineups are identical and Foster may occupy the same position in both lineups, the arrangement of the foils around him may make a difference. Foster's height might be

more noticeable (i.e., Foster might appear to pop out because of a contrast effect) when he is flanked by the two short foils. This effect is analogous to the pop-out phenomenon described by Treisman (see Treisman, 1982; Treisman & Gormican, 1988).

In attempting to create lineups differing in functional size for a previous study (Gonzalez, Ellsworth, & Pembroke, 1993), we observed with considerable frustration that the arrangement of the foils made a substantial difference. In one lineup, the functional size increased from 2.5 to 6.5 when the lineup array was rearranged (62 mock witnesses in each lineup). In other words, different arrangements of the same six people led to different conclusions about the fairness of the lineup. The verbal description given to these mock witnesses was "dark hair, brown eyes, and chubby." When the target was surrounded by two thinner foils, the functional size of the lineup was 2.5; when the target was surrounded by two chubby foils, the functional size was 6.5. We suggest that the feature "chubby" became more salient when the target was surrounded by the two thin foils.

### Verbal Description of the Suspect

In the standard procedure for assessing the fairness of a lineup, mock witnesses are given a description of the suspect and are asked to select the lineup member who most resembles the description. The verbal description becomes part of the context of the mock witness procedure (e.g., the description may make some features more salient than others). To our knowledge, the question of how the content of the verbal description (e.g., the number of features, the ratio of diagnostic to nondiagnostic features, and so forth) influences the choices of mock witnesses has not systematically been explored.

The diagnosticity of a verbal description refers to how easy it is to select a single person from a lineup given only the verbal description. The verbal description "brown hair" can be diagnostic if one brown-haired lineup member is surrounded by five blondes or nondiagnostic if the lineup is composed of six people with brown hair. How would the functional size of a lineup differ if witnesses received a relatively nondiagnostic description as opposed to a more diagnostic description? Does the diagnosticity of the description influence the search process and the selection strategy used by the witness? All other things being equal, the more specific the description, the more difficult it is to construct a fair lineup because foils who match the suspect on more features have to be located. On the other hand, when the description is relatively nondiagnostic it becomes difficult for mock witnesses to find the best match, so one would expect the lineup to have a greater functional size.

A diagnostic description might reduce the arrange-

ment effect described above because mock witnesses have more features on which to base their judgment (cf. the dilution effect discussed by Nisbett, Zuckier, & Lemley, 1981). With more features available, it becomes less likely that adjacent foils provide the sort of contrast effect described earlier because foils might be good matches on some features but poor matches on others. An analogous reduction in the magnitude of the contrast effect with increasing number of features has recently been documented in the domain of choice (Simonson & Tversky, 1992).

### Method

In this study we examined context effects in the standard mock witness procedure for assessing lineup fairness. Mock witnesses observed five lineups and, for each lineup, were given a description of the "perpetrator." Context was manipulated by systematically rearranging the members in the lineup array. We compared two conditions: the foils adjacent to the target were either relatively high or relatively low in similarity to the target. One hypothesis was that in the low-similarity condition the target becomes more salient and will appear to pop out in comparison to the adjacent foils. For example, consider the case when the perpetrator is described as "chubby," three foils are also chubby, and two foils are thinner. Surrounding the chubby target by the two thinner foils makes the chubby feature stand out compared with when the target is surrounded by two chubby foils so that mock witnesses might be more likely to select the chubby target when he is surrounded by the two thin foils.

The diagnosticity of the descriptions was also varied. Some participants were given descriptions that consisted of two or three highly diagnostic features distinguishing the target lineup member from the foils and several features low in diagnosticity for the given lineup (i.e., features that fit all lineup members); other participants were given descriptions that consisted only of the low-diagnostic features that applied to all lineup members. The second hypothesis was that when the verbal description is high in diagnosticity mock witnesses will be less susceptible to the arrangement effect. A relatively high diagnostic description makes the process of finding a best match easier because the search is begun with a list of identifying features that give the mock witness more ways to eliminate from consideration lineup members who do not fit the detailed description. In the case of a description with low diagnosticity, mock witnesses will have a more difficult time finding the best match and might be more influenced by subtle, contextual cues such as arrangement because the description fits all lineup members and does not help reduce the search set.

### Participants

There were 279 participants in this study. They were compensated with extra credit in an introductory psychology course.

### Materials

Over 100 photographs of potential lineup members were taken at a local shopping mall and on the University of Washington campus. The photographs were of White men ranging in age from 18 to 25 years.

