Response Biases in Lineups and Showups

Richard Gonzalez, Phoebe C. Ellsworth, and Maceo Pembroke

The *showup*, or presentation of a single suspect to an eyewitness, is widely believed to be a more biased and suggestive identification procedure than the lineup even though there has been no empirical work on this issue. Results suggest, however, that witnesses at a lineup are less likely to say "not there" than are witnesses at a showup. This tendency is seen in both live and photographic identification procedures, in both laboratory studies and real-world identifications. Showups in the lab resulted in no more mistaken identifications than lineups. Results also suggest that a showup is not equivalent to a poor lineup (i.e., a lineup with a functional size of 1). It is hypothesized that a showup leads to an absolute judgment, whereas a lineup leads to a relative judgment, and that the police pressures on witnesses are unlikely to be any greater for showup than for lineup identifications.

When a witness to a crime points to the suspect and says "That's the man" a crucial link in the prosecution's chain of evidence is forged. Police, prosecutors, and psychological researchers agree with the Supreme Court's conclusion that the pretrial identification of a suspect is a "critical stage" in the proceedings against the defendant, requiring particular scrutiny (*U.S. v. Wade*, 1967). They disagree sharply, however, about the value of different methods of identification.

Police officers commonly use the *showup*, a one-to-one confrontation between the witness and the suspect in which the witness is asked whether the suspect is the person who committed the crime. Psychologists have argued that the showup is overly suggestive (Levine & Tapp, 1973; Malpass & Devine, 1983; Sobel, 1985; Wells, Leippe, & Ostrom, 1979). A procedure believed to be less biased is the *lineup*, in which the witness is asked to select the perpetrator from a group. In a recent survey of psychological experts in the field of eyewitness testimony, 78% of the sample agreed that "the use of a one-person showup

instead of a full lineup increases the risk of misidentification," and 65% felt that the evidence for this proposition was generally reliable or very reliable (Kassin, Ellsworth, & Smith, 1989). The source of this high level of consensus among the experts is somewhat mysterious, however, as there is no empirical research that compares patterns of identification in lineups and showups.

Legal Arguments Regarding the Use of Showups

Police

The police are the strongest proponents of showups, and their argument is largely practical. The police are called to the scene of the crime, the witness describes the criminal, the police immediately initiate a search, and when they find someone who fits the description they bring the witness to the spot and ask him or her whether the suspect is the culprit. This sort of showup is often called an *in-field identification* or a *drive by*. The general rationale is that because of the short delay between the event and the showup, memory will be relatively intact, and so witnesses will be more accurate than they would if they saw the suspect in a lineup hours or days later.

This argument in favor of the showup has nothing to do with the basic task structure of lineups and showups, that is, the number of options given to the witness. Rather, proponents are relying on a variable that covaries with real-world showups: the immediacy of the identification procedure. Consider, for example, the showup in *US v. Wilson* (1970):

> Even though the confrontation is inherently suggestive because of the presentation of a single suspect . . . we think the case is not one of undue suggestiveness, in view of the countervailing considerations that prompt, on-the-scene identifications are likely to promote fairness, by enhancing reliability of the identifications, and permit expeditious release of innocent subjects. (pp. 404–405)

Presumably, if a set of five other people who resembled the suspect could be magically transported to the identification

site, police would have no objection to conducting an "in-field lineup."

Courts

The relative suggestiveness of lineups and showups was first considered by the United States Supreme Court in 1967 (*U.S. v. Wade*, 1967; *Gilbert v. California*, 1967; *Stovall v. Denno*, 1967). The Court recognized that there are substantial risks of bias in the conduct of both lineups and showups but considered the dangers of suggestion to be greater in showups (*Stovall v. Denno*, 1967, p. 302). Thus, the intuitions of the Supreme Court majority coincided with those of the psychologists.

Despite their concerns about the validity of showup identifications, the Court let Stovall's conviction stand, holding that the acceptability of any identification procedure must be evaluated in the context of the "totality of the circumstances surrounding it" (*Stovall v. Denno*, 1967, p. 302). In Stovall's case both the suggestiveness of the showup and the surrounding circumstances were extreme. The victim, Paul Behrendt, was stabbed to death, and the only witness, his wife, was so severely wounded that her survival was in doubt. Two days after the murder, the police brought Stovall, in handcuffs, to Mrs. Behrendt's hospital room and asked her whether he was the man. Quoting the Court of Appeals, the Court accepted the identification because the police had no choice in the circumstances:

> No one knew how long Mrs. Behrendt might live. Faced with the responsibility of identifying the attacker, with the need for immediate action and with the knowledge that Mrs. Behrendt could not visit the jail, the police followed the only feasible procedure and took Stovall to the hospital room. Under these circumstances, the usual police lineup, which Stovall now argues he should have had, was out of the question. (355 F.2d, p. 735)

Since 1967, the Court's definition of *totality of circumstances* has been transformed from one emphasizing circumstances affecting the necessity of the suggestive procedure to one emphasizing circumstances affecting the likely accuracy of the identification. In *Neil v. Biggers* (1972), a rape victim identified her attacker on the basis of a showup 7 months after the crime, and the Court let the conviction stand because she had spent "up to half an hour" with the man, she had been under great stress, she was very confident, and she had not identified anyone in previous lineups or showups. In *Manson v. Brathwaite* (1977), the Court reaffirmed this new view of the totality of circumstances, holding that a single-photograph showup was acceptable because the initial contact lasted for 2 min, the identification was made only 2 days later, and the witness was a trained police officer. Thus, although the Court has continued to describe showups as unnecessarily suggestive, and to regard the lineup as a superior procedure, it has rarely found a showup to be so impermissibly suggestive as to warrant reversal of a criminal conviction.[1]

Psychological Arguments Regarding the Use of Showups

Most researchers in the area of eyewitness testimony have been forceful in their criticism of the showup (e.g., Malpass & Devine, 1983; Wells et al., 1979; Yarmey, 1979). There are two quite different lines of criticism, which are often confused in the literature. It is important to keep them separate if scholars are to think clearly about the relative merits of lineups and showups. The first is a structural, formal problem concerning the nature of the two tasks. It involves problem solving and identification decision strategies and the sort of variables typically studied by cognitive and social–cognitive psychologists. The second is an interpersonal, social, even ethnographic set of considerations involving the state of mind of a real witness, the role of the police in creating situational pressures, the forensic uses of the two procedures, the witness's attributions of beliefs and motives to the police, and the consequences of various types of mistakes for the innocent suspect. These considerations are no less "psychological" than the structural considerations, but they suggest different research methods and different sources of data.

Most research on eyewitness identification has involved compromise designs intended to maintain control over the variables affecting the problem-solving strategies, while introducing various features designed to mimic real-world identification settings. Studies vary enormously in their emphasis on highly controlled abstract tasks, or more engrossing tasks high in experimental realism, or tasks and settings aimed at mirroring the natural context (high in "mundane realism," see Aronson, Ellsworth, Carlsmith, & Gonzales, 1990; Malpass & Devine, 1981a, 1981b). The highly controlled studies can easily be criticized for failing to capture the nature of the eyewitness's task as it actually occurs in the real world; the studies that attempt to bring in real-world elements can frequently be criticized for confounding conceptually distinct variables. These issues are not unique to research on eyewitness testimony; they are standard fare in many research methods books and an inevitable debating point in arguments over basic versus applied research. However, they are particularly troublesome in discussions of the response biases in showups and lineups because there are arguments about biases inherent in the abstract structure of the tasks—biases that would exist whether the stimuli were faces, sentences, shapes, or works of art—and there are also arguments about biases inherent in the concrete interpersonal task involving an eyewitness, a suspect, a police officer, and usually other people. The bottom line of both arguments has typically been that showups are far more suggestive than lineups, but the arguments themselves are quite distinct.

In this article, we describe three studies comparing lineups and showups—two experimental simulation studies and a field study of actual lineups and showups. In introducing the first two studies, we focus on the abstract formal properties of the lineup and showup tasks. In introducing our third study, designed to see whether the pattern of results obtained in the police department resembles that found in the laboratory, we discuss the additional sources of bias that may affect an actual eyewitness faced with a lineup or showup.

