Eyewitness identification evidence and innocence risk

STEVEN E. CLARK AND RYAN D. GODFREY
University of California, Riverside, California

It is well known that the frailties of human memory and vulnerability to suggestion lead to eyewitness identification errors. However, variations in different aspects of the eyewitnessing conditions produce different kinds of errors that are related to wrongful convictions in very different ways. We present a review of the eyewitness identification literature, organized around underlying cognitive mechanisms, memory, similarity, and decision processes, assessing the effects on both correct and mistaken identification. In addition, we calculate a conditional probability we call *innocence risk*, which is the probability that the suspect is innocent, given that the suspect was identified. Assessment of innocence risk is critical to the theoretical development of eyewitness identification research, as well as to legal decision making and policy evaluation. Our review shows a complex relationship between misidentification and innocence risk, sheds light on some areas of controversy, and suggests that some issues thought to be resolved are in need of additional research.

The ever-increasing number of cases in which DNA evidence has led to the exoneration of innocent people has focused a bright light on the problem of false identification. In 161 of the 218 DNA exonerations listed on the Innocence Project Web site (www.innocenceproject.org, last visited August 5, 2008), false identification was part of the evidence brought forward at trial, leading to the wrongful conviction. These DNA exonerations, as well as a number of other archival analyses, have led many to the conclusion that false identification is the primary cause of wrongful convictions in the United States (Gross, Jacoby, Matheson, Montgomery, & Patil, 2005; Huff, Rattner, & Sagarin, 1996; Scheck, Neufeld, & Dwyer, 2000; Wells et al., 1998).

The next question seems obvious: If false identifications are the primary cause of wrongful convictions, what are the causes of false identifications? The logic here is straightforward. If *A* causes *B*, it should be useful to know what causes *A*. However, the relationship between false identification and wrongful conviction is not so straightforward.

Psychological research, going back at least 100 years to Munsterberg's (1908) seminal book *On the Witness Stand*, up to, more recently, Cutler and Penrod (1995), Loftus (1979), and Wells (1993; Wells, Memon, & Penrod, 2006), has demonstrated the frailties of memory and the influence of suggestion, so there is little doubt but that witnesses make mistakes. However, a wrongful conviction requires a particular kind of eyewitness mistake—specifically, a false identification of a person who is suspected by the police but innocent.

The 161 false identifications noted above make the point that the identification of the suspect is not equivalent

to the identification of the perpetrator. This points to the central question of this article: What does it mean when the witness identifies the suspect? Is it a correct identification of the perpetrator, or a false, mistaken identification of an innocent suspect?

A mistaken or false identification can be viewed as a conditional probability—specifically, the probability that the suspect will be identified, given that he is innocent. However, the question we raise points to another important conditional probability: the probability that the suspect is innocent, given that he was identified. For brevity, we will refer to this probability, which lies at the core of the question before the jury, as *innocence risk*.

Simply put, innocence risk depends on two response probabilities: the probability that the suspect will be identified given that he is guilty (a correct identification) versus the probability that the suspect will be identified given that he is innocent (a false identification). Innocence risk is simply the proportion of all suspect identifications that are false identifications of an innocent suspect.

In this article, we will review the eyewitness identification literature, focusing on these three probabilities: correct identification rates, false identification rates, and innocence risk. This review reveals a multifaceted and complex relationship between correct and false identifications and innocence risk arising from memory limitations, similarity relationships, and decision making. As we will show, examination of the literature in this way can bring new insights to areas of uncertainty and controversy and also can raise new questions in areas involving problems that may have been thought of as having been already “solved.”

S. E. Clark, clark@ucr.edu

Table 1
Correct and Incorrect Eyewitness Identification Responses
for Showups and Lineups

Witness Response	Suspect Is Guilty (Target Present)	Suspect Is Innocent (Target Absent)
Suspect identification	Correct response	Incorrect response (false identification, misidentification, or mistaken identification)
No identification	Incorrect response	Correct response
Foil identification	Incorrect response	Incorrect response

Note—Foil identification response applies to lineups only.

FALSE IDENTIFICATION AND INNOCENCE RISK

As a preface, we will begin by briefly describing the procedures for obtaining eyewitness identification evidence and the various possible outcomes of those procedures. Two standard procedures for obtaining identification evidence, showups and lineups, are described in Table 1. In both cases, there are two states of the world that must be considered: The suspect is either guilty or innocent. Laboratory studies on eyewitness identification, which simulate real crimes with staged crimes, often map these two states onto two experimental conditions termed *target present* (guilty suspect) and *target absent* (innocent suspect).

A showup is identical in form to a yes–no test of recognition memory. The witness is presented with a single suspect and is asked, yes or no, whether that person is the perpetrator of the crime. Typically, although not always, showups are conducted when police identify a suspect shortly after the crime. For example, minutes after a convenience store robbery, police may spot a person nearby who fits the witness’s description of the robber. The police might simply drive the witness to the location where the suspect is detained and ask, “Is this the person who robbed the store?” The possible outcomes are shown in the two top panels of Table 1 and are exactly the same as those in a signal detection or yes–no recognition memory experiment. There are two correct responses (hits and correct rejections) and two incorrect responses (false alarms and misses).

An alternative procedure is to present the witness with a photographic or live lineup that includes the suspect and some number of *foils* or *fillers*. Again, the suspect may be guilty or innocent, but typically the foils are known to be innocent. The lineup identification procedure may be viewed as a combination and extension of yes–no and *n*-alternative forced choice testing procedures. The lineup presents the witness with *n* alternatives, but the task is not forced choice. *None of the above* is an acceptable response.

As is shown in Table 1, lineups allow for three different responses: an identification of the suspect, a none-of-the-above response, and one additional response, the identification of a foil, shown in the bottom panel. The correct responses are to identify the suspect if he is guilty and to

give a none-of-the-above response if the suspect is innocent. There are three kinds of incorrect responses: One can incorrectly reject the lineup when the guilty person is in it, one can identify a foil, or one can misidentify the innocent suspect. The distinction between these last two errors is of central importance. The identification of a foil rarely has any consequences for the person who is identified, because typically foils are known to be innocent.¹ The incorrect identification of an innocent suspect, however, is precisely the error that can result in a wrongful conviction. This particular error goes by various terms: *mistaken identification*, *misidentification*, and *false identification*. For clarity, we will not use any of those terms to refer to a *foil identification*.

