A better understanding of eyewitness identification from lineups and photospreads has emerged in the psychological literature over the last dozen years. The structural and procedural aspects of lineups have been likened to the conduct of an experiment with direct analogies to experimenter effects, confoundings, and control conditions. The process of lineup identifications seems to be governed at least in part by relative similarity judgments, which prove to be problematic whenever the actual culprit is not a member of the lineup. Several advances have been made in the instructions, structure, content, and evaluation of lineups and photospreads that can improve the ratio of accurate to false identifications. Combinations of the best features of a lineup can produce well-defined upper limits on false identification rates.

Suppose that you were an eyewitness to a crime. Perhaps it was a theft, a burglary, a mugging, a drive-by shooting, or a robbery. You might or might not have known that a crime was being committed at the time; perhaps you saw someone exit a building that exploded a short time later. Perhaps you were the victim or perhaps you were a bystander. Regardless of the circumstances, there exists some memory trace, however strong or weak, that could have important consequences for the course of justice. Because you have seen the culprit, the police ask you to give a description. Later, perhaps only hours or perhaps months later, you are called to the police station to attempt an identification of the culprit. You are then shown a lineup or a photospread and asked to indicate whether the person you saw on that fateful occasion is one of the people standing or pictured before you on this day.

If a research psychologist were asked what psychology knows about eyewitness identification of culprits from police lineups, what could be said? Fewer than a dozen years ago, the answer would have been rather brief and not highly informative. Most likely, the response would have been something like the following: False identifications occur with surprising frequency in staged-crime experiments and most people seem to place too much faith in eyewitness identification evidence. Today, several thousand staged crimes later, we have no reason to retract that conclusion. Nevertheless, we now have much more to say about the variables controlling the accuracy of eyewitness identification and how accuracy can be improved in actual cases. Research on the structural and statistical properties of lineups and photospreads has greatly enhanced our ability to specify sets of procedures that minimize false identification rates and maximize accurate identification rates.

Familiarity with these developments is becoming increasingly important for psychologists because players in the legal system (e.g., attorneys, judges, police) are now viewing psychologists as their main resource for help on issues of eyewitness identification from lineups and photospreads. Indeed, most major psychology departments in the United States, Canada, the United Kingdom, and Australia have probably been contacted within the past year or two concerning information or expert testimony on eyewitness issues. How many research psychologists are in a position to provide meaningful, research-based information about eyewitness identification?

This article can be useful in developing a sense for the nature of the identification problem and some suggested remedies. It should also be noted that the emphasis here is on eyewitness identification rather than eyewitness testimony in general. Readers are encouraged to consult other sources for overviews of other aspects of eyewitness memory (e.g., Ceci, Toglia, & Ross, 1987; Clifford & Bull, 1978; Doris, 1991; E. F. Loftus, 1979; Wells & Loftus, 1984; Yarmey, 1979) or for discussions of the debate over expert testimony (e.g., Buckhout, 1986; Cutler, Penrod & Stuve, 1988; Fox & Walters, 1986; Kassin, Ellsworth, & Smith, 1989; Lempert, 1986; E. F. Loftus, 1983, 1986; McCloskey & Egeth, 1983; McCloskey, Egeth, & McKenna, 1986; Pachella, 1986).

This article guides readers through a series of logical arguments, theoretical suppositions, and empirical observations that have clear implications for developing an understanding, prediction, and control of eyewitness identification errors. In developing an understanding, it is essential that readers grasp the importance and implications of the various elements of the overall picture that has emerged in the literature. It is critical, for example, that one understand the methods (experimental) and framework (system variable) from which this work flows, the function that lineups serve (uncertainty reduction),

The author thanks Brad Buscian, Meg Gerrard, Rick Gibbons, Anita Kelly, Kevin McKillop, Teddy Warner, Ted Wright, and the social psychology graduate students at Iowa State University for comments on an earlier version of this article. Portions of this article were supported by a grant to the author from the National Science Foundation (SES 9002182).

Correspondence concerning this article should be addressed to Gary L. Wells, Department of Psychology, Iowa State University, Ames, IA 50011-3180.

1 Photospreads are the photographic counterpart to lineups and are distinguished from mugshot searches in two ways. First, photospreads, like lineups, explicitly include a suspect whereas mugshot searches do not have a priori suspects. Second, photospreads and lineups have a relatively low number of members (typically 6–11) whereas mugshot searches might include hundreds of photographs.
the structural features of lineups (target present vs. absent; single vs. all-suspect models), the primary rationale and guiding analogy for suggested improvements (lineup-as-experiment analogy), and the processes giving rise to positive identification decisions (relative-judgment process, social inference, social pressure). This kind of knowledge is required to understand some of the proposed solutions to the eyewitness identification problem, such as the use of sequential lineups, blank lineups, high functional size, and high propitious heterogeneity.

Is There a Problem?
This article assumes the existence of an "eyewitness identification problem," and this assumption is based on three general observations. First, experimental studies using a simulated or staged-crime methodology have tended to find that false identifications occur with surprising frequency (e.g., Brigham & Cairns, 1988; Brigham, Mass, Snyder, & Spaulding, 1982; Buckhout, 1974; Cutler, Perrod, & Martens, 1987a; Davies, Ellis, & Shepherd, 1978; Ellis, Shepherd, & Davies, 1980; Leippe, Wells, & Ostrom, 1978; Lindsay, 1986; Lindsay & Wells, 1980, 1985; Lindsay, Wells, & Rumpel, 1981; E. F. Loftus & Greene, 1980; Malpass & Devine, 1981a; Parker & Caranza, 1989; Shepherd, Ellis, & Davies, 1982; Wells, 1984b; Wells, Ferguson, & Lindsay, 1981; Wells & Leippe, 1981; Wells, Lindsay, & Ferguson, 1979). How high were the rates of false identification in these studies? The rates varied profoundly, from as low as a few percent to greater than 90%. Indeed, part of the central message of this article is that false identification rates vary systematically as a function of many factors, some of which can be directly controlled in actual criminal investigations.

The second general observation supporting the idea that there is an identification problem is that there is a sincerety to most of the false identifications observed in these experiments. The term sincerity is meant broadly here to refer to the fact that many of the eyewitnesses who make false identifications in these types of experiments do so with considerable subjective certainty that their identifications are accurate (see Deffenbacher, 1980; Leippe, 1980; Wells & Murray, 1984). This, in turn, results in eyewitnesses who have made false identifications often being as persuasive in their testimony as are eyewitnesses who have made accurate identifications (Wells et al., 1979). High rates of persuasive false identifications have been observed in staged-crime experiments that maintain the ruse (i.e., are perceived by the eyewitnesses as actual or genuine criminal investigations) throughout identification and testimony phases (Luus & Wells, 1991; Malpass, Devine, & Bergen, 1980; Murray & Wells, 1982). Hence, false eyewitness identifications obtained in staged-crime experiments seem to constitute a genuine phenomenon that likely can be generalized to real cases under a minimal set of reasonable assumptions. In this sense, the experimental data speak to the plausibility of the argument that there is an eyewitness identification problem.

The third general observation bearing on the assumption that there is an eyewitness identification prob-

lem stems from studies of documented cases of wrongful conviction. Analyses of what went wrong in producing more than 1,000 convictions of innocent people have revealed that the single largest factor leading to these false convictions was eyewitness error (Branden & Davies, 1973; Borchard, 1932; Frank & Frank, 1957; Huff, Rattner, & Sagarin, 1986). Of course, the fact that known cases of false conviction are significantly attributable to eyewitness error does not alone prove that there is a significant eyewitness problem. These cases of false conviction were discovered to be false by rare and unpredictable events; hence, we do not know if they are representative of false convictions in general, and there might be no reliable way to make such a determination. Even if we knew that eyewitness error accounted for some constant percentage of all false convictions, say 55%, we would have no clear method for estimating the frequency of false convictions.

Thus, the assumption that there is an eyewitness problem is based on three observations: Eyewitnesses in psychology experiments can often make inaccurate statements about a previously witnessed event, their false statements appear to be sincere, and faulty eyewitness accounts have been implicated as the primary single cause of documented false convictions. It does not necessarily follow from these three observations that there is an eyewitness problem of significant social importance. At the very least, however, these observations are consistent with the idea that there is a significant eyewitness problem and probably are at the heart of the assumptions underlying many eyewitness theorists' programs of research.

Methods and Framework
Throughout the modern research period on eyewitness identification, which began in the 1970s and flourished in the 1980s, the paradigm of choice has been to use simulated events. These events, carefully controlled by the experimenter, have taken three general forms: Slide sequences, film scenarios, and live stagings. Typically, the subjects are unaware that they will have their recollections tested and that they will have to identify someone from a lineup or photospread. In other words, the subjects are caught by surprise. With media-depicted events, such as videotape, the subjects might be led to believe that they are watching to develop "impressions" of some aspect of the visual presentation (e.g., the aesthetics of the scenery). With a live, staged event, subjects might encounter the event unexpectedly in a waiting room or even a public restroom. Regardless of the circumstances, the subject is effectively transformed into an eyewitness by virtue of having observed a critical event.