Five lineups were constructed from the photographs. Each lineup contained six photographs. One photograph was selected as the *target*, the others we called *foils*. Two arrangements of each set of six photographs were created by varying the similarity of the two men who were adjacent to the target. In one lineup the target was flanked by the two foils who looked most like him, and in the other lineup by the two foils who looked least like him. Each photo lineup was presented in a  $2 \times 3$  array on a poster board. Photographs were positioned approximately 1 in. apart. This  $2 \times 3$  arrangement was modeled after the photo-lineup procedure of a Seattle area police department.

Twenty independent raters provided a test of whether the lineups created by the experimenters had the intended characteristics (i.e., the target appeared more salient in one arrangement than the other). For each lineup the raters judged the similarity of each foil to the target on a 21-point scale anchored at *not at all similar* and *extremely similar*. The raters were given a picture of the target and photographs of the five foils. The order of the photographs was randomized for each rater. Data from this pairwise procedure (foil-target pairs) verified our own impressions of the lineups. In all five arrangements hypothesized to make the target stand out, at least one of the foils adjacent to the target received the least median similarity rating of the five foils in the lineup. In all five arrangements hypothesized to make the target less salient, at least one of the foils adjacent to the target received the greatest median similarity rating of the five foils. These data verify that the arrangements varied the similarity of the foils adjacent to the target.

The same raters who made these similarity ratings also provided verbal descriptions of the targets. Half of the raters saw the target for approximately 5–7 s and were then asked to write a description of the target after the photograph had been removed. The other half of the raters were allowed to look at the photograph of the target while they composed the description. The descriptions obtained under the two procedures differed only in level of detail, but they were in agreement as to general features present in each target. These descriptions were used to compose a general description for each target. The general descriptions contained some features that matched all of the members of the lineup but that matched the target the best. We created two types of descriptions for each target, varying the number of diagnostic features that were included. For example, for one target the description with high diagnosticity contained the features "male, 18–25 years old, dark hair, dark eyes, no facial hair, short hair, chubby face, large build," and the less diagnostic description contained the features "male, 18–25 years old, dark hair, dark eyes, no facial hair." The latter description fit all six lineup members, but the three additional features in the former description fit the target the best. The complete list of features for the five lineups appears in Table 1.

### Procedure

Participants were told that this study examined how people match verbal descriptions to faces. They were given the descrip-

tion of the target and shown an array of photographs. Participants were asked to select from the array the photograph that best matched the verbal description; they were given as much time as needed to make their selection. Participants were individually tested or in groups of 2 or 3. When participants were tested in groups, one lineup array was presented to the entire group, and participants made choices independently. The same procedure was followed for all five photo lineups. The order of presentation for the five lineups was randomized for each session. Arrangement (low vs. high contrast) and diagnosticity (low vs. high) for each individual lineup were manipulated between subjects. Because participants observed five lineups, each participant saw at least one lineup from each of the four possible cells.

### Results

Our prediction was that the arrangement of the lineup array should influence judgment more when mock witnesses are given descriptions that are low in diagnosticity. We tested this hypothesis by examining the simple main effects within each level of the diagnosticity variable, that is, the  $(1, -1, 0, 0)$  and  $(0, 0, 1, -1)$  contrasts where, within each contrast, the first two weights correspond to the two low-diagnosticity conditions and the second two weights correspond to the two high-diagnosticity conditions. These contrasts are equivalent to testing the difference between two proportions in the context of logistic regression. Participants who admitted knowing someone in a particular lineup were eliminated from the analysis of that lineup. The median number of participants eliminated across the 20 cells ( $5$  [lineups]  $\times$   $4$  [conditions]) was one, with a maximum of 7 in one of the cells.

The predicted simple main effect, that a different arrangement influences mock witnesses' selection of the target when they are given a description of low diagnosticity, was supported in two of the five lineups. The proportions of mock witnesses selecting the target are shown in Table 2. In Lineup C when the target was surrounded by two similar foils, the percentage of mock witnesses selecting the target was 15%; however, when the foils were rearranged so that the target was surrounded by two dissimilar foils, the selection rate increased to 32%,  $Z = 2.38$ ,  $p = .017$ , for the  $(1, -1, 0, 0)$  contrast. In other words, the functional size in one arrangement was 3.1 but changed to 6.8 under a different arrangement of the same six photographs. The position of the target was held constant in the two arrangements. Lineup E showed a similar arrangement effect: In the arrangement in which the target was surrounded by the two similar foils, the percentage of mock witnesses who selected him was 16%; for the other arrangement, the percentage was 32%,  $Z = 2.16$ ,  $p = .03$ . In terms of functional size this represents an increase from 3.1 for the arrangement that placed the less similar foils next to the target to 6.2 for the arrange-