It will be convenient at times to use abbreviations for these various terms. Thus, *suspTP* denotes a correct identification of the guilty person from a target-present lineup, *foilTP* denotes a foil identification from a target-present lineup, *noidTP* refers to a nonidentification response from a target-present lineup; likewise, *suspTA*, *foilTA*, and *noidTA* refer to target-absent lineups.

Typically, the members of a lineup are presented simultaneously; however, for reasons that will be discussed at length later in this article, a number of jurisdictions in the United States and Canada have begun using a sequential lineup procedure. In the sequential lineup, lineup members are presented one at a time, and the witness makes a *yes* or *no* response after each one.

Our review focuses primarily (although not exclusively) on two responses: correct identifications of the guilty suspect (*suspTP*) and false identifications of the innocent suspect (*suspTA*). These two responses frame the question that faces the members of the jury: Given that the suspect (who is now the defendant) was identified, how likely is it that he committed the crime? Viewed from the other side, how likely is it that he is innocent?

The question is illustrated in Table 2 with data from a typical eyewitness identification experiment (Tunnicliff & Clark, 2000). In this experiment, participants were witnesses to a staged crime: A perpetrator (a paid actor) came to the laboratory and stole an envelope full of cash. Half of the witnesses were later presented with a target-present lineup that included a photograph of the thief, and half were presented with a target-absent lineup that did not include a photograph of the thief but, instead, included a photograph of a person in the role of the innocent suspect. The table gives response probabilities for suspect, foil, and nonidentification responses, but only the suspect identifications are relevant to present purposes. It is clear

Table 2
Example Data From Tunnicliff and Clark (2000)

Witness Response	Target Present (Suspect Is Guilty)	Target Absent (Suspect Is Innocent)
Suspect identification	.531	.125
No identification	.218	.531
Foil identification	.250	.343

Note—Innocence risk is calculated as the probability that the suspect is innocent, given that he is identified: $.125 / (.125 + .531) = .191$.

from the example that a suspect identification was diagnostic of guilt, since they were obtained more often when the suspect was guilty (.531) than when the suspect was innocent (.125).

The innocence risk, the probability that the suspect is innocent, given that the suspect was identified, is simply $.125/ (.125 + .531) = .191$ (assuming equal base rates). It is easy to see that the innocence risk will increase if the false identification rate increases—for example, from .125 to .250 [innocence risk = $.250/ (.250 + .531) = .320$]. Innocence risk also increases if the correct identification rate decreases—from, say, .531 to .250 [innocence risk = $.125/ (.125 + .250) = .333$].

One could also calculate the other conditional probability, the probability that the suspect is guilty, given that he was identified (Clark, Howell, & Davey, 2008; Clark & Wells, 2008; Lindsay & Wells, 1980; Wells & Olson, 2002). This calculation, which is actually more commonly used, simply flips the terms in the equation. This probability is sometimes referred to as *response diagnosticity* (Lindsay & Wells, 1980) and, in more legal terms, represents the *probative value* of the evidence (see *Federal Rules of Evidence*, United States Congress, 2004). We calculate innocence risk, rather than its inverse, because false identification of the innocent has been a driving force in eyewitness identification research, and it is the exoneration of the innocent through DNA analysis that has focused the criminal justice system on issues of eyewitness identification (Wells et al., 2000).

Calculation of diagnosticity, probative value, and innocence risk involve simple applications of Bayes's theorem. The complete, and more formal, calculation of innocence risk—the probability that the suspect is innocent given that the suspect was identified, $p(\text{Susp} = I | \text{SuspID})$ —is given below. Borrowing from the example in Table 2, $p(\text{SuspID} | S = G) = \text{suspTP} = .531$ and $p(\text{SuspID} | S = I) = \text{suspTA} = .125$. The terms $p(\text{Susp} = G)$ and $p(\text{Susp} = I)$ are the prior probabilities of guilt and innocence. In real criminal cases, these prior probabilities might represent the relative frequencies of cases in which police present witnesses with a guilty suspect versus cases with an innocent suspect. These probabilities are unknown in real criminal investigations. For simplicity, in our example and in the calculations to follow, we followed the typical experimental procedure in which target-present and target-absent lineups are administered equally often: $p(\text{Susp} = G) = p(\text{Susp} = I) = .50$. None of our conclusions depend on this simplified equal prior probability assumption. However, after presenting all of our analyses, we will plot innocence risk as a func-

tion of prior probability in order to show the relationship between the prior probability of guilt and the posterior probability of innocence. In the next section, we will present a selective review of the eyewitness identification literature, focusing on correct and false identifications and innocence risk.

REVIEW OF THE EYEWITNESS IDENTIFICATION LITERATURE

The review that follows will present a systematic examination of how correct and false identification rates vary due to changes in memory, similarity, and decision processes, and also how correct and false identification rates combine to change the innocence risk associated with eyewitness identification evidence. In order to calculate innocence risk, our analyses are restricted to studies that have included both target-present and target-absent conditions. This is critical not only for calculation purposes, but also because comparisons based on studies with only target-present lineups or only target-absent lineups can be misleading; a weak manipulation in one study and a strong manipulation in another can suggest differences in target-present and target-absent lineups that may not actually exist.

For some studies, the false identification rates (suspTA) were not reported and, therefore, must be estimated. In these studies, the overall or total identification rates were reported, and we estimate the innocent suspect identification rate by dividing the overall identification rate by the size of the lineup. This estimation procedure, of course, assumes a fair lineup in which the innocent suspect is no more likely to be identified than anyone else in the lineup (a lengthier discussion of the implications of estimating the false identification rate in this way is given by Clark et al., 2008).

Our review of the literature aggregates data using simple meta-analytic techniques but is not a meta-analysis per se. The scope of such a meta-analysis would be much too broad, encompassing the entire eyewitness identification literature. Thus, we must be selective in the subsets of the literature that we review. Although we have been selective in terms of the topics we have chosen to discuss, our reviews of each topic are quite detailed. Also, contrary to the strategy employed in meta-analytic reviews, we have focused on the published literature, rather than on the "gray" literature (Conn, Valentine, Cooper, & Rantz, 2003; Cooper, 2003).