An important feature of eyewitness identification experiments is that an established reality can be compared with the eyewitness's responses. In particular, the actual identity of the culprit is known; hence, identification responses can be categorically scored for their validity. Alternative methods have been used, such as Yuille and Cutsbush's (1986) analysis of multiple eyewitnesses' accounts of an actual shooting incident, but there are two
problems with using data from actual cases. First, sup-
positions must be made about what actually occurred and, because of the danger of witness interaction and other sources of witness-correlated errors, mere consensus among witnesses is a questionable criterion. Second, actual cases do not permit the kinds of experimental manipulations that allow the establishment of cause–effect relations among variables. Accordingly, experiment-generated events are highly favored among eyewitness re-
searchers.

Within this method, a system-variable approach is commonly, although not exclusively, used. The system-variable approach is one in which the critical variables that are manipulated in the experiment can also be controlled in actual cases by people in the criminal justice system. An example of a system variable would be the number of people in a lineup. The number of lineup members is something over which the system (police in particular) has control. System variables can be contrasted with estimator variables, which are sources of eyewitness error or accuracy that are beyond the control of the criminal justice system. An example of an estimator variable would be the amount of time that the eyewitness had the culprit in view during the criminal incident. Although the amount of time that the culprit was in view can be manipulated and controlled in an experiment, it is beyond the system’s control in actual cases and, hence, is not a system variable.

Developing a complete understanding of the social and cognitive processes that can affect eyewitness testimony requires both system- and estimator-variable research (Wells, 1978). The general bias favoring system-variable eyewitness research stems from the practical observation that such research can help to inform the justice system about ways to reduce eyewitness identification errors. In effect, the concept of system variables is synonymous with the concept of preventable error.

It is difficult to prove that the experimental studies on which the scientific literature on eyewitness identification is based provide an adequate parallel to actual criminal cases. Critics will always contend that real eyewitnesses are better or worse or somehow different than the people used in eyewitness experiments. Critics also might charge that eyewitnesses in experiments are not as cautious about their identification decisions because there is no actual consequence to the falsely identified person and that the experimental situation is treated more like a game.

More studies on the issue of real-case generalization need to be conducted. It should be noted, however, that the experimental eyewitness literature has explicitly included young children as well as elderly adults as subjects—witnesses and often these experiments are conducted in the subjects’ natural setting, such as their own homes, workplaces, or schools (see Ceci et al., 1987; Dent & Flin, 1992; Doris, 1991; Goodman & Aman, 1990; Goodman, Bottoms, Schwartz-Kenney, & Rudy, 1991; Goodman & Reed, 1986; Leippe, Romanczyk, & Manion, 1991; Parker & Caranza, 1989; Yarmey, 1984). The vast bulk of the eyewitness experiments do not forewarn subjects that they will be eyewitnesses, thereby capturing the surprise element and the natural, incidental aspect of actual eyewitnesses. In some experiments, the experience is carefully crafted to maintain the ruse of an actual crime throughout the identification phase of the study (e.g., Luus, 1991; Malpass & Devine, 1980; Murray & Wells, 1982).

The subset of experiments using nonstudent subject populations, nonlaboratory settings, and high realism do not reveal data patterns that represent significant problems for generalizing from the typical eyewitness experiment. For example, although it is true that very young children and the elderly tend to be less accurate than the typical college student subject, there do not appear to be any fundamental differences in the principles governing their performance as eyewitnesses. A study by O’Rourke, Penrod, Cutler, and Stuve (1989), for example, found main effects for the age of eyewitnesses (who varied from 18 to 74 years of age) but age did not interact with other manipulated variables in the lineup identification responses of the eyewitnesses. High rates of false identification have been observed as well in studies in which subject–witnesses believed at the time of their identification decision that the crime was real and that there would be consequences to the accused (Luus, 1991; Malpass & Devine, 1980; Murray & Wells, 1982).

In short, there is little or no evidence that the typical eyewitness experiment presents a distortion of what would be expected in actual cases in which the eyewitnesses experience real rather than simulated events. This does not mean that all eyewitness experiments can be readily generalized to actual cases. However, as Bass and Firestone (1980) noted, the psychological attributes of the research may affect the generalizability of the results far more than the study’s objective features, such as demographic characteristics of the sample or the physical nature of the setting (see also Berkowitz & Donnerstein, 1982). If subjects believe that they are witnessing a real crime, for instance, in what important way are they different from people who witness a real crime?

The Function of Lineups

A lineup can be construed as a method for establishing proof about the identity of a suspect or, more precisely, proof that the suspect and the culprit are or are not the same person. Because the identification process is imperfect, however, the actual function served by lineups must be construed in a way that reflects the uncertainty of the proof.

Two general approaches to describing the uncertainty-reduction function of lineups have been advocated. One is to depict the outcome of a lineup task (identification of the suspect or not) as an event that increases or decreases the subjective probability that the suspect is the culprit relative to the probability before the lineup (see Wells & Lindsay, 1980; Wells & Luus, 1990a). This approach lends itself nicely to Bayesian logic and has been used with actual data sets to show the ways that diagnosticity and information gain vary as a function of the
Table 1
Possible Outcomes With the Single-Suspect Lineup

<table>
<thead>
<tr>
<th>Lineup status</th>
<th>Witness identifies</th>
<th></th>
<th></th>
<th>Witness makes no identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Culprit present</td>
<td>Correct (hit)</td>
<td>Cannot occur</td>
<td>Error (distractor identification)</td>
<td>Error (incorrect rejection)</td>
</tr>
<tr>
<td>Culprit absent</td>
<td>Cannot occur</td>
<td>Error (false identification)</td>
<td>Error (distractor identification)</td>
<td>Correct (correct rejection)</td>
</tr>
</tbody>
</table>

eyewitness’s behavior (e.g., identification versus nonidentification, see Wells & Lindsay, 1980), the lineup model being used (e.g., Wells & Turtel, 1986), similarity between distractors (e.g., Lindsay & Wells, 1980), and the separation of lineup from nonlinear evidence (Wells & Luus, 1990a).

A somewhat different approach to describing the uncertainty-reduction function of lineups has been offered by Navon (1990). Navon proposed that the outcome of a lineup task informs us about resemblance between the suspect and the culprit. Even when the identification made by an eyewitness is a false alarm, he pointed out, we have learned that the culprit apparently has physiological characteristics similar to the identified person. Hence, Navon argued, all identification behaviors serve to reduce uncertainty, even false identifications.

Luus and Wells (1991) proposed that this reduction of uncertainty could be understood in the context of a distinction between recall and recognition. Prior to an attempt to make an identification (a recognition memory task), the eyewitness has already given a verbal description of the culprit (a recall task). On occasion, description or verbal recall can be highly diagnostic of a specific individual, such as if an eyewitness says that the culprit had a three-centimeter scar under his left eye, a gold front tooth, and a birth mark on his neck. More commonly, however, a description is much less diagnostic of identity; an eyewitness might describe the culprit as a Caucasian male of average height and weight, blonde hair, and no facial hair. The former (diagnostic) verbal recall would allow one to eliminate almost any innocent suspect and, if a suspect were found who matched the description, it would produce a high level of certainty that the suspect and culprit are the same person. The latter (less diagnostic) description, however, represents a much different situation. Although this description is sufficient to rule out large portions of the population, a mere match between a suspect’s characteristic and the verbal recall description is poor evidence of identity. Under these conditions, the ability of the eyewitness to identify the suspect when the suspect is embedded among distractors of a similar description can be highly informative. The function of a lineup, then, is to learn something from the eyewitness’s recognition memory that the eyewitness was not able to articulate in verbal recall (Luus & Wells, 1991; Wells & Luus, 1990a).

The Structure of Lineups
It is important to understand the basic structural properties of lineups to see how errors occur, how errors are distributed, and how some errors of identification are more harmful than others. The most basic structural property concerns the presence versus absence of the actual target or culprit in the lineup. In some unknown percentage of real cases, police have failed to arrest the actual culprit and have presented the eyewitness with a lineup that contains only innocent people. In other (I hope most) cases, the police have put together a lineup that contains the actual culprit. In most eyewitness identification experiments, this presence–absence structure is a factor in the design of the experiment. Without the inclusion of a presence–absence factor in an eyewitness experiment, it is not possible to adequately assess eyewitness identification accuracy. An experimental manipulation that increases hit rates in a culprit-present lineup, for example, might increase false identification rates in a culprit-absent lineup.

The culprit-present versus culprit-absent distinction is a fundamental structural property for lineups just as signal-plus-noise versus noise-alone trials are fundamental to the structure of Signal Detection Theory (Coombs, Dawes, & Tversky, 1970; Egan & Clarke, 1966; Green & Swets, 1966). There are other structural features of lineups, however, that are somewhat unique to the lineup task. It is important, for example, to distinguish between identifications of distractors and identifications of suspects and to distinguish between single-suspect models and all-suspect models (Wells & Turtel, 1986).