Formal Task Properties of Lineups and Showups

One structural argument against showups involves the psychometric properties of the one-person identification (e.g., Elli-

---

[1] For a lower court example where the totality of the circumstances test failed and the showup was found to be unreliable see *Smith v. Coiner*, 1973.

son & Buckhout, 1981; Malpass & Devine, 1983). Because the witness responds with a yes–no answer, the showup, with only one choice option, makes it difficult to catch guessers (or witnesses who simply want to have someone "put away"). The greater number of options in the lineup decreases the probability that the suspect may be selected by chance (i.e., $1/k$, where k is the number of choice options). In this sense, the showup is analogous to a true–false question and the lineup to a multiple-choice question.

Malpass and Devine argued (1983) that a "lineup is in principle more fair than a showup because it distributes the probability of identification of an innocent suspect across the lineup foils, reducing the risk of an identification error" (p. 85). Two types of error have been identified: misidentification and failure to identify the actual perpetrator (see Levine & Tapp, 1973; Malpass & Devine, 1981a, 1981b; Wells & Lindsay, 1980). These errors correspond to false positives and false negatives, respectively. The mistaken identification of a known foil helps to identify "witnesses who do not possess very good information about the identity of the offender but who feel they must attempt an identification" (Malpass & Devine, 1984, p. 87). In a showup, the only available option for such witnesses is to identify the suspect; the mistaken identification of a foil is impossible. The common assumption is that there is considerable situational pressure on the witness to make an identification.

A second structural argument is that the presence of foils who resemble the suspect causes the witness to exercise greater caution: The witness is less likely to choose a man because he generally resembles the suspect if he is surrounded by others who also generally resemble the suspect. Thus, it is generally believed that witnesses are more likely to make a positive identification—to respond yes—when the suspect is presented alone than when the same suspect is presented along with several foils, assuming that the foils plausibly fit the description of the suspect.

These arguments may be undercut, however, by a difference in the decision-making strategies elicited by the lineup and showup tasks. Lineups call for comparative, or relative, strategies because the witness selects from several alternatives. Showups elicit absolute strategies because the witness must decide if the suspect is or is not the perpetrator. This difference in strategy suggests the opposite prediction—a higher frequency of "no" responses in the showup.

Both lineups and showups are essentially tests of recognition memory. Many theories of recognition memory propose that people decide whether a test stimulus matches an original item on the basis of a judgment along a familiarity dimension, that is, whether it is an "old" or a "new" item (see, for example, Bower, 1967; Gillund & Shiffrin, 1984; Kintsch, 1967, 1970). We hypothesize that at a showup the witness attempts to find a match between the memory trace of the perpetrator and the person presented. If the match between the memory trace and the suspect meets or surpasses some criterion level then the witness responds yes, otherwise no. The showup, then, is a type of absolute judgment (cf. Blumenthal, 1977). This similarity judgment between the memory trace and a physical stimulus has been termed ecphoric similarity (Tulving, 1981).

The number of alternatives in the memory test, however, may influence decision making along the familiarity dimension. When there are several options the witness searches the lineup for the person who most resembles the memory trace of the perpetrator. "The witness often may treat the [lineup task] as one of identifying the individual who best matches the witness' recollection of the culprit even if that match is not perfect, rather than as one of identifying the true criminal" (Woocher, 1977, p. 986). Furthermore, once the witness has successfully accomplished the task (choosing the best match), he or she may lose sight of the fact that there is a second task: deciding whether the best match is good enough. Thus, the lineup is conducive to a comparative judgment (Wells, 1984; Woocher, 1977). We hypothesize that the process of finding the best match (comparative judgment) is more likely than an absolute judgment to result in a positive identification. Thus, we predict more yes responses to the lineup than to the showup. This analysis parallels the well-known finding in the perception literature that the threshold is lower in k-alternative forced-choice discriminations than in yes–no discriminations. The former procedure produces a higher frequency of yes responses (see Blackwell, 1953, for a review).

We conducted three studies to examine the widely believed but never-tested assumption that witnesses are more likely to make an identification in a showup than in a lineup. Given the surprising findings of the first study, we conducted a second study as a somewhat more rigorous replication, including a comparison of showups with lineups of different functional sizes. The third study was designed to discover whether a similar pattern of results would emerge in actual lineups and showups conducted by a local police department, where all sorts of real-world social pressures are added to the formal task properties.

Study 1

Overview

The purpose of the first study was to provide an initial empirical test of response differences in showup and lineup procedures. Following a staged purse-snatching incident in an undergraduate psychology and law course, subjects were shown either a showup or a lineup. The actual perpetrator was presented in half of the lineups and in half of the showups.

Method

Subjects

Sixty-three students in a psychology and law course at Stanford University observed a staged crime. After the incident the instructor informed the students that the crime was staged as part of an experiment on eyewitness identification. Three students declined to continue with the experiment, and 1 subject did not have time to complete it, leaving a total of 59 subjects.

Perpetrator and Lineup Construction

A group of seven African-American male Stanford undergraduates, matched for height, weight, age, color, and facial hair, were selected to pose in the identification procedure. One of them volunteered to be the perpetrator. He was in his early 20s, weighed 140 lb, and was 5 ft 9 in. tall. During the identification procedure all lineup and showup members wore the same clothes: Stanford basketball warmups, dark blue jeans, and white tennis shoes.

Procedure

Fifteen minutes after the beginning of the class the perpetrator entered the classroom and sat in the back row. After 1 min, he stood up and walked slowly down the stairs toward the lecturing instructor. On reaching the front of the classroom he stood in profile view to the class and looked at the instructor for a few seconds. The instructor asked "Can I help you?" Without answering, the perpetrator walked to the side of the instructor's desk and grabbed her purse; he turned around, looked toward the class, and then ran all the way across the front of the classroom and out an exit door. As he was going out of the door, a female student began to run after him—she was stopped by the instructor (this attempt to catch the thief was not part of the planned incident). The perpetrator was in the classroom for 79 s.

Immediately after the incident, the instructor informed the students that the crime was staged as part of a class demonstration on eyewitness identification and asked them not to discuss the incident with other students until the class was over.[2] Subjects then received a consent form and a questionnaire. The experimenter, another African-American male undergraduate, read the consent form to the class. Following completion of the consent form he said,

> You are about to take part in an identification procedure very similar to those used by police today. Please bear in mind that your decision may have an impact on future police identification procedures, so please take this task seriously.

Next, he asked subjects to fill out the pre-identification questionnaire. The questionnaire contained (a) open-ended questions asking subjects to describe the incident and the perpetrator; (b) questions asking subjects to identify specific characteristics of the perpetrator such as age, height, weight, and hair color; and (c) five true–false and three four-alternative forced-choice items (e.g., "Was the perpetrator wearing a jacket?").

Subjects then went to the identification task in randomly assigned groups of 8. A second experimenter (a European-American male) conducted the identification procedure. The instructions were as follows:

> You are being asked to look at a suspect [lineup]. The fact that this suspect [lineup] is being shown to you should not influence your judgment. You should not conclude or guess that this suspect [someone in the lineup] committed the crime. You are not obligated to identify anyone.

These instructions were taken verbatim from the standard forms used by a local police department.[3] Subjects read a printed copy of the instructions while the experimenter read them aloud. The experimenter asked subjects not to talk to other witnesses and not to indicate to others whether they made an identification.

Subjects were shown either a live lineup or a live showup. These two procedures were crossed with the presence (or absence) of the actual perpetrator, creating four conditions. In the *valid showup condition* subjects saw the actual perpetrator alone; in the *valid lineup condition* they saw the actual perpetrator along with five foils (six lineup members in all). In the *perpetrator-absent conditions,* a seventh man, matched in physical appearance to the other six, replaced the real perpetrator. In the *blank showup condition* he was presented alone; in the *blank lineup condition* he took the perpetrator's place in the line with the same five foils.