We have calculated Cohen's h (Cohen, 1988) to evaluate the effect sizes for changes in correct identification

$$p(\text{Susp} = I | \text{SuspID}) = \frac{p(\text{SuspID} | \text{Susp} = I)p(\text{Susp} = I)}{[p(\text{SuspID} | \text{Susp} = I)p(\text{Susp} = I)] + [p(\text{SuspID} | \text{Susp} = G)p(\text{Susp} = G)]}$$

where $p(\text{Susp} = I | \text{SuspID})$ denotes innocence risk, $p(\text{SuspID} | \text{Susp} = I)$ is the probability of a suspect identification given that the suspect is innocent (i.e., a false identification), and $p(\text{SuspID} | \text{Susp} = G)$ is the probability that the suspect is identified given that the suspect is guilty.

rates (h_C), changes in false identification rates (h_F), and the resultant changes in innocence risk (h_{IR}). For consistency, rather than assigning positive and negative signs on the basis of the direction of a particular hypothesis, we have assigned positive signs for increases and negative signs for decreases in correct identification rates, false identification rates, and innocence risk. As a rough guide to the interpretation of Cohen's h as a measure of effect size, $h = .20$ describes a small effect, $h = .50$ describes a medium-sized effect, and $h = .80$ describes a large effect (see Cohen, 1988, pp. 184–185, for details).

In each case, we will present a simple random-effects analysis to evaluate statistical reliability. Random-effects analyses are more conservative than fixed-effects analyses, but allow for broader generalization. For these random-effects analyses, the unit of analysis is the individual study, rather than the participants in those studies, and thus each study is weighted equally, contrary to fixed-effects analyses, which weight large N studies more heavily than small N studies. These random-effects analyses are evaluated by simple one-sample t tests. It is important to be clear, however, that our goal is not to make a series of yes–no determinations as to whether various effects are statistically “significant.” Rather, h and t provide measures of effect sizes and consistency across a set of studies.

Organization of the Review

Reviews of the eyewitness identification literature have typically been organized around a distinction between estimator and system variables (Lindsay & Pozzulo, 1999; Lindsay, Ross, Read, & Toglia, 2007; Wells et al., 2006; Wells & Olson, 2003). The distinction, initially made by Wells (1978), is made clear by noting that *system* refers to the criminal justice system. System variables are those over which the criminal justice system has considerable control—that is, how lineups are constructed and how they are presented. Estimator variables include those elements over which the criminal justice system has little or no control, including the witness's conditions of observation. The estimator–system variable distinction is critical to recent efforts to reform police procedures, because it distinguishes between those variables over which the criminal justice system has control and those over which it does not. However, the distinction does not map well onto the underlying cognitive processes.

Our review is organized around a simple theoretical framework consistent with the general approach of similarity-based memory models, which base recognition decisions on a match between test items and memory. These include global-matching models (see Clark & Gronlund, 1996, for a review), REM (Shiffrin & Steyvers, 1997), SLiM (McClelland & Chappell, 1998), array models (Estes, 1997), and the generalized context model (Nosofsky, 1992). Consistent with the general properties of similarity-based models, our review of the eyewitness identification literature is guided by a simple partitioning of the eyewitness task into memory processes, similarity relationships, and decision processes.²

Memory processes. The witness's memory of the perpetrator may be incomplete or inaccurate due to limi-

tations at the time of encoding and storage. In addition, over time, information in the memory trace may be lost or rendered irretrievable due to interference from other information added to memory. These are familiar and standard accounts of the *general limitations of memory*. In a feature-based memory model, such general memory limitations would include a failure to store features or the storage of incorrect feature values (Shiffrin & Steyvers, 1997), as well as perturbation and drift of feature values over time (Estes, 1997).

Memory may also be altered in very specific ways, due to the witness's exposure to other sources of information after the crime. Here, we refer to what is called *postevent information* (Loftus, 2005). There are many sources of postevent information in real criminal cases: Witnesses can be exposed to postevent information through interviewer questions, news reports, and photographs of the suspect. For the present purposes, it is not necessary to take a stand on the issue of how postevent information alters memory but only to note that it does alter memory, not in the general way described above, but rather by adding specific content to memory—that is, information that is specific to the suspect.

Similarity relationships. In all of the models listed earlier, memory decisions are based on the match or similarity between test items and memory. Thus, a six-person lineup would yield six match values, each representing the match between a given lineup member and the witness's memory of the perpetrator. These match values depend on several other similarity relationships: (1) the similarity of the perpetrator at the time of the identification to the perpetrator's appearance when observed at the crime by the witness; (2) the similarity of an innocent suspect to the actual perpetrator; and (3) the similarity between the foils in the lineup to the witness's memory of the perpetrator. Of course, the similarity of the foils to the witness's memory cannot be manipulated directly but, rather, is mediated by the similarity to the suspect or the match to the witness's description of the perpetrator (because police usually use some combination of the suspect's photograph and the witness's description to select foils).

Decision processes. The match values do not specify the witness's decision. The witness must apply some decision rule on the match values in order to generate a response. It is at this point that suggestive statements or social influence from law enforcement have their effect. Police officers administering the lineup may influence witnesses both as to their willingness to make an identification and as to whom to identify, presumably by inducing witnesses to alter their decision processes.

With this organizing framework, our review will proceed as follows. We first will consider general limitations of memory, followed by suspect-specific alternations of memory due to postevent exposure to suspect information. We then will turn to a discussion of similarity relationships, highlighting three main relationships: the overall similarity of foils, specific strategies for selecting foils, and the absence of foils. The third section will examine influences on witnesses' decisions to make identifications and their decisions regarding whom to identify.

MEMORY LIMITATIONS AND ALTERATIONS

General Limitations of Memory

General memory limitations refer to nonspecific failures to store or retain information. As examples, we will consider studies in which stress, exposure duration, and retention interval have been manipulated. Although it may seem intuitive that witnesses should be less accurate when they have less opportunity to observe the perpetrator, make their observations under stressful conditions, or make identifications after long delays, intuitive predictions have, in some cases, met with counterintuitive data (Read, 1995) and controversy. Ebbesen and Konecni (1996) questioned how the failure to encode or the failure to retain information could increase the likelihood of a misidentification of an innocent suspect. Regarding memory limitations due to short exposure duration, they wrote, "There is little denying the manifest intuition that longer exposure durations should lead to more reliable identifications. However, . . . the literature suggests that this . . . may apply only to hits versus misses . . . and not to false alarms" (p. 7). Regarding the loss of information over time, "Surely memory for a never seen face (i.e., an innocent suspect) does not become stronger as time goes on." The effects of delay have been controversial in the eyewitness memory literature for decades (McCloskey & Egeth, 1983) and continue to generate controversy, since some assert that the effects of delay are intuitive (Wells et al., 2006), whereas others (Dysart & Lindsay, 2007) have asserted that "at this point in time we do not know much about the impact of delay on eyewitness identifica-

tion accuracy" (p. 373). The effects of stress are also not universally accepted (Ebbesen & Konecni, 1996; Yuille & Tollestrup, 1992).