Single-Suspect Model
The single-suspect lineup model is depicted in Table 1. In the single-suspect lineup, there is one suspect and the remaining lineup members are known innocents, distractors, foils, or fillers. The important point about these distractors is that they are not suspected of the crime in question and, if an eyewitness were to identify one of these distractors, the identification would be dismissed by police as an error by the eyewitness. ² Indeed, it has

² Logically and empirically, the identification of a distractor should lower the probability that the suspect is the culprit when compared with the probability that existed before the nonidentification response (Wells & Lindsay, 1980). People’s reactions to nonidentifications of the suspect, however, do not closely follow this normative strategy (Leippe, 1985).
been argued that the single-suspect lineup is both safer and more informative than other models because of the single-suspect model's ability to relegate at least some portion of mistaken identifications into the category of discernible and relatively harmless error (Wells & Turtle, 1986). Hence, the identification of a distractor is not called a false identification; the term false identification is reserved for cases in which a witness has identified an innocent suspect rather than a distractor, thereby preserving the forensic meaning of the term false identification (Lindsay & Wells, 1980).

In a single-suspect lineup, false identifications (in the forensic sense described above) cannot occur when the lineup contains the culprit. Instead, there are only three possible outcomes from a culprit-present lineup using the single-suspect model—a hit, a distractor identification, and an incorrect rejection. In a culprit-absent lineup, the three possible outcomes are a false identification, a distractor identification, and a correct rejection.

Some of the early eyewitness identification experiments counted all mistaken identifications as false identifications rather than just counting identifications of the suspect in the culprit-absent lineup (see Table 1). Assuming a single-suspect lineup, such a counting strategy serves to exaggerate the false identification rate from a forensic perspective because it includes distractor identifications.

**All-Suspect Model**

The other general model for lineups and photospreads is the all-suspect model. An all-suspect lineup is one that has no distractors; instead, each member of the lineup is suspected by the lineup operators of being the perpetrator of the crime. The all-suspect lineup is depicted in Table 2 and, as can be seen by comparing Table 2 with Table 1, the all-suspect lineup has a quite different set of possible outcomes than the single-suspect lineup. Of the eight matrix entries for the single-suspect lineup in Table 1, two cannot occur, namely the identification of an innocent suspect when the perpetrator is present and the identification of the perpetrator when the perpetrator is absent. With the all-suspect lineup, however, there are three outcomes that cannot occur—the identification of the perpetrator or of a distractor when the perpetrator is absent and the identification of a distractor when the culprit is present. There is an outcome that can occur with the all-suspect lineup that cannot occur with the single-suspect lineup, namely the identification of an innocent suspect when the culprit is present.

The implications of using a single-suspect model versus an all-suspect model are considerable. Mathematical analyses show that the likelihood of a false identification (in the forensic sense defined earlier) is much higher with the all-suspect lineup than with the single-suspect lineup (Wells & Turtle, 1986). This is primarily because there is no opportunity for the all-suspect lineup to reveal eyewitness error whereas the single-suspect lineup has a known-error category (i.e., any distractor identification). This difference in the likelihood of false identification is applicable to lineupwise error rates (which are cumulative across suspects and analogous to experimentwise error rates in research), but the likelihood that any given innocent suspect will be falsely identified (analogous to per-comparison error rates in experiments) is largely unaffected by the model being used.

Although field data are limited, indications are that perhaps one-fourth of the lineups used by police in the U. S. Midwest are all-suspect lineups and about that same proportion are mixed models (i.e., more than one suspect but also some distractors, see Wells & Turtle, 1986). The all-suspect model has been likened to a multiple-choice test for which there could be no wrong answer; any identification can be construed by police as a hit (Wells & Turtle, 1986). The advantage of the single-suspect model, on the other hand, is that it provides a grading key of sorts in that the identification of a distractor can be readily dismissed as a wrong answer. An all-suspect lineup might be acceptable in multiple-witness cases when the purpose of presenting an all-suspect lineup to one eyewitness is to reduce the number of suspects to one; the remaining eyewitnesses can then view a single-suspect lineup. Throughout the remainder of this article it will be assumed that the single-suspect model is being used.

**Lineups-as-Experiments Analogy**

One of the guiding heuristics for exploring possible improvements to lineup methods is the analogy between conducting a lineup and conducting a psychology experiment (Wells & Luus, 1990b). The analogy is relatively rich. Consider the parallelisms between a police lineup

---

3 At this point I am still dealing with situations in which there is only one perpetrator. The term all-suspect lineup does not refer to a lineup in which there are multiple perpetrators; rather, it means that each lineup member is suspected of being the single perpetrator in question.
and an experiment: Police have a hypothesis (i.e., that the suspect is the culprit); a design is created (e.g., place the suspect in the fourth position among seven other people); questions are asked (e.g., Do you see the culprit among those people?); responses are recorded (e.g., yes or no); and the original hypothesis is reevaluated in light of the response data.

When considered in light of appropriate experimental methodology, actual police lineups typically lack the kinds of controls that are essential to making clear inferences from data. An obvious example of this is sample size. With actual criminal cases, inferences are based on small samples, often only one eyewitness’s response to a lineup. From the perspective of a psychological experiment, a sample of one is meaningless. Unlike a psychology experiment, however, police cannot simply decide to boost their sample size to provide a more reliable test of their hypothesis that the suspect and the culprit are the same person. There are, however, other features of a good experiment that can be incorporated into lineup tasks as forms of control.

**Blank lineup control.** One such control has been proposed whereby an eyewitness is first presented with a blank lineup—a lure or screen of sorts (Wells, 1984b). A blank lineup contains no suspect; all members are known-innocent distractors. The eyewitness does not realize that there is a second lineup, and the critical measure is whether the eyewitness identifies someone from this blank lineup. This is analogous to creating a noise-alone trial in a signal detection task. Those who get beyond the lure without making an identification and then go on to identify someone from the real lineup that follows are much more likely to have made an accurate identification than those who go directly to the real lineup (Wells, 1984b).

**Mock witness control.** Although police cannot increase their sample size of eyewitnesses (and often must use a sample of one to test their hypothesis that the suspect is the culprit) it is possible to use a control sample of mock witnesses to which the actual eyewitness’s response can be compared. Mock witnesses were first proposed by Doob and Kirshenbaum (1973) and the procedures and numerical indices for mock-witness data have been refined by others (e.g., Brigham, Ready, & Spier, 1990; Malpass, 1981; Wells, Leippe, & Ostrom, 1979). The general idea is to show the lineup (or a photographic representation of the lineup) to a set of people who are given only limited information about the crime. If these mock witnesses tend to choose the suspect as the lineup member that they believe is the culprit, then it raises the possibility that the actual witness’s identification was based on the same limited information. In other words, mock eyewitnesses, who have never actually seen the culprit, ought to spread their guesses in a relatively equal frequency across the lineup members. Indeed, there usually is (or should be) a photographic record of a lineup in actual cases; hence this is a common method by which many psychologists have been helpful to the courts in providing empirical evidence that can help clarify the meaning of a given identification. Caution is required in using this method, however, because it is important to know what information the mock witnesses should and should not receive prior to their guesses about which lineup member is the suspect. This is discussed in greater detail in the later sections on functional size and propitious heterogeneity. The general point, however, is that it is possible to generate a large sample of mock witnesses that serve, in effect, as a sort of control condition.

The analogy between a properly conducted police lineup and a properly conducted psychology experiment gives credence to the argument that any well-trained experimental psychologist could evaluate the adequacy of procedures used by police in securing an eyewitness identification. Indeed, the concepts of experimental demand (e.g., Rosenthal, 1976), response bias, confoundings, double-blind procedures, and so on (see also Orne, 1962) can be readily transferred from the research lab to the police station.

However, not all knowledge of lineup identification flows directly from an understanding of experimental research methods. As discussed in the previous section on the structure of lineups, for example, it is necessary to know something about the system of lineups, such as the ways that different lineup models map onto various outcome possibilities (e.g., three different types of false alarms). As is apparent in the next section, a theoretical understanding of the identification process that is somewhat unique to lineup identifications is required to fully appreciate some of the suggested improvements to the lineup procedure.

**The Identification Process**

Empirical evidence suggests that eyewitness identification of faces is a holistic process rather than a feature-based or particularistic process. For example, when subjects are required to break down a face into its component features, such as nose, eyes, mouth, and so on, while encoding the face, their subsequent ability to identify that face among a set of distractors is severely impaired relative to those who make global judgments (e.g., judging the honesty of the face; see Bower & Karlin, 1974; Warrington & Akroyd, 1975; Wells & Hryciw, 1984; Wells & Turtle, 1988; Winograd, 1976).

What is meant by holistic processing rather than processing based on individual features of faces is not entirely clear. The general idea is that faces are processed not just as sets of separate features but as interactive systems of features that include interfacial properties such as distance between features, relative sizes, and other topographical types of information (Wells & Hryciw, 1984). The same nose on two different faces can appear to be a quite different nose; changes in hair style can make chin lines or lips appear to change; and a simple featural change (e.g., loss of a moustache) can lead a person to notice that a face now looks different, but the person cannot necessarily specify what feature has changed. All of these observations suggest that face recognition represents something more than the sum of its features, that there is a holistic level of some sort that strongly influences the
face recognition process (e.g., Wells & Turtle, 1988). The process might be different for young children, who seem to process faces in a more piecemeal or feature-extraction manner than do adults (Carey & Diamond, 1977; Diamond & Carey, 1977).