In the lineup conditions the six men stood in a straight line facing the witnesses. The numbers 1 (left) to 6 (right) were posted over their heads. The perpetrator or his substitute appeared in Position 2. The foils maintained the same serial position across all lineups.

Subjects were given three options: (a) They could make a positive identification of the perpetrator (in the lineup condition they also indicated the perpetrator's position in the lineup), (b) they could indicate that the perpetrator was not there, or (c) they could indicate that

they did not remember the perpetrator well enough to make a decision either way. We refer to the three options as *yes, no,* and *I don't know,* respectively. We allowed an I don't know option because previous research on lineups suggests that the inclusion of such an option might lower the false-alarm rate (e.g., Köhnken & Maass, 1988).

Results and Discussion

The study had a 2 (identification procedure) × 2 (presence or absence of the actual perpetrator) between-subjects design. There were no main effects or interactions for the subjects' (a) estimates of the duration of the incident (in seconds); (b) estimates of the perpetrator's age, height, or weight; or (c) proportion correct on the objective portion of the pre-identification questionnaire (all $Fs < 1.43$). Because these items were assessed before the identification procedure, we conclude that there were no initial group differences on these variables.

The primary purpose of this study was to examine differences in positive and negative responses across showup and lineup identification procedures. Thus, the dependent variable of interest was whether the subject said the perpetrator was there or not there, not whether the subject made an accurate identification. We computed the proportion of positive (yes) and negative (no) responses given that the subjects made a definite response (i.e., the conditional probabilities P[yes/yes or no] and P[no/yes or no]).[4] Limiting analyses to subjects who made a definite decision allows a simple, clear presentation of the results and a more rigorous test of the hypotheses (see Wickens, 1989, for a similar procedure using logits). This constraint is also forensically relevant because definite yes or no responses on the part of the witness have important consequences for the suspect. The merits of yes rates (or, equivalently, no rates) to assess response bias are reviewed by Macmillian and Creelman (1990). However, a direct application of standard response bias measures was not possible in this context because, among other things, each subject contributed only one data point and hit or false-alarm rates must be computed between subjects. Results on accuracy rates are presented after the response bias data.

Response Bias

The results presented in Table 1 show a striking tendency for subjects to respond no to the showup but yes to the lineup. Collapsing over valid and blank identification procedures, 4 of 22 subjects (18%) who made a definite response in the showup conditions responded yes, whereas 17 of 22 subjects (77%) in the

---

[2] As in most studies of eyewitness identification, subjects knew they were participating in a research study. There is some evidence that such knowledge may inflate the rate of choosing (Köhnken & Mass, 1988). To have affected the results of the present study, this inflation would have had to occur only in the lineup conditions. Whether this is a plausible alternative explanation is an empirical question.

[3] This was the same police department that was examined in Study 3, so the instructions were the same in the lab and field research reported here.

[4] There were no significant differences across the experimental conditions in the proportion of I don't know responses. Furthermore, computing proportions of responses over all three response categories (i.e., so that P[yes] + P[no] + P[I don't know] = 1) does not affect the pattern of statistical significance.

Table 1
Frequency of Choice as a Function of Identification Procedure for Study 1

Identification procedure	Choice (%)				Conditional proportions (%)	
	Yes	No	I don't know		P(y/y or n)	P(n/y or n)
Showup						
Valid	21	50	29		30	70
n	3	7	4			
Blank	7	73	20		8	92
n	1	11	3	Marginal	18	82
				n	4	18
Lineup						
Valid	64	0	36		100	0
n	9		5			
Blank	50	31	19		62	38
n	8	5	3	Marginal	77	23
				n	17	5

Note. y = yes; n = no.

lineup conditions responded yes ($p = .0002$ by Fisher's exact test).[5]

The same response bias was observed whether the actual perpetrator was present or not. When shown the actual perpetrator in a showup, only 3 of 10 subjects responded yes; the other 7 said he was not the right man. When shown the actual perpetrator in a lineup, all 9 of those who were sure enough to make a definite response said yes. This difference is significant at $p = .003$ by Fisher's exact test. Similarly, when confronted with the wrong man in a showup, only 1 of 12 subjects mistakenly responded yes, whereas 8 of 13 subjects (62%) mistakenly responded yes when presented with the blank lineup ($p = .011$ by Fisher's exact test).

Accuracy

How did accuracy differ in the lineup and showup conditions? First, consider the procedures that included the actual perpetrator. In the valid showup, 3 of 10 subjects (30%) correctly identified the perpetrator, whereas in the valid lineups, 6 of 9 subjects (67%) correctly chose the perpetrator. Thus, the lineup yielded a higher proportion of correct identifications than did the showup.

A complete assessment of accuracy, however, must also include an examination of correct "not there" answers. In the blank showup, 11 of 12 subjects (92%) correctly stated that the man was not the perpetrator. In the blank lineup, 5 of 13 subjects who made a choice (38%) correctly stated that the perpetrator was not there. Thus, the showup yielded a higher rate of correct rejections than the lineup. The differential pattern of correct identifications and correct rejections reflects the base-rate response pattern in the showup and in the lineup.

This response bias, in turn, results in different kinds of errors in lineup and showup identifications. That is, the tendency to respond no when faced with a showup produces a high miss rate (70% of the subjects who were sure enough to make a decision when faced with the actual perpetrator said he was not the man). Similarly, the bias to respond yes when faced with a lineup leads to a high false-alarm rate (62% of the subjects who were sure enough to make a decision incorrectly selected some-

one from the blank lineup). If these results apply to real-world identifications, the consequences for showups are fairly clear: Many criminals are not identified, but relatively few innocent people are misidentified. The consequences for lineups are much more complicated. In the blank lineup condition, 2 subjects selected the perpetrator substitute, and the remaining 6 were split, 4 and 2, between two of the foils. The person we arbitrarily selected as a substitute for the perpetrator was selected no more often than would be expected by chance. Had we happened to select the foil most often picked to be our perpetrator substitute, however, he would have had a 31% chance of being misidentified (or a 25% chance if we include the don't knows in the calculation).[6]

Did the presence of the actual perpetrator in the identification procedure have any effect on the subjects' responses? Obviously, there should have been a higher frequency of no responses in the perpetrator-absent procedures than in the perpetrator-present procedures. Unfortunately, the effect was not clear. Not one subject responded no to the valid lineup, but 5 of 13 subjects in the blank lineup (38%) responded no ($p = .054$ by Fisher's exact test). However, within the showup conditions the proportion responding no to the valid showup (70%) did not differ statistically from the proportion responding no to the blank showup (92%; $p = .293$ by Fisher's exact test), although the trend was in the right direction: higher rates of "no" responses in perpetrator-absent conditions.

Study 2

The results of Study 1 suggest that people respond differently to lineups and showups. In particular, people appear to be much more cautious about making an identification when confronted with one individual. Because these results fail to con-

---

[5] All significance levels reported in this article, including Fisher's exact test, are two-tailed.

[6] A reviewer argued that this different foil would also have been picked more often in a showup, most likely 25% of the time. However, because the real perpetrator was only picked 21% of the time in Study 1 and 0% of the time in Study 2, this seems unlikely.

firm the generally accepted notion that witnesses are *more* likely to falsely identify an innocent suspect in a showup than in a lineup, we carried out a second study designed to test the same hypothesis using a different crime, a different modality, and a different perpetrator. In particular, we felt that subjects in Study 1 might have been reluctant to accuse a lone African-American man for fear of appearing racist, so we wanted to see if the results would replicate with a European-American perpetrator.[7]

A second goal of Study 2 was to examine the effects of the similarity of the foils to the suspect, which has been an issue of concern in the construction of lineups (e.g., Doob & Kirshenbaum, 1973; Lindsay & Wells, 1985; Malpass & Devine, 1983, 1984; Wells et al., 1979). A lineup is considered fair if all lineup members fit the general description of the perpetrator equally well. In a fair lineup each lineup member should have an equal chance of being selected as the perpetrator by someone who did not witness the crime but is simply selecting on the basis of a verbal description of the suspect.