By considering stress, exposure duration, and retention interval together, we are, of course, assuming that they may be treated as different pathways to the same outcome—that is, less accurate information in the memory trace. Thus, Table 3 presents the relevant data as comparisons between better (lower stress, longer exposures, shorter retention intervals) and worse memory conditions. Correct (suspTP) and false (suspTA) identification rates and innocence risk [$\text{suspTA} / (\text{suspTA} + \text{suspTP})$] are presented, along with the corresponding effect sizes (h_C , h_F , and h_{IR} for correct identifications, false identifications, and innocence risk).

Table 3 shows consistent decreases in correct identifications [$h = -.445$, $t(12) = -3.453$, $p = .005$] and generally small and inconsistent results for false identifications, with an average effect size near zero [$h = .018$, $t(12) = 0.404$, $p = .693$]. The decrease in correct identifications, with small and inconsistent changes in false identifications, combine to increase the innocence risk for suspect identifications [$h = .239$, $t(12) = 3.192$, $p = .008$]. General limitations in memory decrease the correct identification rate, have very little effect on the false identification rate, and increase innocence risk.

This combination of results leads to a conclusion that appears self-contradictory: Conditions that result in reduced accuracy or incompleteness of the memory trace do not reliably increase the likelihood of a false identification but do increase the likelihood that the identification of the suspect is, in fact, a false identification. As

Table 3
Correct and False Identifications and Innocence Risk (IR) As a Function of Nonspecific Memory Limitations

Study	Better Memory Conditions			Worse Memory Conditions			Effect Sizes		
	Correct	False	IR	Correct	False	IR	h_C	h_F	h_{IR}
Stress									
Morgan et al. (2004, live)	.621	.033	.050	.267	.030	.101	-.729	-.017	.196
Morgan et al. (2004, photo)	.761	.038	.048	.337	.030	.082	-.881	-.044	.139
Exposure Duration									
Memon, Hope, & Bull (2003)	.950	.068	.067	.290	.150	.341	-1.553	.268	.724
Boyce, Lindsay, & Brimacombe (2008, Experiment 1)	.294	.154	.344	.088	.118	.573	-.544	-.105	.464
Boyce et al. (2008, Experiment 2)	.610	.148	.195	.225	.178	.442	-.804	.081	.540
Read (1995, high similarity)	.467	.070	.130	.385	.022	.054	-.166	-.238	-.269
Read (1995, low similarity)	.273	.078	.222	.176	.083	.320	-.234	.018	.221
Retention Interval									
Krafka & Penrod (1985)	.273	.017	.059	.308	.091	.228	.077	.351	.505
Shepherd, Ellis, & Davies (1982, p. 97)	.300	.058	.162	.091	.035	.278	-.546	-.110	.282
Yarmey, Yarmey, & Yarmey (1996)	.486	.162	.250	.324	.140	.302	-.332	-.062	.116
Dekle (1997)	.299	.051	.146	.320	.097	.233	.045	.178	.223
Memon, Bartlett, Rose, & Gray (2003)	.410	.099	.194	.350	.075	.176	-.124	-.085	-.046
Brewer, Caon, Todd, & Weber (2006)	.300	.058	.162	.300	.059	.164	.000	.004	.005
<i>M</i>	.465	.080	.156	.266	.085	.253	-.445	.018	.239
<i>SD</i>							.465	.164	.269
<i>t</i> (12)							-3.453	0.404	3.192
<i>p</i>							.005	.693	.008

Note—Correct identifications are suspect identifications from target-present lineups, false identifications are suspect identifications from target-absent lineups. Shepherd et al. (1982) presented many experiments that were not numbered. The data presented here are from their Tables 11.2 and 11.3, pp. 96–97.

we will argue later, the self-contradicting sound of this conclusion may lead jurors to ignore evidence related to general memory limitations when evaluating eyewitness identification evidence.

For now, we will turn our attention to the asymmetric pattern of results for correct and false identification rates, the possible role of criterion shifts, and how such criterion shifts may obscure the decline of identification accuracy due to the passage of time.

The asymmetric pattern for correct and false identifications is inconsistent with the ubiquitous mirror effect found for recognition memory: For two conditions A and B, where performance is better for A than for B, the results typically show a mirror pattern—that is, an increase in hit rates that mirrors the decrease in false alarm rates for Condition A relative to Condition B (Glanzer & Adams, 1985, 1990). The asymmetry found in the present analyses may be explained in part by a difference in target-present and target-absent lineups. A weak memory trace makes the target (guilty suspect) less distinguishable from the lineup foils, resulting in lower correct identification rates. However, in a target-absent lineup, even if the correct rejection rate decreases (i.e., the mirror pattern), the increased error will be evenly distributed across lineup members, assuming a fair and unbiased lineup. Thus, the correct rejection rate in a target-absent lineup may fall from .60 to .20, for example, but the increase in error (from .40 to .80) will be distributed across all lineup members in such a way that the increases in false identifications of the innocent suspect (from .067 to .133) will be fairly small.

This analysis predicts that the effects of general memory limitations should be more symmetric (and show a mirror effect) when there are no foils—that is, when showup procedures are used, rather than lineup procedures. Indeed, this appears to be the case. Yarmey, Yarmey, and Yarmey (1996) compared lineups and

showups either immediately or after a 30-min, 2-h, or 24-h delay. Their results are shown in Figure 1. Panel A shows that the correct identification rate dropped off quickly over time and that the false identification rate increased sharply at the 30-min delay, but then decreased and remained fairly constant through the 24-h delay. The showup results, in contrast, showed not only a steep drop in correct identification rates, but also a large increase in false identifications as well. After a 2-h delay, correct and false identification rates had converged, showing chance performance. The innocence risk calculations shown in panel C show roughly equal levels of risk for showups and lineups at short retention intervals, but at 2- and 24-h delays, the innocence risk for showups was much higher than that for lineups. Dekle (1997) showed a similar pattern—that is, a larger effect of delay for showups than for lineups—although the pattern was not as pronounced as in the Yarmey et al. results.³ These lineup and showup results suggest that Ebbesen and Konecni's (1996) skepticism about increases in false identifications may apply to lineups but not to showups. This may largely be the case, although we note that five studies in Table 3 (Boyce et al., 2008; Dekle, 1997; Krafka & Penrod, 1985; Memon, Bartlett, Rose, & Gray, 2003; Read, 1995) did show false identification increases. The mirror pattern may arise in lineups, but the distribution of responses over multiple lineup members may cause the false identification side of the mirror to be small and blurry.