The question of whether faces are processed holistically or piecemeal is somewhat related to the question of whether faces are processed in parallel or serially. Some researchers have argued against the parallel processing model for faces. Bradshaw and Wallace (1971), for example, showed that it takes people longer to decide if two faces are different if the two faces differ on few (e.g., two) features than if they differ on several (e.g., seven). E. E. Smith and Nielsen (1970), however, argued that “sameness” judgments (in contrast to judging whether two faces are “different”) are the result of parallel processing because the latency of such judgments does not vary as a function of the number of features to be processed. Townsend (1971), Orenstein and Hamilton (1977), and others, however, have argued that it is usually impossible to distinguish unequivocally between serial and parallel processing with latency measures. Matthews (1978) argued that the perception of faces involves a combination of serial and parallel processing wherein outer features are processed in parallel whereas inner features are scanned sequentially. Ellis (1981) has argued that most of these studies are too far removed from eyewitness identification situations to be highly informative about the eyewitness identification process (e.g., the faces are artificial composites and the unique tasks presented to these subjects can produce artificial strategies suited to the particular situation).

Although the questions of holistic versus feature processing and parallel versus serial processing of faces are related somewhat, it does not necessarily follow that holistic processing implies a parallel process. The idea that faces are processed holistically simply means that features interact such that they might not be recognized individually or might appear different when some other feature is mutated. Both a serial and parallel processing model can explain the holistic processing view.

There are relatively consistent data indicating that the upper portion of faces, especially hair and eyes, are more important to recognition than the inner features of faces (e.g., Davies, Ellis, & Shepherd, 1981; Goldstein & Mackenberg, 1966; Matthews, 1978). Intuitively, we might expect that such findings, which come from face fragmentation studies and feature alteration studies, would also surface in analyses of eye fixation patterns. In fact, however, although hair seems to be the most reliable fragment to recognition and causes the most disruption when altered, eye fixation patterns show that noses and mouths are more heavily looked at when people process faces (e.g., Cook, 1978; Walker-Smith, Gale, & Findlay, 1977). Shepherd et al. (1982) have concluded that fragmentation, alteration, eye fixation, and other studies of feature saliency do not resolve the question of whether face recognition is holistic or piecemeal because the design and premises used in such studies impose a piecemeal strategy on subjects.

Some faces are more easily recognized than are other faces and the dimension of distinctiveness or typicality is usually implicated as a critical mediating variable. When faces are rated for typicality, those that are rated as atypical rather than typical surface as the faces that are more easily recognized (e.g., Courtis & Mueller, 1981; Light, Hollander, & Kayra-Stuart, 1981; Light, Kayra-Stuart, & Hollander, 1979). Along these lines, it has been suggested that a good line drawing that accentuates the distinctive features of the individual, a face caricature, can capture the person better than can a spontaneous photograph. Consistent with this view, Ryan and Schwartz (1956) found that cartoon drawings were correctly interpreted at shorter tachistoscopic exposures than were actual photographs. Rhodes, Brennan, and Carey (1987) found that caricatures were more easily recognized than were line drawings or anticaricatures. Other studies, however, have failed to find evidence that drawings are superior to photographs (e.g., G. Loftus & Bell, 1975) and the caricature superiority hypothesis for faces seems not to have clear support (see Hagen & Perkins, 1983; Perkins, 1975; Tversky & Baratz, 1985).

Some researchers have argued that humans’ ability to perceive and store faces is so unusually good that there must be a special processing mechanism for the task (e.g., Geschwind, 1979; Whitely & Warrington, 1977; Yin, 1969). In fact, however, there is uncertainty about this argument (see Ellis, 1975). One of the strongest arguments for the special processing mechanism view has been the research on patients suffering from prosopagnosia, a disorder ostensibly characterized by a specific inability to recognize faces. Meadows (1974) and Ellis (1975), however, have found that this disability does not show specificity to faces. Bornstein (1963) showed that an ornithologist who developed an inability to recognize faces following brain surgery also could no longer identify different birds. A similar phenomenon was observed in a farmer who could no longer recognize his individual cattle (Bornstein, Sroka, & Munita, 1969). Ellis (1981) argued that prosopagnosia is most likely a more general deficit in the ability to make complex discriminations from memory. Although there might not be a special face processing mechanism, the social and survival significance of face recognition is clear; one must develop competence at an early age to distinguish between family members and outsiders, friends and strangers. The kind of face recognition ability that impresses some researchers, however, need not be unique to human faces. Bateson (1977) tested a subject who demonstrated an ability to individually identify each of 450 swans, which she had previously named, at a level of performance that was virtually perfect.

Shapiro and Penrod’s (1986) meta-analysis of face recognition studies shows that eyewitness identification experiments produce reliably lower hit rates and reliably higher false identification rates than do traditional face recognition studies. Indeed, lineups and photospreads
represent a somewhat different task for eyewitnesses than that often encountered by subjects in most face recognition experiments. Consider, for example, a relatively well known study by Bahrick, Bahrick, and Wittlinger (1975). Their work showed that people’s recognition memories for their high school classmates’ faces were highly accurate after 48 years. Several factors tend to make this study largely irrelevant to the eyewitness situation, however. One factor is that the recognition test was of people that the subjects actually knew at one point. Their original exposure to these faces was not a one-time event that occurred briefly. The presence of these people in classes, hallways, and dances, as well as having some exposure during earlier years, makes the situation quite different from what an eyewitness experiences. Reviews of one’s high school yearbook is a common pastime. Indeed, most of the recognition in Bahrick et al.’s study could be portrait recognition rather than person identification.

Portrait recognition occurs when a subject’s recognition of a face is tested by using the same photograph at test as was used for encoding. In portrait recognition, nothing changes from the original stimulus to the test stimulus. Hence, the subject is recognizing the particular portrait rather than the person. With portrait recognition tasks, facial expressions, lighting, and other ephemeral features are identical at test and acquisition. Person identification, however, is the recognition of someone who was viewed previously under a somewhat different set of transitory factors. Many of the studies showing strong face recognition performance are actually portrait recognition studies rather than person identification studies, and the stimulus stability (from encoding stimuli to test stimuli) in these studies is therefore extremely high.

In lineups and photospreads, this kind of stimulus stability does not exist. A person’s face changes daily, weekly, and monthly. Facial expressions, hair style and length, weight, tiredness versus alertness, clothing, acne, age lines, and so on are quite variable. Police do not have the luxury of freezing the appearance that the culprit had at the time of the witnessed event in order to help the witness at the time of the lineup or photospread test. Photographs are commonly outdated. Delays of months are common between the witnessed event and the lineup test. Indeed, given the considerable variations in the actual appearance of people over time, it is surprising that people can perform the kind of perceptual calculus that is required to perceive stability under such dynamic changes. This is one of the problems that has plagued the visual robotics industry; although it is not difficult to get a robot to recognize the exact visual stimulus that was presented previously, it is very difficult to get it to recognize that stimulus under a set of dynamic but irrelevant changes without also promoting false alarms to other stimuli.

There are many other factors that can make the eyewitness identification experience importantly different from the traditional face recognition experiment. It is not uncommon, for example, for actual eyewitness cases to involve cross-racial identifications (see Bothwell, Brigham, & Malpass, 1989). In addition, stress at encoding and during the identification task, as well as the distracting or arousing effects of weapons, can have considerable effects on the reliability of identification (see Cutler, Penrod, & Martens, 1987a; Cutler, Penrod, O’Rourke, & Martens, 1986; Dent, 1977; Johnson & Scott, 1976; Kramer, Buckhout, & Eugino, 1990; E. F. Loftus, Loftus, & Messo, 1987; Maas & Kohnken, 1989; Tooley, Brigham, Maas, & Bothwell, 1987; Steblay, in press).

Relative-Judgment Process

Despite an incomplete understanding of the cognitive processes involved in person identification, it appears that the concept of relative judgments will need to be a part of any adequate theory of eyewitness identification (Wells, 1984b). In the context of lineups, a relative-judgment process is one in which the eyewitness chooses the lineup member who most resembles the culprit relative to the other members of the lineup. This is not a surprising proposition and would seem to be both rational and functionally appropriate on initial consideration. But the relative-judgment process can also be dysfunctional.

To understand how the relative-judgment process can be dysfunctional, one need only readdress the earlier discussion of the structure of lineups, with particular reference to culprit-absent lineups. Choosing the lineup member who best resembles one’s memory of the culprit is a useful and unflawed strategy as long as the actual culprit is present in the lineup. But what happens when the culprit is not present in the lineup? Unfortunately, the relative-judgment process will still produce an affirmative answer even in the absence of the actual culprit because there remains one lineup member who resembles the culprit better than do the other lineup members. Hence, the relative-judgment process is seductive yet dangerous.