Researchers frequently use a mock-witness procedure to assess the fairness of a lineup (e.g., Doob & Kirshenbaum, 1973). Mock witnesses are given a brief description of the perpetrator (e.g., "brown hair, brown eyes, and chubby") and asked to select the lineup member who best fits the description. If all lineup members are selected at equal rates the lineup is considered fair. In the extreme, if only one lineup member is chosen by the mock witnesses the lineup is considered biased (i.e., the evidence suggests only one lineup member fits the description). A simple measure of lineup fairness, called *functional size,* is the reciprocal of the proportion of mock witnesses who select the suspect (see, for example, Wells et al., 1979). For example, if the nominal size of a lineup is six and 25% of the mock witnesses choose the suspect, the lineup is assigned a functional size of four (i.e., 1/0.25).

Some researchers have argued that a showup is the equivalent of a lineup with a functional size of one (Malpass & Devine, 1983; Wells et al., 1979). The argument rests on the assumption that if the foils in a lineup can easily be eliminated from consideration then that lineup is equivalent to presenting the suspect alone (cf. Wall, 1965). However, to the extent that a showup leads to an absolute judgment and the lineup leads to a relative judgment, the responses to the two procedures may differ.

To compare people's responses to showups with their responses to lineups of various functional sizes, we created three lineups: one in which the foils closely resembled the perpetrator (high similarity), one in which they bore a moderate resemblance to him (moderate similarity), and one in which everyone in the line had blond hair and blue eyes except for the brown-haired, brown-eyed perpetrator (low similarity). Half of the time the perpetrator was present in the lineup (valid lineup) and half of the time another brown-haired, brown-eyed man was substituted (blank lineup). Thus, there were six lineup conditions, forming a 2 (presence or absence of the perpetrator) × 3 (similarity of foils to the perpetrator) factorial design.

Identification rates in these six lineup conditions were compared with three showup conditions: (a) a *valid* showup condition in which the person shown was the actual perpetrator, (b) a *high similarity* showup in which the person shown closely resembled the perpetrator, and (c) a *moderate similarity* showup in which the person shown, though fitting the general descrip-

tion, did not closely resemble the perpetrator. Given the very low rate of positive showup identifications in Study 1 we decided that little would be added by running a low-similarity showup.

The design of the second study allowed us to test the claim that a showup procedure is equivalent to a lineup with a functional size of one. We believe that in viewing a lineup, people find the best match, and thus are making a comparative judgment rather than an absolute one. Hence, we predicted the lineup with a functional size of one would produce a higher frequency of positive identifications than the showup.

Method

Subjects

One hundred eighty subjects participated in this experiment. Approximately half the subjects received partial credit for their introductory psychology course, and the other half were paid $4 for their participation. The paid subjects were recruited from several junior- and senior-level dormitories at Stanford University. Paid and nonpaid subjects were equally distributed across conditions.

Perpetrator and Lineup Construction

Perpetrator. The perpetrator was a European-American male undergraduate in his early 20s with brown hair, brown eyes, and a stocky build. A 4 × 6 color photograph of the perpetrator was used in the identification procedures.

Selection of foils. Possible foils were approached at a shopping mall and at Stanford University's student union. To create varying degrees of similarity between the perpetrator and the foils, three sets of five foils were compiled, all photographed in the same format as the perpetrator.

The foils in the low-similarity condition were five blond, blue-eyed male students who clearly did not match the description of the perpetrator. To select the foils for the high-similarity and moderate-similarity lineups, we photographed 14 men in their early 20s with brown hair, brown eyes, and a stocky build. Six independent raters judged the similarity of the perpetrator to these 14 foils. The raters viewed the photograph of the perpetrator for approximately 30 s. After a 10-min distractor task (responding to 100 general knowledge questions) they rated the physical similarity of each of the 14 photographs to the perpetrator. The photograph of the perpetrator was not presented a second time. Thus, judgments of physical similarity involved the retrieval of the perpetrator's appearance from memory, or judgments of ecphoric similarity. The photographs of the 14 men were presented individually and in a different random order for each rater.

Similarity was rated on a 21-point scale with the anchors *not at all similar* (1) and *very similar* (21). The average rating for each of the 14 photographs was computed and converted to ranks. The two ranks in the middle of the distribution (Ranks 7 and 8) were discarded, creating two groups of 6 photographs: a high-similarity and a moderate-similarity group. The second-ranked photograph in the high-similarity group was used as the substitute for the perpetrator in the blank lineups and in the high-similarity showup. The second-ranked photograph in the moderate-similarity group was discarded, and the photograph ranking fourth in the remaining set of 5 was used in the moderate-similarity showup. The remaining two groups of 5 foils comprised the high-similarity and moderate-similarity lineups.

---

[7] Only a handful of the subjects in these studies were African-American, so it was impossible to conduct meaningful analyses of same-race versus cross-race identifications.

As a check on the reliability of this method, a second group of six raters also judged the similarity of the 14 foils to the perpetrator, with both the perpetrator's photograph and the photograph to be rated presented concurrently. The same 21-point scale was used, and each rater viewed a different random order of the 14 photographs. The two groups of 5 foils generated from these raters were essentially identical to those derived from the first procedure.[8]

The five photographs for each lineup were placed on a poster board in a 2 × 3 array. The first position in the second row (Position 4) was reserved for a photograph of either the perpetrator or his substitute. The position of the perpetrator and his substitute was held constant because extensive pretesting showed that different arrays of the same six photographs produced lineups of different functional size (Gonzalez & Ellsworth, 1993), and it was important to hold functional size constant to assess the effects of similarity. Thus, rather than counterbalancing the position of the perpetrator (or his substitute) within Study 2, we simply varied the position across studies by selecting a different position from that used in Study 1.

Lineup assessment. We also assessed the functional size of the high-similarity and moderate-similarity perpetrator-present lineups. One hundred and twenty-four raters (or mock witnesses) were shown a lineup and told that the suspect was accused of stealing a purse. The perpetrator was described to them as a man with dark hair, brown eyes, and a chubby build. The raters were then asked to select the photograph of the person who best fit the description. The functional size for the high-similarity valid lineup was 3.4, and the functional size for the moderate-similarity valid lineup was 1.0. A similar procedure was used for the two perpetrator-absent lineups—here the assessed bias was toward the perpetrator substitute. The functional sizes were 6.5 for the high-similarity blank lineup and 1.2 for the moderate-similarity blank lineup.

Procedure

Subjects witnessed a videotaped staged crime either individually or in groups as large as 10. Before the tape was shown the experimenter informed the subjects that they would watch a video recording of a social interaction and then answer a few questions about it.

The staged crime took place at a food stand in the psychology department. The videotape begins with a view of the food counter from the perspective of the clerk. Two undergraduate women walk to the counter and order food. While the women are talking to each other the perpetrator enters the scene and grabs the purse of one of the women. While grabbing the purse, the perpetrator is in full view of the camera for about 2 s. He then runs out of sight. The victim does not realize her purse has been stolen until 4 s later, when she reaches over for her wallet to pay for the food, looks around frantically and yells, "Where's my purse?"

Subjects were told that the interaction they had just witnessed was a staged crime and that in a few minutes they would be asked a few questions about the incident. They were then given a 10-min distractor task consisting of general knowledge questions (e.g., What is the capital of Colombia?). At the end of the distractor task any group of subjects larger than 5 was divided into two groups. Thus, no group larger than 5 entered the identification phase of the experiment. Subjects were told

> Earlier you witnessed a staged crime on videotape. You will be asked to make an eyewitness identification and answer some questions about your memory for the staged crime. The identification procedure is very similar to those used by police today. Bear in mind that your decision may have an impact on procedures of future police identifications, so please take this task very seriously.

Subjects were then given a pre-identification questionnaire containing (a) two open-ended questions asking for free descriptions of the

incident and the perpetrator, (b) eight specific questions about particular characteristics of the perpetrator such as age, race, weight, and so forth; and (c) nine yes–no questions, such as "Was the victim wearing eye glasses?"