The results shown in Table 3 are also complicated by the possibility that witnesses may utilize different decision criteria for better and worse memory conditions. This is relevant, of course, to the current controversy regarding strength-based mirror effects. Specifically, the reduction in false positive responses with increased strength, which is half of the mirror pattern (longer exposure durations, shorter retention intervals), may be due to an upward criterion shift (Singer & Wixted, 2006; Stretch & Wixted, 1998) but may

Figure 1. Forgetting in showups and lineups (Yarmey, Yarmey, & Yarmey, 1996). (A) Correct (filled circles) and mistaken (open circles) identification rates for lineups. (B) Correct (filled squares) and mistaken (open squares) identification rates for showups. (C) Innocence risk calculations for lineups (closed circles) and showups (closed squares).

Table 4
Overall Identification Rate and Identification Accuracy in Target-Present Lineups

Study	Better Conditions		Worse Conditions	
	ID Rate	ID Accuracy	ID Rate	ID Accuracy
Memon, Hope, & Bull (2003)	.975	.923	.755	.424
Boyce, Lindsay, & Brimacombe (2008, Experiment 1)	.529	.556	.441	.200
Boyce et al. (2008, Experiment 2)	.780	.781	.650	.346
Read (1995, high similarity)	.934	.500	.692	.556
Read (1995, low similarity)	.637	.429	.293	.601
Krafka & Penrod (1985)	.364	.750	.462	.667
Shepherd, Ellis, & Davies (1982, p. 97)	.600	.500	.318	.286
Shepherd et al. (1982, p. 85) ^a	.800	.813	.250	.400
Shepherd et al. (1982, p. 90) ^a	.453	.458	.627	.438
Yarmey, Yarmey, & Yarmey (1996)	.771	.630	.794	.407
Dekle (1997)	.433	.690	.590	.542
Memon, Bartlett, Rose, & Gray (2003)	.640	.641	.350	1.000
Brewer, Caon, Todd, & Weber (2006)	.475	.632	.550	.545

Note—ID rate = overall identification rate = suspTP + foil TP; identification accuracy = suspTP/(suspTP + foilTP). ^aStudies not included in Table 3.

also be due to a differentiation process that reduces the match for nonstudied test items (Criss, 2006; see also McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997).

We do not presume to resolve this controversy here. However, it is useful to examine the relationship between the *overall identification rate* and *identification accuracy* as a function of memory strength, due to variation in exposure duration and retention interval. To that end, we calculated two additional response rates for the 11 target-present lineup conditions in the experiments listed in Table 3, plus two additional experiments (Shepherd, Ellis, & Davies, 1982, pp. 85, 90) not included in Table 3 (because they did not include a target-absent lineup condition). We calculated the overall identification rate by summing suspect and foil identifications suspTP + foilTP, and we calculated the accuracy of those identifications as suspTP/(suspTP + foilTP).

Table 4 shows five comparisons for exposure duration and eight for retention interval. In all five exposure duration studies, the overall identification rate was higher for the longer exposure condition. However, the longer exposure duration does not always lead to higher accuracy of those identifications. In both conditions in the study by Read (1995), witnesses made more but less accurate identifications with longer exposure times. This pattern of results suggests that witnesses's willingness to make an identification outstretched their ability to make an accurate identification. The long exposures in the Read study were quite long, with face-to-face interaction. Under such conditions, people may believe that their memories are much better than they really are, leading to a high rate of identification, but with lower accuracy.

The retention interval studies showed a different pattern of results. In five of eight cases, the overall identification rate either increased or stayed the same at the longer retention interval. In all five cases, however, the accuracy of those identifications decreased. Increasing the overall identification rate can give the appearance that memory has not faded over time, with correct identification rates that decrease very little (Krafka & Penrod, 1985) or actually increase slightly (Dekle, 1997) over time.

Because the effects of delay have been (McCloskey & Egeth, 1983) and continue to be (Dysart & Lindsay, 2007) a source of controversy, it is important to discuss delay effects in greater detail.⁴ In our analysis, four of six studies showed increased risk with longer delays, and two showed no change in risk due to the longer delay. On the whole, innocence risk increased with longer delays, not because of the increase in false identification rates, but rather because of the decrease in correct identification rates. This may shed light on other results that have led some (Dysart & Lindsay, 2007; McCloskey & Egeth, 1983) to question the effects of delay on face recognition. Specifically, the asymmetric pattern may account for the noneffects of delay reported by Laughery, Fessler, Lonerovitz, and Yoblick (1974) and Mauldin and Laughery (1981), whose test procedures consisted of 149 and 130 test trials, respectively, of which all were distractors with one target. To the extent that there is an effect for targets more than for distractors, and almost all test trials were distractors, it is not hard to see how null results might be obtained.

Suspect-Specific Alterations of Memory

Work by Loftus and her colleagues (Loftus, 1979; Loftus, Miller, & Burns, 1978; see Loftus, 2005, for a review) has shown that memory for a given event can be influenced or distorted by exposure to postevent information. If the postevent information is correct, the witness's accuracy in response to subsequent questions will increase, but if the postevent information is incorrect, the witness's accuracy will decrease (Loftus et al., 1978).

In eyewitness identification experiments, the most straightforward implementation of the postevent information design exposes the witness to the suspect after the staged crime and before the lineup. For example, witnesses may be presented with intervening lineups or mugshots prior to the critical lineup identification task. A procedural replication of the Loftus et al. (1978) study would present witnesses with a photograph of either the perpetrator (correct information) or an innocent person (false information) in the intervening postevent phase. However, nearly all of the relevant studies have presented

only innocent people in the postevent phase (see Deffenbacher, Bornstein, & Penrod, 2006). These experiments show increases in false identifications and increases in innocence risk. However, they tell only half of the story.