Undoubtedly, some of the allure of the relative-judgment process owes to the failure of the eyewitness to fully entertain the possibility that the culprit might not be among the set of alternatives. Indeed, the mere act of telling eyewitnesses that the actual culprit might or might not be present in the lineup serves to diminish the rate of choosing in culprit-absent lineups with little or no damage done to hit rates in culprit-present lineups (Malpass & Devine, 1981a). There are both theoretical and empirical reasons to believe that the relative-judgment process continues to exert a role even when the “might or might not be present” instructions are made salient (Gonzalez, Ellsworth, & Pembrooke, 1990; Wells, 1984b).

The relative-judgment process does not deny the role played by ephoric similarity. Ephoric similarity is a similarity judgment between a memory trace and a physical stimulus (Tulving, 1981). Relative judgments assume that an ephoric-similarity process is operative, otherwise the eyewitness could not decide who most looks like the culprit. However, the relative-judgment process assumes that the threshold for an affirmative response (i.e., for making a positive identification) is governed at least in part by the mere existence of greater ephoric similarity for one lineup member than there is for the other lineup.
members. The relative-judgment concept implies that, as long as one lineup member shows greater ephoritic similarity than do the others, there is a propensity to positively identify that lineup member even if the absolute level of ephoritic similarity is only modest.

Perhaps the simplest demonstration of the use of relative judgments can be shown by observing rates of choosing under conditions in which a target is removed without replacement from a set of distractors. Such data are shown in Table 3. The important phenomenon to observe in these data is what happens to choices of other lineup members when the actual target is removed. If subject—eyewitnesses choose the target when the target is present but then shift to another lineup member when the target is removed, then there must be some relative-judgment processes operating. The data in Table 3 were obtained from just such a study. In this study 100 eyewitnesses viewed a lineup containing a staged-crime culprit and 100 viewed the exact lineup with the culprit removed. All were told that the culprit might or might not be present. What would the 54% who identified the target in the target-present lineup have done if the target was removed without replacement? Some (21%) would have identified the lineup member who was the second-best choice in the target-present lineup and who apparently became the best choice in the target-removed lineup.

The pattern of data in Table 3 is consistent with what a relative-judgment process would predict. The actual target should be the lineup member who best resembles the culprit because the target is the culprit; but if the target is removed, someone else in the lineup becomes the one who best resembles the culprit. At the same time, however, the data in Table 3 indicate that the relative-judgment process is not totally dominant. If the process were fully one of relative judgments, the rate of no-choice responses would be equal between target-present and target-absent lineups, whereas both these data and other data show that no-choice rates are consistently higher when the target is absent rather than present (e.g., Lindsay & Wells, 1980, 1985; Lindsay, Wells, & O'Connor, 1989; E. F. Loftus, 1976; Malpass & Devine, 1981a; Pigott, Brigham, & Bothwell, 1990; Wells, 1984a; Wells & Lindsay, 1980; Wells & Turtale, 1986; also see Shapiro & Penrod, 1986, for a metaanalysis on this variable). Hence, in addition to being governed by a relative-judgment process, there are other, threshold-like processes governing the identification decision.

Among the practical consequences of understanding that eyewitness identification is influenced by a relative-judgment process has been the development of sequential lineups. Using the notion of relative judgments outlined by Wells (1984a), Lindsay and Wells (1985) reasoned that a sequential decision task would force the eyewitness to abandon relative judgment strategies and instead make decisions based more purely on threshold-like criteria.

**Sequential Lineups**

A sequential lineup is a procedure in which each lineup member is presented to the eyewitness individually in sequence. The eyewitness must decide at the time of each initial presentation whether that lineup member is the culprit. A simultaneous lineup, which is traditionally used by police, presents the eyewitnesses with all lineup members at once. Borrowing on the idea that eyewitnesses are prone to making relative judgments (Wells, 1984a), Lindsay and Wells (1985) reasoned that a sequential lineup would largely nullify the ability of eyewitnesses to use a relative judgment strategy. Specifically, although an eyewitness could reason that a given lineup member (e.g., Number 3) was a relatively better match to the culprit than was a previously presented member (i.e., better than either Number 1 or Number 2), the witness could not be certain that a subsequent lineup member (yet to be viewed) would not prove to be an even better match to the culprit than the one being currently viewed. As a result, the eyewitness is forced to abandon the relative judgment strategy and use a more "absolute" strategy when confronted with a sequential presentation procedure.

In the original experiment testing the sequential versus simultaneous lineup procedures, Lindsay and Wells (1985) did not tell the sequential procedure witnesses about the number of lineup members they would eventually view. The results showed that the sequential procedure produced fewer false identifications in culprit-absent lineups with no significant reduction in accurate identifications for culprit-present lineups. Cutler and Penrod (1988) subsequently replicated the sequential-superiority effect under conditions in which eyewitnesses knew the number of lineup members that they were going to view. In a direct comparison of conditions in which eyewitnesses were either told or not told how many lineup members they were going to view in a sequential procedure, Lindsay, Lea, and Fulford (1991) showed that the

---

**Table 3**

Rates of Choosing Lineup Members When a Target is Present Versus Removed Without Replacement

<table>
<thead>
<tr>
<th>Lineup member</th>
<th>1</th>
<th>2</th>
<th>3 (target)</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>No choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target present</td>
<td>3%</td>
<td>13%</td>
<td>54%</td>
<td>3%</td>
<td>3%</td>
<td>3%</td>
<td>21%</td>
</tr>
<tr>
<td>Target removed (without replacement)</td>
<td>6%</td>
<td>38%</td>
<td>—</td>
<td>12%</td>
<td>7%</td>
<td>5%</td>
<td>32%</td>
</tr>
</tbody>
</table>

May 1993 • American Psychologist
sequential procedure was superior to the simultaneous procedure regardless of the witnesses’ assumptions about the number of lineup members to be viewed. However, the sequential procedure worked significantly better if the eyewitnesses were not aware of the number of lineup members to be viewed. Presumably, the eyewitnesses who knew that they were close to the end of the sequence of options tended to fall back on making some guesses about the likelihood that the one or two remaining lineup members would show a greater relative resemblance to the culprit than the one being viewed at that moment.

The patterns observed in these comparisons of sequential versus simultaneous lineup procedures, along with the added benefits that derive from keeping eyewitnesses blind about the number of lineup members to be viewed, represent further support for the idea that false identifications owe at least in part to the relative-judgment process. There are two further empirical observations concerning sequential presentations that support the idea that false identifications from lineups are at least partly governed by relative-judgment processes. First, a recent experiment by Lindsay, Lea, Nosworthy, Fulford, Hector, LeVan, and Seabrook (1991) showed that sequential lineup procedures serve to minimize the influence of two critical lineup biases, namely low functional size (see following section) and failure to warn eyewitnesses that the culprit might not be in the lineup. This is an important observation from a theoretical perspective because these two lineup biases are themselves presumed to be problematic only if the eyewitnesses are using a relative-judgment strategy. Thus, it is logical that a procedure that inhibits relative judgments should also tend to cancel the effects of these two lineup biases.

If lineup techniques that force eyewitnesses to use something other than a relative-judgment strategy tend to produce better decisions by eyewitnesses, then we might also predict that a show-up procedure would be better than a simultaneous lineup procedure. A show-up is the presentation of one and only one suspect to the eyewitness; in effect, a one-member lineup. Such a prediction seems on surface to be incongruous with the idea of a good identification procedure, and yet, a show-up totally prevents the relative-judgment process from being used by eyewitnesses and therefore might not be a poor procedure. Using this logic, Gonzalez et al. (1990) examined hit rates and false identification rates using show-ups versus lineups. The results showed that eyewitnesses were more likely to identify someone from a lineup than from a show-up and that this effect held even when the culprit was absent from the lineup or show-up. It appears that the presence of distractors (in lineups) seduces witnesses into making a positive identification of an innocent person whom they would not have identified if that person was shown to them without distractors. The presence of distractors apparently served as comparisons and the relative-judgment process served to trigger a positive, albeit mistaken, identification. Gonzalez et al.’s study suggests that a lineup with a functional size of 1.0 (see next section) is probably more dangerous for the innocent suspect than is a show-up, even though courts of law and psychologists have routinely argued that show-ups are the most biased of all identification procedures (Gilbert v. California, 1967; Levine & Tapp, 1973; Malpass & Devine, 1983; Sobel, 1985; Stovall v. Denno, 1967; U.S. v. Wade, 1967; Wells, et al., 1979). Gonzalez et al.’s results should not be interpreted as a green light for the use of show-ups, however. As demonstrated in a recent study contrasting lineups with show-ups, a good lineup is better than a show-up because false positive responses can be distributed to nonsuspect distractors (Wagenaar, 1992).

**Functional Size**

The concept of functional lineup size was developed in the late 1970s to refer to the number of viable lineup members, as opposed to nominal size, which refers merely to the number of people in the lineup (Wells et al., 1979). In the extreme, the distinction between functional and nominal lineup sizes is obvious, such as when the witness describes a Black perpetrator and a Black suspect is placed in a lineup with five other people who are White. In this rather absurd example, the nominal size is six, but it is functionally a lineup of only one person.