Table 1  
*Low- and High-Diagnostic Descriptions Given to Mock Witnesses*

Condition	Description
Lineup A	
Low diagnostic	White male, 20–25 years old, high forehead, no facial hair
High diagnostic	White male, 20–25 years old, <i>bushy eyebrows</i> , high forehead, no facial hair
Lineup B	
Low diagnostic	White male, 18–25 years old, dark hair, dark eyes, no facial hair
High diagnostic	White male, 18–25 years old, with dark, <i>short hair</i> , dark eyes, <i>chubby face</i> , <i>large build</i> , no facial hair
Lineup C	
Low diagnostic	White male, 20–25 years old, dark hair, dark eyes
High diagnostic	White male, 20–25 years old, dark hair, <i>angular jaw</i> , <i>oval face</i> , dark eyes
Lineup D	
Low diagnostic	White male, 18–25 years old, average build, short hair, high forehead
High diagnostic	White male, 18–25 years old, average build, short hair, <i>full lips</i> , high forehead
Lineup E	
Low diagnostic	Male, dark hair, dark eyes, 20–24 years old, average build
High diagnostic	Male, dark hair, dark eyes, <i>rounded chin</i> , <i>squarish face</i> , 20–24 years old, average build

Note. The features added in the high-diagnostic description are italicized.

ment that placed the similar foils next to the target. The differences in the remaining three lineups did not reach statistical significance.

In the high-diagnosticity condition, we found that none of the five comparisons between the two arrangements reached statistical significance, three  $Z$ s < 1 and two  $Z$ s < 1.35. Thus, it appears that for these five lineups the diagnostic description inoculates against the arrangement effect. Further, the interaction contrast (i.e., 1, -1, -1, 1) was statistically significant for Lineup C,  $Z = 2.13$ ,  $p = .03$ , but failed to reach statistical significance for the remaining four lineups ( $Z$ s < 0.84).

### Discussion

The mock witness procedure for assessing fairness is widely used in studies of lineup identifications, with

functional size as the most common measure of fairness. A high functional size is generally taken as an indicator that the investigator has succeeded in choosing foils who resemble the target. Our data indicate, however, that functional size is sensitive to other variables besides choice of foil. The functional size of a lineup can change when the same lineup is rearranged. This finding suggests that the outcome of a mock witness procedure for testing lineup fairness reflects not only who is in the lineup but how the lineup is arranged. In two lineups the functional size went from 3 to 6 after the foils were rearranged. The direction of the change was predicted by the similarity of the foils immediately adjacent to the target. This effect was moderated by the nature of the verbal description given to the mock witnesses: A detailed, diagnostic description appears to inoculate against the lability caused by arrangement. Thus, the description given to the mock

Table 2  
*Proportion of Mock Witnesses Selecting the Target as a Function of the Diagnosticity of the Description (Low vs. High) and the Similarity of the Adjacent Foils (Low vs. High)*

Lineup	Low diagnosticity				High diagnosticity			
	High-similar foil		Low-similar foil		High-similar foil		Low-similar foil	
	<i>P</i>	<i>n</i>	<i>P</i>	<i>n</i>	<i>P</i>	<i>n</i>	<i>P</i>	<i>n</i>
A	.03	69	.06	68	.07	68	.06	66
B	.12	69	.12	69	.78	69	.72	60
C	.15	68	.32	68	.42	72	.38	61
D	.67	63	.64	69	.59	71	.48	65
E	.16	68	.32	69	.11	61	.20	69

Note. *P* = proportion.

witnesses also affects the outcome of tests for assessing lineup bias.

The effective size measure (Malpass, 1981) did not fluctuate as much as the functional size measure for the lineups in this study. For example, in Lineup C the effective size for the arrangement when the two similar foils were next to the suspect was 3.9, and the effective size for the arrangement when the two less similar foils were next to the target was 4.4. Recall that the functional size for these two lineups was 6.8 and 3.1, respectively. So, if anything, the effective size went in the opposite direction. It is not clear, however, that this stability across different arrangements is an attractive feature of effective size because the target was more than twice as likely to be selected in the latter lineup (15% vs. 32%, respectively). The effective size measure seems to have missed a critical difference between these two lineups.

One possible criticism of these results is that the descriptions in the low-diagnostics condition were too vague. We agree that the low-diagnostics descriptions used in the study were vague, but they are comparable to the descriptions given to mock witnesses in other studies. For example, Malpass and Devine (1983) presented mock witnesses with a description that consisted of these features: male with dark eyes, medium build and height, and wavy brown hair of medium length. Doob and Kirshenbaum (1973) used the description "rather good looking." Overall, we were surprised by the lack of attention given to the verbal description in the literature—often empirical studies that report functional size do not provide the verbal description given to mock witnesses. Descriptions actual witnesses provide can also be general and of low diagnosticity. There are several examples in the legal literature of vague descriptions. For example, in *People v. Caruso* (1968), the witness to a robbery described the assailant as "big, with wavy hair and a dark complexion." Additional examples of vague descriptions appear in Loftus (1979).