It is critical, of course, to consider not only the possibility that the suspect is innocent (which is what most studies have done), but also the possibility that the suspect is guilty. The Deffenbacher et al. (2006) review listed only one study—by Brown, Deffenbacher, and Sturgill (1977)—that presented witnesses with both guilty and innocent people during the postevent phase. The experiments had some methodological problems that limit the conclusions and generality.⁵ With those caveats, their results showed increases in both correct and false identifications due to the presentation of mugshots during the intervening phase. Specifically, in their Experiment 2, correct identification rates increased from .51 to .65, and false identification rates increased from .08 to .20. In their Experiment 3, correct identification rates increased from .24 to .45, and false identification rates increased from .18 to .29. Thus, the innocence risk due to the presentation of mugshots increased in Experiment 2 from .136 to .263 ($h = .322$) but decreased slightly in Experiment 3, from .429 to .392 ($h = -.075$). The general result—that correct identification rates increase with an intervening presentation of the guilty suspect and false identifications increase with an intervening presentation of the innocent suspect—is consistent with results from postevent information studies (Loftus et al., 1978) and with results from repeated-testing procedures in recognition memory experiments (Challis & Roediger, 1993; Payne & Roediger, 1987).⁶ However, because of the departures from the eyewitness identification paradigm, the quantitative aspects, and the innocence risk calculations in particular, should be interpreted with caution. Additional research, extending the full design of the postevent information paradigm to eyewitness identification, is clearly needed.

Summary and Synthesis

We have reviewed two very different kinds of memory alterations in terms of how they affect correct and false identifications and innocence risk. The clearest result that emerged from this review is that nonspecific limitations of memory, due to failures to encode, store, or maintain information over time, decrease the probative value of suspect identifications and increase innocence risk—in most cases (but not all), through a substantial decrease in the correct identification rate and little change in the false identification rate. There were some potentially interesting deviations, however, from that general pattern of results.

In a few studies (Dekle, 1997; Krafka & Penrod, 1985), correct identification rates increased, or decreased only very little, with the passage of time. On the surface, it might appear that these studies showed no loss of accuracy over time. However, in both studies, the correct identification rate appears to have been “propped up” by an increased willingness to pick; the overall identification rate increased, although the accuracy of those identifications decreased. This general pattern was shown in several studies. This raises two questions. First, do witnesses

consciously or unconsciously compensate for the loss of information over time by lowering their decision criterion? Second, why are the patterns different for variations in exposure duration? We do not have answers for these questions but pose them for future research.

The results of stress, exposure duration, and delay studies, taken together, show that general memory limitations may contribute to increased innocence risk in two ways: *directly*, due to increases in the false identification rate (Boyce, Lindsay, & Brimacombe, 2008; Memon, Bartlett, et al., 2003; Yarmey et al., 1996), but also *indirectly*, due to decreases in the correct identification rate, *even if the misidentification rate remains constant*. The last point is of particular importance and again constitutes a rather awkward conclusion: The likelihood that the suspect is innocent given that he was identified may increase even if the likelihood that the suspect is identified given that he is innocent is unchanged. How do, and how should, the consumers of eyewitness identification evidence (police, judges, lawyers, juries) reason about identification evidence, given this complex and indirect association between memory and false identification?

The results of memory alteration, through postevent exposure to the suspect, show a very different pattern. The seminal study of postevent information effects (Loftus et al., 1978) showed both increases in accuracy following exposure to correct postevent information and decreases in accuracy following exposure to incorrect postevent information. In only one study (Brown et al., 1977) have both sides of postevent information effects due to preidentification exposure to both innocent and guilty suspects been examined. The results were inconsistent, showing an increase in innocence risk in one experiment and a decrease in the other. The inconsistency in outcomes, combined with the problems of generalization, makes a clear argument for additional research in this area.

SIMILARITY RELATIONSHIPS

Three similarity relationships will be considered in this section: the similarity of the perpetrator as he appears in the lineup to the perpetrator as he appeared at the time of the crime, the similarity of an innocent suspect to the perpetrator, and the similarity of the foils to the perpetrator. The effects of varying the first two similarities are fairly predictable and will be discussed only briefly. The third similarity, foil similarity, will be considered in much more detail. We also will review the effects of removing the foils entirely by comparing lineups and showups.

Suspect and Perpetrator Similarity

We will consider first two similarity relationships: (1) the similarity of the perpetrator as he appeared at the time of the crime to the perpetrator as he appears at the time of the identification and (2) the similarity of the innocent suspect to the perpetrator. In both cases, the results are rather straightforward, and we will summarize them quickly: If the perpetrator changes his appearance, the correct identification rate decreases and the innocence risk increases (Pozzulo & Marciniak, 2006; Read, 1995).⁷

Table 5
Correct and False Identifications and Innocence Risk (IR) in Lineups With More Similar or Less Similar Foils

Study	More Similar Foils			Less Similar Foils			Effect Sizes		
	Correct	False	IR	Correct	False	IR	h_C	h_F	h_{IR}
Lindsay & Wells (1980)	.580	.310	.348	.710	.700	.496	.273	.801	.301
Lindsay, Wallbridge, & Drennan (1987, Experiment 1)	.667	.111	.143	.778	.278	.263	.249	.431	.301
Lindsay et al. (1987, Experiment 2)	.833	.278	.250	.778	.389	.333	-.139	.236	.183
Lindsay et al. (1987, Experiment 3)	.444	.235	.346	.529	.470	.470	.170	.499	.253
Lindsay, Martin, & Webber (1994)	.793	.250	.240	.806	.500	.383	.033	.524	.311
Wells, Rydell, & Seelau (1993)	.667	.119	.151	.714	.429	.375	.102	.724	.520
Carlson, Gronlund, & Clark (2008)	.314	.163	.342	.706	.644	.477	.806	1.032	.276
<i>M</i>	.614	.209	.260	.717	.487	.400	.213	.607	.306
<i>SD</i>							.296	.264	.104
<i>t</i> (6)							1.905	6.074	7.807
<i>p</i>							.105	.001	.000

If the innocent suspect is more similar to the perpetrator, the false identification rate increases and the innocence risk also increases (Clark & Tunnicliff, 2001). These similarity effects are not surprising, and we include them primarily for completeness. The main concern, however, is with the similarity of foils, which we will take up next.