Under some set of assumptions about human behavior, one might not care who the distractors are or whether they resemble the description that the witness gave to the police. It might be argued, for example, that as long as the witness is told that the perpetrator might not be in the lineup, the witness can still make a sound judgment as to whether the Black member of the lineup is the person he or she saw commit the offense. In general, however, such an assumption would conflict with psychology’s theoretical models of social cognition and social influence. It is known that people will base decisions on inferences and that conformity, obedience, and compliance pressures can be especially strong phenomena in situations in which ambiguity and authority are prominent (e.g., Cialdini, 1988). Applying these theoretical ideas to the low functional size lineup, one can hypothesize that the witness can infer that the police suspect the only person who matches the description that the witness had previously provided. Hence, mere deduction tells the witness which lineup member the police expect him or her to choose; actual recognition is not needed to make this deduction. Because the police would be unlikely to exert the effort, time, and cost of conducting the lineup unless they had good reasons (e.g., other evidence) to do so, a witness might infer that the suspect is probably the perpetrator. Not identifying the suspect under these circumstances would be a rather uncooperative response and harmful to the investigators’ case against the suspect. If, in addition, the witness has a somewhat vague memory of the perpetrator so that clear rejection seems implausible, the inference process and social pressures can prevail to yield a positive identification based on something other than recognition memory.

Although the functional size issue is obvious in the Black-suspect–White-distractors example, it is more likely
to surface in a less obvious (although perhaps no less robust) form in actual cases. Suppose, for example, a witness describes the perpetrator as about six feet tall, male, White, with no facial hair, and about 160 pounds. Suppose further that the suspect matching this five-feature description is placed in a lineup in which the distractors each match four of the five description features but do not match the other descriptor. Thus, one distractor matches each descriptor except he is only five feet six inches tall, another matches each descriptor but weighs 200 pounds, and so on. Because the witness knows the description that he or she gave to the police, he or she can perhaps safely infer that those who violate that description are not actually suspected by the police and deduce that the police suspect only the person who does match the description. In this case, as much as in the Black-suspect–White-distractor lineup, the functional size is one.

There is yet another reason why a low functional size can promote false identifications. As discussed at earlier points in this article, it is assumed that eyewitnesses are prone to using a relative-judgment process (Wells, 1984a). Hence, even an innocent suspect who does not highly resemble the actual culprit at some absolute level of ephorich similarity might resemble the culprit relatively more than do the other lineup members and thereby elicit a feeling of recognition in the mind of the eyewitness. The danger of an innocent suspect showing greater relative similarity to the culprit than do the distractors is higher when functional sizes are low rather than high.

Experimental evidence supports the view that low functional sizes promote false identifications when the actual target is absent from the lineup (e.g., Lindsay & Wells, 1980). In other words, an innocent person who merely matches the description of the culprit will receive a much higher rate of false identifications in the context of distractors who fail to match the description (low functional size) than in the context of distractors who do match that description (high functional size). The ability to identify the actual culprit in a culprit-present lineup is not significantly harmed by increasing functional size (Lindsay & Wells, 1980).

Although there is some disagreement over the best way to construct a numerical index of functional size, lineup researchers agree that mock witnesses can provide the essential data for calculating the functional size of a lineup (Buckhout, Rabinowitz, Alfonso, Kanellis, & Anderson, 1988; Doob & Kirshenbaum, 1973; Malpass, 1981; Malpass & Devine, 1983; Wells et al. 1979). Mock witnesses are people who are given the eyewitness’s description of the culprit, shown the lineup, and asked to choose the person that they think the police suspect of the offense. If mock witnesses can use this limited information to infer which lineup member is suspected by the police, then one cannot assume that the eyewitness himself or herself used anything beyond such scanty information to make an identification of the suspect.

The first and most widely used index of functional size was proposed by Wells et al. (1979). The index of functional size is simply $N/NS$ where $NS$ is the number of mock witnesses who choose the suspect and $N$ is the total number of mock witnesses. Hence, if every mock witness chooses the suspect, functional size is 1.0; if half choose the suspect, functional size is 2.0; and if 20% choose the suspect, functional size is 5.0.

Malpass and Devine (1983) proposed other numerical indexes from mock witness data to express lineup bias and defendant bias. The indexes proposed by Wells et al. (1979) versus those of Malpass and Devine tend to produce similar conclusions in most cases, but can yield different solutions in some cases (see Brigham & Brandt, 1992; Brigham et al., 1990). Regardless of the index proposed, however, there is no disagreement that the distractors should match the description that the witness gave of the perpetrator. As discussed in the next section, however, matching distractors to the witness’s prelineup description should not be confused with the idea of selecting distractors who resemble the suspect.

**Propitious Heterogeneity**

Previous treatments of the issue of lineup bias have led to some confusion regarding how to select distractors for lineups and what constitutes a biased lineup (Luus & Wells, 1991). Some police scientists, and some eyewitness researchers, have come to think that distractors should be selected on the basis of their resemblance to the suspect. Indeed, some readers of this article might think that the previous section on functional size advocates such a selection strategy. If so, then it is particularly important to note the critical distinction between a strategy of selecting distractors because of their resemblance to the suspect and selecting distractors because of their match to the witness’s description of the culprit. Hereinafter these are called the suspect–resemblance strategy and the culprit–description strategy, respectively.

Arguments have been launched for selecting distractors according to the culprit-description strategy and against the suspect-resemblance strategy (Luus & Wells, 1991; Wells, Seelau, Rydell, & Luus, in press). The general argument is that the culprit-description strategy offers the same level of protection for the innocent suspect that the culprit-description strategy offers and yet the culprit-description strategy is more likely to produce an accurate identification of the actual culprit than is the suspect-resemblance strategy. Specifically, as long as each member of the lineup fits the eyewitness’s prelineup description of the culprit, there is no reason to believe that the lineup is biased against the suspect; any further attempt to make the distractors look like the suspect is merely pursuing gratuitous similarity, serving no justifiable purpose. In fact, additional attempts to select the distractors who resemble the suspect (beyond the level of the eyewitness’s description of the culprit) can serve to confuse the eyewitness by eliminating the perceptual distinctions across the recognition test set that are necessary for the operation of recognition memory (Gibson, 1969).

Variation in the physical appearance of lineup members that aids the process of accurate recognition
while not biasing the lineup against an innocent suspect is referred to as *propitious* (helpful or beneficial) *heterogeneity* (Wells et al., in press). Such heterogeneity in the appearance of lineup members is importantly distinct from forms of heterogeneity that serve to reduce functional size. Although the strategy of selecting distractors who resemble the suspect and the strategy of selecting distractors who match the eyewitness's description of the culprit both serve the laudable purpose of ensuring adequate functional size, the suspect-resemblance strategy also serves the undesirable purpose of eliminating propitious heterogeneity.

The argument that the suspect-resemblance strategy for selecting distractors serves the undesirable purpose of eliminating propitious heterogeneity becomes readily apparent when one considers the "paradox of the clones" (Luus & Wells, 1991). If the goal in selecting distractors is to select people who resemble the suspect, then a lineup of the suspect's clones ought to be a truly superior lineup. This would be absurd, of course, but this hypothetical lineup of clones serves to illustrate the problem—deciding which of the lineup members' physical features should be common and which features should be variable across members of the lineup set.

Lineups are conducted on the assumption that there is information to be gained from a recognition (lineup identification) task that could not be obtained from a recall (description of culprit) task (Wells & Luus, 1990a). Hence, a proper set of distractors for the recognition task must control for recall, but allow nonrecallable features to vary across the recognition set (Luus & Wells, 1991; Wells et al., in press). Because eyewitness descriptions of culprits tend to be general, incomplete, and interval-like, distractors can look quite different from each other and from the suspect while matching the description that the eyewitness gave to the police. Experimental data reveal that subject—witnesses' physical descriptions of people tend to involve only six or seven features on average, tend to be general rather than specific (e.g., gender, race, height, weight, hair, age), and tend to include ranges or intervals (e.g., 170–190 pounds or early to mid 30s). These experimental data (e.g., see Pigott & Brigham, 1985; Pigott et al., 1990; Wells, 1985) are in close agreement with archival analyses of eyewitnesses' descriptions of culprits from actual criminal cases (see Sporer, 1992).

It is important to note that special circumstances can arise that do not permit one to totally ignore the physical characteristics of the suspect when selecting distractors. Suppose, for example, the suspect's appearance does not match the eyewitness's description; if the distractors match the description whereas the suspect does not, then the suspect might stand out as a distinctive member in the lineup. This problem would be revealed by conducting a mock witness test of functional size. Suppose that the eyewitness's prelineup description says nothing about height, but the suspect is six feet tall. As in the former example, the critical question is whether a lineup that uses only the prelineup description for selecting distractors produces a bias against the suspect.

Again, the question of whether this creates a bias against the suspect can be answered by conducting a mock witness test of functional size. The general point is that selecting distractors in a way that promotes propitious heterogeneity is permissible only to the extent that such selections do not reduce the functional size of the lineup. Other potential problems related to the pursuit of propitious heterogeneity are discussed in greater detail elsewhere (see Luus & Wells, 1991; Wells et al., in press).