When subjects had completed the pre-identification questions they read the identification instructions. The identification procedure (lineup or showup) was described, and subjects were instructed not to discuss their choice with others in the room or to indicate whether they had made an identification. Following local police practice, we included the following caution in the instructions:

> The photograph[s] you are viewing may or may not depict the hair style or the facial hair similar to that of the person who committed the crime, since hair style, beards, and moustaches are easily changed. Note also that the true complexion of a person is not always depicted in photographs; the individual may be lighter or darker than shown.

> The fact that a suspect [lineup] is being shown to you should not influence your judgment. You should not conclude or guess that the person shown to you [someone in the lineup] committed the crime. You are not obligated to make an identification.

The experimenter reviewed the instructions with the subjects and explained the response categories. Subjects had three response options: (a) to state that the perpetrator was present and, in the lineup conditions to indicate which lineup member he was, (c) to state that the perpetrator was not present, or (c) to indicate that they could not remember the appearance of the perpetrator well enough to recognize whether his photo was present. Subjects were then presented with the photographs and given as much time as they needed to make their identification. There were three showup and six lineup conditions. The three showup conditions consisted of (a) the perpetrator, (b) a high-similarity foil, and (c) a moderate-similarity foil. The six lineups resulted from crossing the presence or absence of the perpetrator with the three levels of similarity.

Subjects were then given a postidentification questionnaire that consisted of four items. The first item asked subjects for a confidence rating ("How confident are you in the identification decision you made?"). The second item assessed their willingness to volunteer to testify in court ("Would you volunteer to testify in court regarding your identification decision?"). The third item attempted to assess prior expectations that the true perpetrator would be present in the identification procedure. We asked, "Before you actually saw the picture[s], did you think the perpetrator's picture would be included?"[9] The fourth item assessed subjects' estimates of how much weight would be given to their identification decision in actual legal proceedings ("If evidence of your identification decision were presented in court do you think the suspect would be convicted?"). All four of these items involved ratings on 9-point scales.

---

[8] We computed the reliability of the ratings by converting each rater's responses into ranks and then, within each set of six raters, calculated the average rank for each picture. The correlation between the two sets of average ranks was .83. The correlation between the average of the ratings (raw scores) for the two sets of raters was .85.

[9] We checked the data on this question after running half the subjects and discovered an insignificant trend suggesting that lineup subjects were more likely than showup subjects to believe that there was a better than even chance that the true perpetrator would be shown ($p \geq$ .19 by median test, with p computed through Fisher's exact test). The remaining 90 subjects were told that there was a 50% chance that the real perpetrator would actually be shown, and the trend disappeared. The change in instructions produced no differences in choice frequencies.

Results and Discussion

Response Bias

Response frequencies as a function of identification procedure and presence or absence of the perpetrator are shown in Table 2. Because similarity had little effect on the response patterns, descriptive statistics are collapsed across similarity conditions. Choice frequencies were analyzed through linear contrasts because we had specific predictions involving more than two proportions, and thus Fisher's exact test could not be used. We used the arcsine transformation method to test the contrasts (Gonzalez, 1993; Langer & Abelson, 1972; Mosteller & Tukey, 1969). Computing the contrasts on the raw (untransformed) proportions using the method recommended by Rosenthal and Rosnow (1985) or through logit analysis (Wickens, 1989) did not affect the pattern of results. As in Study 1, there were no significant differences across conditions in the frequency of I don't know responses, and statistical analyses were conducted on conditional probabilities (e.g., P[no/yes or no]).

The results show the same response bias as in the first study —more subjects said no when viewing a showup than when viewing a lineup. When the real perpetrator was shown, the proportion of no responses to the showup differed from the proportion of no responses to the three valid lineups, contrast $= (-3, 1, 1, 1)$; $z = 3.59$, $p < .001$. All 10 subjects who saw the perpetrator alone responded no, whereas, 10 of the 25 subjects (40%) who saw him in a lineup responded no. The results were similar when the showup was compared with the three lineup conditions (high, moderate, and low similarity) separately. The three pairwise comparisons between the valid showup and each of the valid lineup conditions (i.e., the three levels of foil similarity) were also significant, the three $(-1, 1)$ contrasts yielded $zs > 2.70$, $p < .007$.

If we look at the number of positive identifications elicited by each procedure, none of 10 subjects who saw the actual perpetrator in a showup responded yes. However, 15 of 25 subjects who saw him in a lineup responded yes. Surrounding the photograph of the perpetrator with pictures of other people increased the chance that a witness would say he was there.

A similar response bias occurred when the perpetrator was not shown. The two perpetrator-absent showups (high- and moderate-similarity foil) produced a higher rate of "no" responses (12 of 13 subjects, or 92%) than the three blank lineups (19 of 31 subjects, or 61%), contrast $= (-3, -3, 2, 2, 2)$; $z = 2.39$, $p < .017$. With one exception (described in the next section), the pairwise comparisons between the blank showups and blank lineups exhibited a similar pattern.

Showups Versus Lineups With a Functional Size of One

A major purpose of Study 2 was to compare the showup with a lineup with a functional size of one. When confronted with the perpetrator alone (valid showup) none of the 10 subjects identified him. However, when the same man was surrounded by implausible foils (low-similarity lineup), 6 of 9 subjects who were sure enough to make a definite response identified him. Thus, in the valid conditions a showup was not equivalent to a lineup with a functional size of one. The presence of foils who could easily be eliminated from consideration changed the nature of the response in a manner consistent with the distinction we proposed between absolute and comparative judgments.

When the perpetrator was not shown, however, the response to the low-similarity lineup was the same as the response to the showup: In both situations 12 of 13 subjects sure enough to make a definite choice correctly said that the perpetrator was not there. In general, subjects confronted with lineups of low functional size performed appropriately: They were more likely to make an identification when the true perpetrator was shown than when he was not shown ($z = 2.68$, $p < .01$). In the showup conditions there were no differences in the responses of subjects who saw the right man and subjects who saw the wrong man. Thus, the data from this study suggest that in general showups are not like bad lineups. Showups create a response bias to say no whether or not the actual suspect is shown; bad lineups lead to the rejection of the wrong person but not of the right one. The yes response bias characteristic of our other lineups disappeared in lineups with a functional size of one.

Table 2
Frequency of Choice as a Function of Identification Procedure for Study 2

Identification procedure	Choice (%)				Conditional proportions (%)	
	Yes	No	I don't know		P(y/y or n)	P(n/y or n)
Showup						
Valid	0	50	50		0	100
n		10	10			
Blank	2.5	30	67		8	92
n	1	12	27	Marginal	4	96
				n	1	22
Lineup						
Valid	25	17	58		60	40
n	15	10	35			
Blank	20	32	48		39	61
n	12	19	29	Marginal	48	52
				n	27	29

Note. y = yes; n = no.

Accuracy

Moving from the question of response bias to the question of accuracy, the results resemble those found in Study 1—correct identifications are more frequent in the lineups, but correct rejections are more frequent in the showups. When confronted with the actual perpetrator in a showup, not one of the subjects correctly identified him; when asked to choose the actual perpetrator from a lineup, 32% of the subjects sure enough to make a definite response correctly identified him. When confronted with the wrong man in a showup, 92% of the subjects who were sure enough to make a definite response correctly stated that he was not the perpetrator; faced with a blank lineup, 61% of the subjects correctly stated that the perpetrator was not there (collapsing across foil similarity).

The rate of false positives in the valid lineup was 28%. That is, 7 of 25 subjects sure enough to say either yes or no incorrectly selected a foil instead of the actual perpetrator. In the blank lineups 39% of the subjects identified someone even though the real perpetrator was not there. Five of these subjects (42%) selected the perpetrator substitute.

Confidence Measures

After testing half the subjects it became apparent that the confidence measure, as worded, would not be useful because only 1 subject in the showup condition had made a positive identification. Thus, we changed the format of the confidence measure for the remaining subjects. Subjects in the second half of the experiment rated the likelihood that the person shown to them was the actual perpetrator on a 9-point scale with the anchors *definitely not the perpetrator* (1), *even chance he is or is not the perpetrator* (5), and *definitely the perpetrator* (9). Subjects in the showup conditions rated the person shown to them; subjects in the lineup conditions rated all six lineup members. Note that the change in procedure occurred after the subject had made the identification and hence could not affect the identification data.