Foil Selection: Higher Versus Lower Similarity

On the surface, the selection of foils seems straightforward. If foils are selected in such a way that only the suspect is a plausible choice, the task seems inherently unfair and biased. Intuitions on this point are the stuff of cartoons, such as the one that appeared in Loftus's (1979) book *Eyewitness Testimony*, in which an angry witness points at the only plausible suspect standing conspicuously in a lineup with a blind man with a cane, an elderly woman in a wheelchair, and a little kid blowing a bubble with chewing gum (p. 145).

The intuition underlying the cartoon is that if foils can be easily ruled out, the likelihood that the suspect will be identified increases in a way that may be quite meaningless, given that it is the result of the lineup composition, rather than of the witness's memory. It seems intuitive also that the likelihood of a suspect identification should increase both when the suspect is guilty (target-present lineup) and when he is innocent (target-absent lineup). In both cases, poor foil selection should make the suspect stand out. Of course, probative value and innocence risk are determined by the proportional increases in correct and false identifications, so it is not obvious a priori how or whether innocence risk will change as a function of foil similarity.

Foil similarity manipulations have been instantiated in a variety of ways, but primarily by including foils who mismatch one or several elements of the perpetrator's description, although in one study foil bias was introduced by having only the suspect wear clothing worn by the perpetrator. Seven comparisons are shown in Table 5. The number of comparisons may seem small; however, as in the previous analyses, we will restrict our analyses only to those cases that presented both target-present and target-absent lineups. Six of the seven comparisons showed the same pattern: When less similar foils were included in the lineup, both correct and false identification rates increased. The increase in false identification rates was larger than the increase in correct identification rates, pro-

ducing an overall increase in innocence risk.⁸ The statistical details are as follows: for correct identifications, $h_C = .213$, $t(6) = 1.905$, $p = .105$; for false identifications, $h_F = .607$, $t(6) = 6.074$, $p = .001$; and for innocence risk, $h_{IR} = .306$, $t(6) = 7.807$, $p < .001$.

Foil Selection: Suspect Matched Versus Description Matched

In the studies listed in Table 5, foil similarity was defined in terms of mismatch to the target's description. However, it is common for police officers to select foils on the basis of their match to a photograph of the suspect (Wogalter, Malpass, & McQuiston, 2004). Luus and Wells (1991) suggested that a photograph of a suspect, because it contains more information than a witness's verbal description of the perpetrator, may lead police officers to select foils that are more similar—and perhaps unnecessarily similar—to the perpetrator. Assuming this to be the case, a comparison of suspect-matched and description-matched foil selection should be equivalent to a comparison of more similar versus less similar foils and should produce results like those in the seven studies shown in Table 5.

There are, however, other differences between suspect-matched and description-matched lineups, which have been discussed in detail elsewhere (Clark, 2003; Clark et al., 2008), and we will summarize the issues only briefly here. One important difference is that when foils are selected on the basis of their similarity to the suspect, they will be similar to the guilty suspect in a target-present lineup but will be similar to the innocent suspect in a target-absent lineup. In the target-absent lineup, this may produce what Clark and Tunnicliff (2001) called a *backfire effect* (see also Navon, 1992). Specifically, the similarity of the foils to the perpetrator in a target-absent lineup is mediated by the similarity of the innocent suspect to the perpetrator in such a way that the foils cannot (on average) be more similar to the perpetrator than is the innocent suspect, making the innocent suspect the person in the lineup who is most likely to be identified. Tunnicliff and Clark (2000) suggested that description-matched foil selection should reduce (or possibly eliminate) the backfire effect as well.

Clark et al. (2008) have recently reported a meta-analysis of four studies in which suspect-matched and description-matched lineups were compared directly. The

Table 6
Correct and False Identifications and Innocence Risk (IR) for Lineups
With Suspect-Matched Versus Description-Matched Foils

Study	Suspect Matched			Description Matched			Effect Sizes		
	Correct	False	IR	Correct	False	IR	h_C	h_F	h_{IR}
Wells, Rydell, & Seelau (1993)	.214	.119	.357	.667	.119	.151	.950	.000	-.483
Lindsay, Martin, & Webber (1994)	.655	.083	.112	.793	.250	.240	.311	.463	.341
Juslin, Olsson, & Winman (1996)	.438	.094	.177	.521	.094	.153	.166	.000	-.065
Tunnicliff & Clark (2000, Experiment 1)	.531	.031	.055	.531	.125	.191	.000	.369	.431
Tunnicliff & Clark (2000, Experiment 2)	.333	.083	.200	.313	.188	.375	-.043	.313	.391
<i>M</i>	.434	.082	.180	.565	.155	.222	.277	.229	.123
<i>SD</i>							.402	.216	.393
<i>t</i> (4)							1.541	2.373	0.700
<i>p</i>							.198	.077	.522

results, along with those of one additional study that did not satisfy the inclusion criteria for the Clark et al. analysis, are presented in Table 6. The results are variable, with some studies showing a reduction in innocence risk and others showing an increase. In two studies (Tunnicliff & Clark, 2000, Experiment 1, and Lindsay, Martin, & Webber, 1994), the innocent suspect was selected in such a way as to create biased lineups; however, description-matched foil selection did not reduce the problem, and even if those two studies are set aside, the results of the remaining studies are not consistent. Comparing these studies with the foil similarity studies listed in Table 5 suggests that the suspect-matched–description-matched distinction involves much more than similarity.

No Foils: Showups Versus Lineups

We will consider the innocence risk for showup procedures under the heading of foil similarity because the showup–lineup comparison is essentially one between N foils and no foils. Steblay, Dysart, Fulero, and Lindsay (2003) conducted a meta-analysis of eight studies that compared lineup and showup procedures. They concluded that witnesses, on the whole, were more accurate for showups than for lineups, a conclusion that is inconsistent with that of a number of legal writings that have pronounced showups to be “inherently,” “automatically,” or “impermissably” suggestive (American Law Institute, 1975; Calandra & Carey, 2006; National Institute of Justice, 1999). Steblay et al.’s (2003) conclusion was based on the sum of correct identifications in target-present lineups and correct rejections in target-absent lineups. This dependent measure may place lineups at a disadvantage, because foil identifications contribute to the error rates for lineups but not for showups. Also, because probative value and innocence risk are calculated with respect to a given response, and the measure used by Steblay et al. (2003) summed across two different responses, a reanalysis of the probative value and innocence risk associated with suspect identifications may lead to a different conclusion.