Readers should note that the idea of selecting distractors in a manner that fosters both high functional size and high propitious heterogeneity might prove to be controversial. Although there are no published rebuttals to the propitious heterogeneity concept at this time, a 1992 symposium revealed that some eyewitness researchers have concerns about whether strategies for achieving propitious heterogeneity might create biases against the innocent suspect inadvertently (Swartz, 1992).

**False Certainty**

It can be argued that it is not false identifications per se that create a concern for the miscarriage of justice but rather the *certainty* with which such false identifications are sometimes made (Wells, Lindsay, & Ferguson, 1979). A number of eyewitness studies have examined the extent to which accurate identifications are associated with greater certainty by the identifier than are false identifications. The question of the relation between eyewitness certainty and eyewitness accuracy and an understanding of variables that govern the confidence of an eyewitness is of significant forensic import. There is a strong consensus in the empirical literature that the certainty expressed by an eyewitness regarding his or her identification decision is a robust determinant of the perceived credibility of the eyewitness (e.g., Leippe & Romanczyk, 1987, 1989; Leippe, Manion, & Romanczyk, 1992; Lindsay et al., 1981; Turtle & Wells, 1989; Wells et al., 1981; Wells & Murray, 1984). To the extent that eyewitness certainty is determined by variables that are unrelated to accuracy, eyewitness certainty represents a low validity cue to accuracy that is heavily used by jurors (see Leippe, 1980; Wells, 1984a; Wells & Lindsay, 1985). The results of studies directed at the certainty-accuracy relation are varied. In some experiments, false identifications are as-

---

4 For example, Brigham, Maas, Snyder, and Spaulding, 1982; Clifford and Hollin, 1981; Cutler, Penrod, and Martens, 1987a, 1987b; Cutler, Penrod, O'Rourke, and Martens, 1986; Fleet, Brigham, and Bothwell, 1987; Gorenstein and Ellsworth, 1980; Greenberg, Wilson, Ruback, and Mills, 1979; Hildengard and Irving, 1978; Hosch and Cooper, 1982; Hosch, Leippe, Marchioni, and Cooper, 1984; Jenkins and Davies, 1985; Kassin, 1985; Krafka and Penrod, 1985; Leippe, Wells, and Ostrom, 1978; Lindsay and Wells, 1985; Lindsay, Wells, and Rumpel, 1981; Malpass and Devine, 1981a, 1981b; Murray and Wells, 1982; Pigott and Brigham, 1985; Sanders and Warnick, 1980, 1981; Shepherd, Ellis, and Davies, 1982; V. L. Smith, Kassin, and Ellsworth, 1985; Wells, Ferguson, and Lindsay, 1981; Wells and Leippe, 1981; Wells, Lindsay, and Ferguson, 1979; Vuille and McEwan, 1985). Also, see Bothwell, Deffenbacher, and Brigham (1987) for a meta-analysis.
associated with somewhat less certainty on average than are accurate identifications, and in other experiments, the certainty of false and accurate identifiers is virtually equivalent. It is not my purpose in this article to attempt an estimate of the true magnitude of the mean differences in certainty between accurate and false identifications. It is instructive, however, to explore the question of why the certainty–accuracy relation in eyewitness identifications is apparently weak enough to yield controversy over whether such a relation exists at a level that has any practical utility.

There are two important differences between lineup identification tasks, which show little certainty–accuracy relation, and other tasks (such as signal detection) that seem to show consistent evidence of a certainty–accuracy relation. First, in an eyewitness identification experiment, each subject contributes only one data point for accuracy and one data point for certainty (i.e., the eyewitness is attempting to identify only one person). Having only one accuracy score and one certainty score per witness in eyewitness identification experiments is not just an arbitrary experimental design decision; it is a decision based on considerations of ecological validity. In actual cases, an eyewitness's certainty about an identification from a lineup is not compared with that same eyewitness's certainty in prior or subsequent lineup tasks. Being an eyewitness in an actual case is not a repeated measures affair as it might be in some memory experiments in which the subject is shown perhaps 25 or 50 stimuli at encoding and is tested for recognition memory using all or half of the original stimuli. Hence, unlike many other tasks in cognitive psychology, the certainty–accuracy relation is between subjects rather than within subjects and there are no repeated measures. This means, among other things, that individual differences between witnesses in their overall certainty cannot be separated from the individual response that they make (see V. L. Smith, Kassin, & Ellsworth, 1989). A witness who makes a false identification is compared with a different witness who made an accurate identification, rather than comparing a witness's false identification on one trial to that same witness's accurate identification on some earlier or later trial.

Second, the lineup identification situation is not a forced-choice test. If a witness is not certain enough to make a positive identification, then no identification is attempted. This free choice process has two effects. It serves first to restrict the range of certainty among those making accurate and false identifications by eliminating those who are relatively uncertain and, hence, deleting that portion of the sample who know that they don't know. Free choice also might serve to induce a dissonance (Festinger, 1957) or self-perception (Bem, 1972) process in the witnesses who make positive identifications that could elevate their certainty for reasons that are independent of their accuracy (Leippe, 1980). Were the lineup-identification situation a forced-choice task, those who know that they don't know would be forced to guess, would operate at close to chance, and would show little confidence. Of course, it would be unacceptable to suggest that lineups be made into forced-choice situations merely to produce a certainty–accuracy relation; the cost of a forced-choice task would be an increase in false identification rates.

The Malleability of Certainty

In an earlier section of this article (concerning the relative-judgment process), I used the idea of ephorism similarity to describe part of the process of lineup identification. A simple theory of eyewitness identification certainty would state that eyewitness certainty is a positive function of ephorism similarity judgments. That is, the greater the perceived similarity between the witness's memory trace (for the culprit) and the physical appearance of the identified person, the greater the certainty that the eyewitness will express in that identification.

Undoubtedly, there is considerable value to this ephorism-similarity source of eyewitness certainty. Nevertheless, it is equally clear that eyewitness identification certainty is influenced by factors that can be independent of the similarity between the witness's memory trace and the appearance of the identified person. In a robust demonstration of nonephoric influences on eyewitness certainty, Luus (1991) provided staged-crime eyewitnesses with randomly determined information about whether their counowitness to a theft had identified the same person they had identified from a lineup or had made a different decision. All witnesses believed they had witnessed an actual theft and, following their attempted identification of the thief, were told of the identification decision of their counowitness before they made statements about their identification to a police officer (who was also a confederate of the experimenter). Results showed that the certainty with which the witnesses delivered their statements to the police interviewer varied robustly as a function of whether they believed that their counowitness (a) identified the same person (M = 8.77 on a 10-point scale), (b) identified a different person (M = 4.67) or (c) had decided that the actual thief was not present (M = 3.57). It is important to note that these robust changes in the witnesses' expressed certainty in their identifications occurred after they had made an identification; hence, whatever level of ephoric similarity during identification existed in one condition also existed at that level in every other condition. As Luus (1991) pointed out, eyewitness identification certainty is the eyewitness's strength of belief about whether he or she was right or wrong in the identification; similarity between the witness's recollection of the physical appearance of the culprit and the physical appearance of the identified person is but one factor determining the strength of this belief. Other factors include extramemorial sources of social influence and heuristic inference strategies. If a witness makes a tentative identification and subsequently learns that the identified person was found in possession of the stolen goods, the witness's certainty in the identification is bound to inflate even though the witness's memory has not improved and the physical resemblance between the culprit and the identified person has not changed. Hence, data and logic
suggest that eyewitness identification certainty is not a simple product of ephoric similarity judgments. Eyewitness identification certainty is influenced by postidentification information, which probably involves a combination of normative and informational social pressures (a la Deutsch & Gerard, 1955).

A shift in eyewitness certainty as a function of postidentification information could be construed as a beneficial phenomenon to the extent that it improves the relation between accuracy and certainty. Kassin and his colleagues have provided some recent evidence that an increase in the confidence-accuracy relation for high public self-consciousness eyewitnesses can occur as a result of reviewing a videotape of themselves making their identification (Kassin, 1985; Kassin, Rigby, & Castillo, 1991).

**Base Rates and Error Rates**

Although this article is based in large part on the premise that there is an eyewitness identification problem, it is also based on the system-variable proposition; namely, that considerable control can be obtained that can minimize this problem. In this section, some numerical estimates are derived from a supposition that all of what is now known about proper ways to conduct lineups is applied to actual cases. The purpose of this section is to demonstrate the potential for holding down the rate of false identifications using a set of realistic assumptions about base rates in conjunction with optimal lineup procedures. It is not assumed that the error rates estimated in this section represent the rates of false identification being obtained in actual practice at this time.

Suppose that all of the best recommendations that eyewitness researchers can currently make regarding the conductance of a lineup were implemented. For example, suppose there were 1,000 independently conducted lineups, each using the single-suspect model, proper sets of instructions, distractors selected to match the lineup description of the culprit, a sequential presentation procedure, and so on. How many false identifications would be expected? Although there is no way to estimate the actual figure, there is an upper limit that can be specified under certain sets of assumptions.