Do the number and similarity of the foils affect the likelihood rating for the actual perpetrator? A one-way analysis of variance of the likelihood ratings for the actual perpetrator in the valid showup and lineups (over the three levels of foil similarity) revealed no treatment effect ($F < 1$). The mean likelihood rating for the perpetrator in the valid showup was 5.4 ($SD = 2.06$). The mean ratings for the perpetrator in the three valid lineup conditions were 4.3 ($SD = 1.95$), 4.6 ($SD = 2.27$), and 5.13 ($SD = 2.13$) for the high, moderate, and low levels of foil similarity, respectively. Mean likelihood judgments in lineups were inversely related to the similarity of the foils, with the showup producing the highest level of confidence. Clearly, the small sample sizes (approximately 10 per cell) do not provide a powerful significance test. However, power analysis revealed that, given our sample estimates, the sample size would have to be at least 80 per cell to approach acceptable levels of power ($\beta = .20$).

The result that the average likelihood rating for the perpetrator did not differ statistically across conditions is interesting in light of the data on actual choices. It appears that subjects believed the suspect was as likely to be the real perpetrator when they saw him alone as they did when they saw him in a lineup.

However, they were substantially less willing to make a positive identification in the showup. This inconsistency between the likelihood rating and the identification response is consistent with the absolute–comparative distinction, suggesting that the decision-making strategies used in the showup and the lineup involve different criteria.

Discussion of Study 1 and Study 2

Contrary to the common assumption of legal and psychological experts (but consistent with the views of the police), our results suggest that the formal task structure of a one-person showup does not create an unacceptable increase in the risk that an innocent suspect will be identified as the perpetrator. Across both studies, only 2 subjects (3.7%) identified an innocent suspect in a showup as the perpetrator, whereas 7 subjects (9.2%) picked an innocent suspect from a six-person lineup. The protection afforded by embedding the suspect among five innocent foils is largely dissipated by the much greater willingness of the witness to select a lineup member than a lone suspect. Subjects confronted with six men were fairly likely to "find" the man they had seen. Subjects confronted with one man were highly unlikely to say that he was the one they had seen. The same pattern of results was found for both live and photographic identification procedures and for African-American and European-American suspects. This response bias also produces differences in the types of errors observed. The showup has a relatively high rate of misses, whereas the lineup has a relatively high rate of false positives.

These results are consistent with our idea that lineups and showups elicit different processing strategies. The structure of the lineup elicits comparative strategies because the witness must select one member from several alternatives. In the lineup the witness must first select the member who best matches the perpetrator. If a best match is found, the witness must then decide whether that person is the perpetrator, that is, whether the best match is good enough. The problem is that the initial process of finding a best match may bias the witness to make an identification. This bias may involve both motivational and cognitive processes. The witness may feel a sense of accomplishment at having found a best match, or the process of eliminating lineup members may increase the salience of the best match. The presence of five people who look less like the perpetrator may make the match appear better than it actually is. Confirmatory biases may also result from attempts to explain why the particular lineup member is the best match. In the showup the only task of the witness is to decide if the suspect presented is or is not the perpetrator; there can be no biases resulting from finding a best match.

The effects of such processing differences become evident when comparing a showup with a lineup that contains only one plausible choice (i.e., a lineup with a functional size of one). Researchers in the field have made the intuitively plausible argument that a lineup that surrounds the suspect with foils who do not resemble him or her is equivalent to presenting the suspect alone (a showup). Lineup members who do not fit the general description of the perpetrator can be rejected from consideration, effectively transforming a lineup into a showup. However, the results of Study 2 show that the lineup consisting of one plausible member led to a different response pattern

than the showup. Witnesses were still more likely to identify someone from a lineup than from a showup. The process of eliminating foils appears to influence the identification response. We take this as evidence that the two procedures elicit different processing strategies, a comparative strategy in the lineup and an absolute strategy in the showup (Shafir, Osherson, & Smith, 1989, used similar concepts in the context of decision making under risk).

The claim that lineups lead to comparative judgments is not new (Cutler & Penrod, 1988; Lindsay & Wells, 1985; Wells, 1984). For example, Lindsay and Wells attempted to reduce the comparative nature of judgments in lineups by means of a *sequential lineup*, in which the witness observes the lineup members one at a time and is asked to make a yes–no decision for each one. The sequential lineup eliminates spatial proximity. The witness is not informed of the exact number of judgments he or she will make (though in the study by Lindsay & Wells subjects saw the stack of 12 photographs they would judge). Lindsay and Wells found a higher percentage of "no" responses in the sequential lineup (0.48 and 0.65) than in the traditional simultaneous lineup (0.30 and 0.42, for perpetrator-present and perpetrator-absent lineups, respectively). Thus, by presenting one person at a time, the sequential lineup minimizes the opportunity for comparative judgments that exists when a witness observes all lineup members simultaneously.

Although Lindsay and Wells (1985) did not include an I don't know option, their data parallel the present results. The response pattern in the sequential lineup is similar to the pattern we observed in the showup. Both the sequential presentation of lineup members and the forced yes–no decision for each member suggest that the sequential lineup leads to an absolute judgment. A more appropriate name for the sequential lineup may be *multiple showup*. The sequential lineup, then, may be one solution to the biases stemming from the comparative nature of lineup judgments.

Eyewitness Identification and Recognition Memory

Although eyewitness identification procedures are basically tests of recognition memory, with the showup analogous to a yes–no test and the lineup analogous to a *k*-alternative forced-choice test, fundamental differences between the task requirements of a laboratory memory test and the requirements of a police lineup limit our ability to generalize from one to the other. Paradigms used in memory research separate the act of selecting from the act of deciding. That is, in a yes–no test the subject decides whether the new item is the same as the old item (as a witness does in the showup); in a forced-choice memory test the subject knows the old item is present in the set and must decide which one it is. The police lineup confounds these two processes. Because the police do not know whether the suspect is the actual criminal, the lineup is not a true forced-choice test. The task of the witness who observes a lineup is twofold: (a) to select someone from the lineup and (b) to decide if he or she is the perpetrator. Also the eyewitness identification paradigm is not conducive to the standard analyses performed in memory research. Measures of sensitivity and bias (such as d' and β) cannot be computed because each witness provides only one data point.

These differences limit the direct application of the findings

and techniques of memory research even to the formal problems involved in the eyewitness identification task. Nevertheless, signal detection theory may provide a useful framework to explore and test the absolute–comparative distinction. Indeed, our analysis of error rates, suggesting that false negatives are more frequent in showups but false positives are more frequent in lineups, is a step in this direction. Models and experimental paradigms used in the study of categorization and perception may also be useful (e.g., Ashby & Gott, 1988; Warren, 1985). A signal detection model, for example, suggests that the difference between absolute and comparative judgment strategies is not the only possible explanation for our results. Malpass and Devine (1984) pointed out that signal detection theory also directs attention to the possibility of differences in the payoff matrix associated with the particular response alternatives. It is possible, for example, that the risks of false identification are more obvious to the witness in showups than in lineups, leading to different estimates of values and subjective probabilities and ultimately to the bias against saying yes we observed in showups.[10]

A different view of recognition memory (as opposed to models that incorporate a familiarity dimension) is that people make similarity judgments between features of the new stimulus and the old stimulus. In the context of eyewitness memory this model suggests that the witness judges the similarity between the suspect and his or her memory of the perpetrator. In similarity judgments it is useful to distinguish the subject from the referent (see Tversky, 1977). Judgments of the form "assess the degree to which A [subject] is similar to B [referent]" do not always lead to the same judgment as "assess the degree to which B is similar to A" (Tversky, 1977, p. 333). That is, the similarity relation is asymmetric.