One implication of these results for applied researchers centers around the practice of counterbalancing the order of the lineup array. In general, it seems intuitively "right" to counterbalance the order of the lineup members because researchers do not want "artifactual" order effects to contaminate results. The present data, however, suggest that counterbalancing the arrangement of lineups can raise problems. As we have shown, a different arrangement can produce a different functional size; in other words, rearrangement creates a new stimulus. The arrangement effect described here is an alternative explanation to an order effect observed by Sporer (1993) across his counterbalanced lineups.

One way to deal with order effects is to counterbalance the position of the target across studies (as was done in Gonzalez et al., 1993). For example, in a replication

study the target might be placed in a different position than in the original study. By replicating an effect with the target in different positions, the researcher establishes generality with respect to the position of the target.

We recommend that those responsible for constructing lineups (police and defense attorneys) consider carefully the arrangement of the lineup. Our data suggest that simply matching on a pairwise basis (either to the physical features of the suspect or to the description originally given by the witness) may lead to different perceptions of the lineup with differing levels of bias, depending on the arrangement and diagnosticity of the description. It has been our experience that people who construct lineups do so on a pairwise basis (e.g., looking for similar individuals in jail cells and among off-duty police officers on an individual-by-individual basis). While this is certainly an efficient strategy for finding lineup foils, careful attention should be paid to how the lineup members are arranged and any context effects that might occur. Further, when constructing an actual lineup it is important to keep in mind the description given by the witness(es). How diagnostic is the description, given the lineup at hand? Recall that diagnosticity of the description cannot be evaluated independent of a particular lineup—a description that is diagnostic in one lineup (has at least one feature that distinguishes the suspect) may not be diagnostic in another lineup (there are no features that help distinguish an individual).

Our study suggests that lineup fairness depends not just on who is in the lineup (in the sense of similarity) but on how the members are arranged and on how they are described. These results also suggest that merely assessing the similarity between the target and each foil separately is not sufficient to ensure a fair lineup because the perception of similarity depends, in part, on the context created by the arrangement of the lineup array and the description of the target. While a great deal of attention has been given to the choice of similar foils and some to the placement of the target in the array, almost no attention has been directed to the description of the target in the mock witness procedure. The present results show that the description can have a substantial influence on mock witnesses' selections.

## References

- Brigham, J. C., Ready, D. J., & Spier, S. A. (1990). Standards for evaluating the fairness of photographic lineups. *Basic and Applied Social Psychology, 11*, 149-163.
- Davis, J., Gonzalez, R., & Ellsworth, P. (1993). *A survey of police departments: Lineup and showup procedures*. Unpublished manuscript, University of Washington and University of Michigan.
- Doob, A. N., & Kirshenbaum, H. M. (1973). Bias in police

- lineups—Partial remembering. *Journal of Police Science and Administration*, 1, 287–293.
- Foster v. California, 394 U.S. 440 (1968).
- Gonzalez, R., Ellsworth, P., & Pembroke, M. (1993). Response biases in lineups and showups. *Journal of Personality and Social Psychology*, 64, 525–537.
- Loftus, E. F. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.
- Luus, C. A. E., & Wells, G. L. (1991). Eyewitness identification and the selection of distractors for lineups. *Law and Human Behavior*, 15, 43–57.
- Malpass, R. S. (1981). Effective size and defendant bias in eyewitness identification lineups. *Law and Human Behavior*, 5, 299–309.
- Malpass, R. S., & Devine, P. G. (1983). Measuring the fairness of eyewitness identification lineups. In S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating witness evidence* (pp. 81–102). New York: Wiley.
- Nisbett, R., Zuckier, H., & Lemley, R. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology*, 13, 248–277.
- People v. Caruso, 65 Cal. Rptr. 336 (1968).
- Simonson, I., & Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, 24, 281–295.
- Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology*, 78, 22–33.
- Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 194–214.
- Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95, 15–48.
- Wall, P. M. (1965). *Eyewitness identification in criminal cases*. Springfield, IL: Charles C Thomas.
- Wells, G. L., Leippe, M. R., & Ostrom, T. M. (1979). Guidelines for empirically assessing the fairness of a lineup. *Law and Human Behavior*, 3, 285–293.

Received February 22, 1994

Revision received June 10, 1994

Accepted September 9, 1994 ■