Assuming optimal lineup procedures, the upper limit for false identification rates in actual cases is $1/N$ where $N$ represents the number of lineup members. This follows from the fact that mistaken identifications of distractors are not forensically false identifications. Assume for the moment that the suspect is innocent. Why should an eyewitness show any greater propensity to select the suspect than to select one of the distractors? As long as the distractors match the eyewitness's prior description of the culprit (and other proper procedures have been followed), the only reason for the witness to prefer the suspect is by mere chance. That is, by chance (or bad luck) the suspect might happen to resemble the actual culprit (or the witness's misperception) more than do the distractors. But this chance should be equally distributed across all lineup members over the hypothetical 1,000 independent cases. Hence, across 1,000 properly conducted lineups involving one suspect embedded among nine distractors, one would expect no more than 100 false identifications.

The situation would be different, of course, if there were multiple witnesses in each of the 1,000 cases. Suppose, for example, there were an average of two witnesses for each of the 1,000 independent cases. In general, the upper limit formula is $W/N$ where $W$ is the number of witnesses and $N$ is the number of lineup members. Across the 1,000 lineups, then, we would expect no more than 200 false identifications when there are two witnesses per lineup. Assuming independence of errors, of the 1,000 lineups there would be only 20 cases in which both eyewitnesses made false identifications; the remaining 180 lineups would involve one false identification and one distractor identification.

Returning to the situation of one eyewitness per lineup, it can be argued that the upper limit $1/N$ is further reduced by two additional factors, namely the probability of witnesses making no identification and the base rate for the culprit being present versus absent in the lineup. Consider first the base rate issue. The upper limit of $1/N$ assumes that the police have placed an innocent suspect in the lineup—that is, that in all 1,000 lineups the suspect is in fact not the culprit in question. Suppose one were to assume the opposite, that all of the 1,000 suspects are in fact the actual culprits. Under this assumption, the probability of a false identification is zero by definition. (Recall that, in a single-suspect lineup, forensically false identifications cannot occur if the suspect is the culprit; see Table 1.) Neither of these assumptions (i.e., that all or none of the suspects are the actual culprits) is realistic. It would be more realistic to assume that perhaps one half of these 1,000 single-suspect lineups include actual culprits and one half do not. Although the true base rate is indeterminant, it is clear that the likelihood of false identification is a function not only of the number of distractors but also of the base rate for suspects being culprits. If only 50% of the suspects across these 1,000 lineups are in fact the culprits, then the maximum number of false identifications would be 50 (or 10% × 50% × 1,000). If, as police might argue, 80% of the suspects are the actual culprits in question, then the false identification frequency drops to a mere 20 out of the 1,000 lineups, or a 2% false identification rate.

These rates of false identification are impressively low when one considers that they assume that the eyewitness has no recognition memory information at all and is merely guessing. This speaks to the power of a

---

2 This assumption of independence is probably not realistic. If one of the two eyewitnesses prefers the innocent suspect over the distractors, the other eyewitness might also show such a preference. In general, the chance factors that made the innocent suspect resemble the culprit more than do the distractors is likely to be replicated in the other witness. This is the main reason that police and courts should be cautious about making too much of the fact that two or more witnesses happen to identify the same person and why it is recommended that each witness view lineups with different distractors or, at least, with the lineup members appearing in a different order (Wells, 1988).
properly conducted lineup in terms of its ability to protect an innocent suspect. What happens if one assumes that eyewitnesses have a clear enough memory of the culprit to perform above chance on the lineup task? Consider, for example, a quite modest level of above-chance performance in which 20% of the eyewitnesses recognize the absence of the culprit in a culprit-absent lineup with the remaining 80% making an identification of someone from the lineup. That is, for 20% of the target-absent lineups the eyewitness responds with something to the effect that “The guy is not there” or “I can’t make a positive identification of any of these people.” Using the assumption that the lineup’s suspect is the culprit 50% of the time, there would be only 500 of the 1,000 lineups from which there could be a false identification, 100 of these 500 would produce no identification, and 360 of the remaining 500 would result in identifications of known-innocent distractors. The result is that only 40 of the 1,000 total suspects would be falsely identified. This 4% rate of false identifications would be further reduced to a mere 1% if the baserate for suspects being the actual culprits was 80%. Various combinations of base rates and choosing rates are depicted in Table 4 in terms of the maximum false identification associated with those conditions.

Previous analyses of the role of base rates in the eyewitness identification literature point to the importance of having “reasonable cause for suspicion” to justify placing a suspect in a lineup (Wells & Lindsay, 1980; Wells & Turtle, 1986). Imagine two police jurisdictions. In one, a suspect is placed in a lineup for possible identification only if the prelineup evidence indicates that it is more likely than not (i.e., greater than 50%) that the suspect is the culprit. The other jurisdiction, however, places a suspect in a lineup any time that there is a nonzero chance that the suspect is the culprit. Suppose, given these strategies, the former averaged around an 80% base rate for suspects being the actual culprits whereas the latter averaged 40%. The result is quite straightforward: False identifications will run twice the rate in the latter jurisdiction compared to the former jurisdiction over the long run. This difference in false identification rates will hold even if both jurisdictions are equally careful in their construction and conduct of the lineups and even though the eyewitnesses are equally competent or incompetent in these two jurisdictions. Perhaps it is time for the criminal justice system to require probable cause to place a suspect in a lineup just as it requires probable cause for other investigating behaviors (such as searches and seizures) that can place innocent persons in jeopardy. No matter how well an identification procedure is conducted, there remains some nonzero risk that an innocent suspect will be identified. Although a suspect cannot be forced to participate in a lineup, the use of photospread lineups (rather than a live lineup) can be conducted without a suspect’s consent or knowledge and without benefit of legal counsel. Also, refusal to participate in a live lineup can be used against a defendant in a criminal trial.

Conclusions and Prospectus

The psychological literature on eyewitness identification is less concerned today with the general thesis that eyewitnesses can be unreliable and more concerned with the question of what conditions foster reliable and unreliable identifications. The difficulty of separating accurate from false identifications post hoc, owing in large part to a weak and unstable certainty-accuracy relation, has led to a premium being placed on methods of preventing false identifications from occurring in the first place.

The single-suspect lineup, by virtue of its use of known-innocent foils, has the capability of shifting eyewitness errors from the harmful false-identification category to the relatively harmless selection of known-innocent distractors. Concepts such as functional lineup size and propitious heterogeneity can guide the lineup constructors’ selection of distractors to protect the innocent suspect without creating undue homogeneity in the appearance of the lineup members. The use of mock witnesses as a control condition in actual cases as well as the use of the blank lineup control, unbiased lineup instructions, and double blind techniques, follow analogously from existing knowledge in experimental psychology about what constitutes a properly designed and conducted experiment. Explicit recognition and exploration of this analogy has rich implications for defining the properly-conducted lineup in actual cases.

Although the process of recognizing a stranger who has been seen only once before is not fully understood (e.g., in terms of feature vs. holistic processes and parallel vs. serial processes), it seems clear that lineup identifications are influenced in part by relative judgments. In particular, positive lineup identifications are not purely epiphenomenal in that the identification decision is influenced by the extent to which a lineup member looks like the culprit relative to the alternative lineup members. Sequential lineups seem to reduce eyewitnesses’ propensities toward making relative judgments and serve to lower false identification rates without harming accurate identification rates.

The single-suspect lineup model, when properly executed in terms of distractor selection, instructions to the
witnesses, and so on, can produce relatively low rates of identification of innocent suspects. The prospects for further advances in the understanding, prediction, and control of eyewitness identification errors are considerable. The recent work of Loftus and her colleagues on the qualitative differences between "real" and "unreal" memories, for example, holds promise for distinguishing between accurate and false identification testimony (see Schoeller, Gerhard, & Loftus, 1986). The idea of generic cognitive retrieval mnemonics, which appear to be successful in verbal recall by eyewitnesses (e.g., see Geiselman, Fisher, MacKinnon, & Holland, 1985; Fisher, Geiselman, & Amador, 1989) but have not yet proved effective for identification tasks, hold some promise for improving lineup instructions given to eyewitnesses. Recent research showing significant increases in the accuracy-certainty relation as a function of retrospective self-awareness manipulations could enhance the utility of eyewitnesses' statements of their certainty (Kassin, 1985).

A false eyewitness identification can create a real-life nightmare for the identified person, friends, and family members. The innocent suspect need not be convicted to experience immense negative life events (see E. Loftus & Ketcham, 1991). False identifications also mean that the actual culprit remains at large—a double injustice. Scientific research methods used by psychologists in conjunction with bodies of knowledge in memory, cognition, social perception, and social influence provide powerful methods and theories that make research psychologists uniquely well suited to contribute solutions to the eyewitness identification problem.

4 Hypnosis, on the other hand, probably does not enhance eyewitness memory and might risk other problems, such as increased suggestion (see Orne, Soskis, Dingee, & Orne, 1984; Putnam, 1979; M. C. Smith, 1983.)

REFERENCES


570

May 1993 • American Psychologist