The showup and the lineup may differ on whether the suspect is the subject of the similarity judgment or the referent. Because the showup presents one person to the witness, a natural strategy for the witness is to treat the lone suspect as the subject and his or her memory trace of the perpetrator as the referent (i.e., comparing the suspect with the memory trace). On the other hand, in the lineup the natural strategy is to treat the memory of the perpetrator as the subject to which each lineup member is a referent (i.e., comparing the memory trace with the lineup members). This yields an efficient strategy for finding a best match among several alternatives. Recent research suggests that features of the referent are more mutable than features of the subject (Houston, Sherman, & Baker, 1989; Kahneman & Miller, 1986); if these results generalize to the domain of eyewitness memory, sources of distortions and biases may be different in lineups and showups. The lineup may lead to selective attention to the features of the lineup members, whereas the showup may lead to selective recall of the features of the perpetrator. It is not clear that the subject–referent distinction explains the response pattern observed in the present studies, but the distinction in the context of eyewitness identification may be useful and is worth further investigation.

Study 3

Do these surprising laboratory results have any applicability to real-world lineups and showups? Perhaps the showup's repu-

---

[10] This idea was suggested to us by Roy S. Malpass.

tation for impermissible suggestiveness has nothing to do with the fact that one person is shown but with the police's behavior and the witness's state of mind. The witness, still shaken by the recent crime, is confronted by a frightened or hostile suspect restrained by the police, who are very much hoping to catch the perpetrator and close the case quickly. Although the formal properties of the lineup and showup tasks do not support the idea that showups lead to a greater willingness to make an identification, perhaps these task properties are of trivial importance in the interpersonal context of an actual encounter between a witness and a suspect in the presence of the police.

Although arguments about the highly suggestive nature of police showups have been expressed with great confidence by both psychological researchers and Supreme Court Justices, it is important to point out that these arguments, like the arguments about task properties, are based entirely on intuition and anecdote. There are no empirical studies of identification patterns in actual lineups and showups, just as there were none of simulated lineups and showups. There are, of course, examples of egregiously suggestive showups, like Stovall's (*Stovall v. Denno*, 1967) hospital room confrontation, but there are also plenty of examples of egregiously suggestive lineups where the suspect is the only plausible choice or is singled out for special attention (Loftus, 1979). There is in fact no actual evidence of the relative suggestiveness of lineups and showups.

Assuming that appropriate foils are chosen, how could a lineup possibly be more suggestive than a showup? Even given strong pressure to cooperate with the police, there is only a one in six chance that a witness will choose the perpetrator, and other choices are usually harmless. However, there are perhaps more different ways to bias a lineup than a showup, without the bias becoming blatant: choice of foils, placement of the suspect, timing, subtle forms of extra attention to the suspect, reinforcement or nonreinforcement of tentative answers, and so on.

Probably most important, by the time the police conduct a lineup, they are usually convinced that they have the right person and are looking for confirmation. Often an eyewitness identification is essential to go forward with the prosecution, for example, if the evidence against the suspect consists of the statements of an informant, illegally seized evidence, or an inadmissible confession (Gross, 1987). In such a situation all the hypothesis-confirming biases identified by Rosenthal (1966) are likely to operate in full force as the police communicate their hypothesis to the witness.

When the police stop someone on the street and conduct an in-field showup, they do not know whether it is the right person. In this context they sincerely want information, not confirmation, from the witness. Therefore police pressure on the witness to make an identification may be considerably less in the typical showup than in the typical lineup.

Thus, there are arguments and anecdotes on both sides. An important first step is to find out how often real witnesses actually make identifications in lineups and showups. To this end, we enlisted the aid of a police detective in a town near Stanford and asked him to keep track of all the identifications he was involved in, both lineups and showups, until he reached 50 lineups.[11] The notes for each case included a description of the crime, the arrest, the identification procedure, the suspect (or suspects), the witness (or witnesses), and the outcome. Usually the detective also added a few comments or reactions.

All of the lineups were photo lineups, almost all of the 172 showups involved showing the actual suspect to the witness.[12] The lineups were conducted from 3 hr to 3 weeks after the crime (mode = about a week), showups from 15 min to 2 days after the crime (or a later police contact; mode = within 2 hr). Showup suspects were usually located on the basis of witnesses' descriptions of the person or clothing, sometimes on the basis of "suspicious behavior" such as fleeing or jettisoning incriminating objects. Lineup suspects were located (in order of frequency) because the description matched that of someone already known to the police, because car owners were tracked down from descriptions of the car or license plate, because of tips from informants, because of other descriptions, or because the police arrested them for another crime. In showups, the witness was usually taken to the place where the police had stopped the suspect. In lineups, either the witness was brought to the police station or (less frequently) the photo array was taken to the witness's home or workplace. The data must be regarded as approximate because some of the lineups involved multiple witnesses, some involved the same witness viewing multiple lineups, and some involved both, so the observations are not independent. Nonetheless, the data provide an initial check on whether our counterintuitive laboratory findings are totally at odds with what goes on in real identifications. If there were many more positive identifications made in showups than in lineups, for example, we would have to conclude that the differences we have identified in judgment processes are overshadowed and reversed by other influences in the real world. Our task would then be one of identifying the suggestive interpersonal and situational forces at work.

Table 3 shows that in this sample, as in our two experimental studies, witnesses were more likely to make positive identifications of suspects in lineups than in showups (56% vs. 22%) and were far more likely to assert that the suspect was *not* there in showups than in lineups (31% vs. 6%). If we limit the data to only those lineups that we can be sure were completely independent (i.e., different suspect, different foils, and different witness), positive identifications were made in 18 of 24 lineups (75%). All 29 of the witnesses who identified a member of the lineup as the perpetrator chose the suspect. None of the foils was ever misidentified as the culprit.

These field data also address two possible criticisms of the laboratory studies. First, in our own discussion with colleagues unfamiliar with criminal investigation we are frequently asked "Why would anyone care about the difference between showups and lineups since a showup would never occur in the real world?" (television and movies almost always portray lineups). In our sample showup identifications were over three times more common than lineups, and follow-up research currently underway in Washington and Michigan suggests that showups are frequently used. The second criticism concerns the high rate of I don't know responses observed in the two studies. The I

---

[11] We are grateful to James Zboralski for collecting these data. There were actually 52 lineups because the last case investigated by Zboralski involved several lineups.

[12] This confound reflects reality. Live lineups are rarely used any more in this jurisdiction. Photo lineups are less trouble, and there is no right to counsel at photo lineups.

Table 3

Frequency of Choice as a Function of Identification Procedure for Field Study

Identification procedure	Choice (%)			
	Yes	No	I don't know	Total
Showup	22	31	47	
n	38	54	80	172
Lineup	56	6	38	
n	29	3	20	52

don't know response rate in the field data is comparable with that observed in Study 1 and not far off from the rate in Study 2. The rate of I don't know responses observed in our laboratory studies may not be unrealistic.

Although the data in Study 3 are consistent with our hypothesis, it would be wrong to base any strong conclusions on them. We conducted this study simply to find out if data on identification procedures in the real world would falsify our hypothesis, in Popper's (1959) terms. That the hypothesis is not falsified does not mean that it is confirmed. Much more needs to be known about real-world identification procedures to truly assess their suggestibility. We need to know, for example, the reasons why lineups and showups are conducted, the witnesses' beliefs about these reasons, the methods of selection of the target (and foils), and the typical time lapse between the crime and the identification procedure. In addition, we speculate that by the time the police conduct a lineup they are more likely to have the right person, so the base rates of target-present lineups are higher than for target-present showups.

General Discussion

Our studies provide no evidence that one-person showups are more suggestive than full lineups. Instead they suggest that lineups may provide less protection against misidentification and showups less danger of misidentification than is commonly believed. People appear to approach showups more cautiously; they are more reluctant to say that the person they see is the perpetrator, even when he or she is. This reluctance was also apparent in our field data, suggesting that any social or situational pressures to make a showup identification are insufficient to overcome witnesses' hesitation. In both the laboratory and the field, witnesses are far more likely to pick out a member of the lineup than they are to identify a lone suspect.

Both task factors and real-world factors may contribute to this bias. As we have discussed earlier, witnesses faced with a lineup may choose the person who most resembles their memory of the perpetrator. Having identified the best match, they name him or her, without addressing the additional task of deciding whether in absolute terms the best match is good enough. In the showup, witnesses are forced to make an absolute judgment—is this the same person?

Contrary to popular belief, we also hypothesize that police pressures to make an identification—the "right" identification —may be greater in the lineup than in the showup. When the police conduct a showup, typically they have no great confi-

dence that they have the right person—they've picked up someone who seems to fit the description, or who seems out of place in the neighborhood or who is running—and they honestly need information. A no answer from the witness is as acceptable as a yes answer. Usually by the time the police present the witness with a lineup they are no longer looking for information; they are looking for evidence that can be used to compel a plea or persuade a jury (Gross, 1987). The person conducting the lineup knows who the suspect is and often "knows" that the suspect is guilty. If psychological researchers, presumably committed to objectivity, can communicate their expectations to subjects and bias the data toward their hypothesis (Rosenthal, 1966), so can police officers. Some of the comments of the detective who collected our field data provide anecdotal support for our intuitions about the conduct of lineups: "We took another photo, and once again the suspect was not identified. Without any other leads to pursue . . . the investigation was stalled at this point. We never did solve the case which really made us mad because we know we had the right crook." "We got a photo and showed our 3 victims the lineup. Two of them identified the suspect; one did not. The woman who did not ID the suspect was quite nervous and embarrassed and didn't want to get involved. I think she purposely didn't identify the suspect."

The fact that every person who made a lineup identification identified the suspect may provide some further evidence of pressure and bias in the lineup. Certainly, it suggests that the odds against a mistaken identification of the suspect in a lineup are not what they appear to be. It is often argued that most mistaken identifications in lineups are harmless because they are identifications of foils known to be innocent; thus, lineups provide protection against overeager but unreliable witnesses that showups do not afford. In our sample there were no harmless mistaken identifications; if there were any mistaken identifications, they were mistaken identifications of the suspect.

Is it possible that this particular police department conducts laughably poor lineups? We were able to examine three of the photo lineups; they looked quite good to our somewhat experienced eye, and informal tests for functional size did not suggest that the suspect was particularly likely to be chosen. Thus, there are two possible explanations for the fact that the suspect was always chosen. First, the police may have arrested the right person in all 29 cases, and the witnesses recognized him or her. Second, subtle and not-so-subtle pressures in the identification procedure may have biased witnesses to choose the suspect. It is probably the case that by the time the police conduct a lineup they usually have the right person (Gross, 1987), and it is probably also the case that their strong belief that they have the right person substantially increases the risks of misidentification when that belief is unfounded.

Obviously, these hypotheses about the psychology of police officers and witnesses are mere speculations at this point. They seem plausible to us, implausible to some of our colleagues. It is somewhat astonishing that after 20 years of intensive research on eyewitness identification, we know next to nothing about actual practice. Our current research is designed to explore police officers' assumptions about the identification procedures they use, their purposes, and the circumstances surrounding the choice and conduct of lineups and showups.

References

Aronson, E., Ellsworth, P. C., Carlsmith, J. M., & Gonzales, M. (1990). *Methods of research in social psychology* (rev. ed.). New York: McGraw-Hill.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 33–53.

Blackwell, H. R. (1953). *Psychophysical thresholds: Experimental studies of methods of measurement.* Ann Arbor, MI: Engineering Research Institute at University of Michigan.

Blumenthal, A. L. (1977). *The process of cognition.* Englewood Cliffs, NJ: Prentice-Hall.

Bower, G. H. (1967). A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (pp. 229–325). San Diego, CA: Academic Press.

Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. *Journal of Applied Psychology, 73,* 281–290.

Doob, A. N., & Kirshenbaum, H. M. (1973). Bias in police lineups—partial remembering. *Journal of Police Science and Administration, 1,* 287–293.

Ellison, K. W., & Buckhout, R. (1981). *Psychology and criminal justice.* New York: Harper & Row.

Gilbert v. California. 1967. 388 US 263.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Gonzalez, R. (1993). *Contrasts on binomial data for between-subjects designs.* Unpublished manuscript.

Gonzalez, R., & Ellsworth, P. (1993). *Assessing lineup fairness: Psychological factors in functional size.* Manuscript in preparation.

Gross, S. (1987). Loss of innocence: Eyewitness identification and proof of guilt. *Journal of Legal Studies, 16,* 395–453.

Houston, D. A., Sherman, S. J., & Baker, S. M. (1989). The influence of unique features and direction of comparison on preferences. *Journal of Experimental Social Psychology, 25,* 121–141.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93,* 136–153.

Kassin, S. M., Ellsworth, P. C., & Smith, V. L. (1989). The "general acceptance" of psychological research on eyewitness testimony: A survey of the experts. *American Psychologist, 46,* 1089–1098.

Kintsch, W. (1967). Memory and decision aspects of recognition learning. *Psychological Review, 74,* 496–504.

Kintsch, W. (1970). *Learning, memory and conceptual processes.* New York: Wiley.

Köhnken, G., & Maass, A. (1988). Eyewitness testimony: False alarms or biased instructions. *Journal of Applied Psychology, 73,* 363–370.

Langer, E. J., & Abelson, R. P. (1972). The semantics of asking a favor: How to succeed in getting help without really dying. *Journal of Personality and Social Psychology, 24,* 26–32.

Levine, F., & Tapp, J. (1973). The psychology of criminal identification: The gap from *Wade* to *Kirby. University of Pennsylvania Law Review, 121,* 1079–1131.

Lindsay, R. C., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70,* 556–564.

Loftus, E. F. (1979). *Eyewitness testimony.* Cambridge, MA: Harvard University Press.

Macmillian, N. A., & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" indexes. *Psychological Bulletin, 107,* 401–413.

Malpass, R. S., & Devine, P. G. (1981a). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology, 66,* 482–489.

Malpass, R. S., & Devine, P. G. (1981b) Realism and eyewitness identification. *Law and Human Behavior, 4,* 347–358.

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.

Malpass, R. S., & Devine, P. G. (1984). Research on suggestion in lineups and photospreads. In G. Wells & E. Loftus (Eds.), *Eyewitness testimony* (pp. 64–91). Cambridge, England: Cambridge University Press.

Manson v. Brathwaite. 1977. 432 US 98.

Mosteller, F., & Tukey, J. (1969). Data analysis, including statistics. In G. Lindsey & E. Aronson (Eds.), *Handbook of social psychology* (2nd ed.) (pp. 183–199). Reading, MA: Addison-Wesley.

Neil v. Biggers. 1972. 409 US 188.

Popper, K. (1959). *The logic of scientific discovery.* New York: Harper Torchbooks.

Rosenthal, R. (1966). *Experimenter effects in behavioral research.* New York: Appleton-Century-Crofts.

Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analysis: Focused comparisons in the analysis of variance.* Cambridge, England: Cambridge University Press.

Shafir, E. E., Osherson, D. N., & Smith, E. (1989). An advantage model of choice. *Journal of Behavioral Decision Making, 2,* 1–23.

Smith v. Coiner. 1973. 473 F.2d 877.

Sobel, N. R. (1985). *Eyewitness identification: Legal and practical problems* (2nd ed.). New York: Clark & Boardman Co.

Stovall v. Denno. 1967. 388 US 293.

United States v. Wade. 1967. 388 US 218.

United States v. Wilson. 1970. 435 F. 2d 403.

Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning and Verbal Behavior, 20,* 479–496.

Tversky, A. (1977). Features of similarity. *Psychological Review, 24,* 327–352.

Wall, P. M. (1965). *Eye-witness identification in criminal cases.* Springfield, IL: Charles C Thomas.

Warren, R. M. (1985). Criterion shift rule and perceptual homeostasis, *Psychological Review 92,* 574–584.

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14,* 89–103.

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.

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin, 88,* 776–784.

Wickens, T. (1989). *Multiway contingency tables analysis for the social sciences.* Hillsdale, NJ: Erlbaum.

Woocher, F. D. (1977). Did your eyes deceive you? Expert psychological testimony on the unreliability of eyewitness testimony. *Stanford Law Review, 29,* 986–1030.

Yarmey, A. D. (1979). *The psychology of eyewitness testimony.* New York: Free Press.