Suspect identification rates and innocence risk were calculated for six of these studies (excluding those with children as witnesses), with a total of 15 showup–lineup comparisons. As before, when necessary, false identification rates were calculated by dividing by the number of lineup members. The suspect identification rates and innocence

risk calculations for lineups and showups are shown in Table 7. Collapsed over all 15 showup–lineup comparisons, showups produce slightly lower correct identification rates [$h_C = -.078$, $t(14) = -0.769$, $p = .455$] and higher false identification rates [$h_F = .163$, $t(14) = 1.847$, $p = .086$], which combine to produce an increase in innocence risk [$h_{IR} = .269$, $t(14) = 2.193$, $p = .046$].

Summary and Synthesis

Two results emerge with clarity from our review: First, low-similarity foils increase innocence risk by increasing false identification rates, with consistent but smaller increases in correct identification rates. The consistent increase in correct identification rates due to the inclusion of low-similarity foils in the lineup should not be overlooked. It has been suggested that low-similarity foils do not lead to significant increases in correct identification rates (Lindsay & Pozzulo, 1999); however, six of the seven comparisons reviewed here showed an increase in the correct identification rate, very small in some cases, but quite large in other cases.

The second clear finding is that the absence of foils—the characteristic that defines the showup procedure—leads to a substantial increase in innocence risk, relative to the risk observed in lineup procedures. This increase in innocence risk arises from an increase in the false identification rate and a very slight decrease in the correct identification rate. Unlike other analyses that have shown increases in both correct and false identifications, the showup–lineup analysis is unique in its clear demonstration of costs with no benefits.

One issue that is not clear concerns suspect-based versus description-based foil selection. Analyses and arguments by Luus and Wells (1991), Navon (1992), and Clark and Tunnicliff (2001) make a compelling case for description-based foil selection, and this preference is clear in the National Institute of Justice’s (1999) guidelines. The data, however, are not as clear, and more research will be needed to address the disconnect between the guidelines and the relevant data.

DECISION BIASES AND DECISION RULES

One of the central questions for eyewitness identification research concerns the decision processes that

Table 7
Correct and False Identifications and Innocence Risk (IR) in Lineups and Showups

Study	Lineup			Showup			Effect Sizes		
	Correct	False	IR	Correct	False	IR	h_C	h_F	h_{IR}
Wagenaar & Veefkind (1992, K = 2)	.556	.125	.184	.347	.111	.242	-.423	-.043	.142
Wagenaar & Veefkind (1992, K = 6)	.500	.072	.126	.347	.111	.242	-.311	.136	.303
Wagenaar & Veefkind (1992, K = 10)	.424	.048	.102	.347	.111	.242	-.158	.238	.379
Wagenaar & Veefkind (1992, Experiment 2)	.750	.049	.061	.500	.146	.226	-.524	.338	.492
Gonzalez, Ellsworth, & Pembroke (1993, Experiment 1)	.429	.125	.226	.214	.067	.238	-.467	-.199	.028
Gonzalez et al. (1993, Experiment 2)	.133	.083	.384	.000	.025	1.000	-.747	-.267	1.805
Dekle, Beal, Elliot, & Huneycutt (1996, no delay)	.299	.051	.146	.280	.040	.125	-.042	-.053	-.061
Dekle (1997, 2- to 3-day delay)	.300	.063	.174	.350	.050	.125	.107	-.056	-.138
Dekle (1997, 7-day delay)	.320	.097	.233	.300	.130	.302	-.043	.104	.156
Yarmey, Yarmey, & Yarmey (1996, 0 h)	.486	.162	.250	.697	.176	.202	.433	.037	-.115
Yarmey et al. (1996, 0.5 h)	.382	.333	.466	.636	.441	.409	.514	.222	-.115
Yarmey et al. (1996, 2 h)	.364	.143	.282	.543	.579	.516	.362	.954	.483
Yarmey et al. (1996, 24 h)	.324	.140	.302	.556	.528	.487	.472	.860	.381
Lindsay, Pozzulo, Craig, Lee, & Corber (1997, simultaneous)	.548	.057	.094	.500	.070	.123	-.096	.053	.093
Lindsay et al. (1997, sequential)	.621	.042	.063	.500	.070	.123	-.244	.123	.209
<i>M</i>	.429	.106	.206	.408	.177	.307	-.078	.163	.269
<i>SD</i>							.392	.342	.476
<i>t</i> (14)							-0.769	1.847	2.193
<i>p</i>							.455	.086	.046

Note—*K* denotes lineup size.

witnesses use in making identification decisions. Decision processes have been described primarily through a distinction made by Wells (1984) between absolute and relative judgments. Briefly, an absolute judgment is an assessment of how closely a lineup member matches the witness's memory for the perpetrator, whereas a relative judgment is an assessment of that match relative to other lineup members. In the strongest version of a relative judgment rule, the witness simply picks the best match in the lineup, with little or no regard to how good that match is.

A witness's willingness to make an identification has often been described as a shift from absolute to relative judgments (Lindsay & Wells, 1985; Steblay, 1997; Wells, 1984, 1993). It can, of course, also be described as a simple criterion shift (Clark, 2005; Flowe & Ebbesen, 2007; Malpass & Devine, 1981; Meissner, Tredoux, Parker, & MacLin, 2005).

Our purpose is not to distinguish between decision models but, rather, to consider how the innocence risk associated with a suspect identification varies as a function of manipulations in variables that (presumably) alter the witness's decision processes. Decision processes have been studied primarily using two manipulations: varying the instructions given to the witness prior to the lineup and varying whether the lineup is presented simultaneously or sequentially. These two manipulations influence the witness's decision as to *whether* to pick or *how* to pick. Following our review of instructions and simultaneous and sequential lineups, we also will include a brief discussion regarding lineup administrator influence as to *whom* to pick.

Biased Versus Unbiased Instructions

In many (but not all) jurisdictions, police present witnesses with very standard instructions prior to a showup

or lineup. The instructions often have two key components: (1) that the perpetrator may or may not be in the lineup, and (2) that the witness is not obligated to pick anyone. Such instructions are considered to be *unbiased* with respect to the perpetrator's presence in the lineup and the responses that witnesses may give. By contrast, the instructions are considered to be *biased* if they state or imply that the perpetrator is in the lineup (i.e., "we've got the guy who stole your purse") or fail to acknowledge that *none of the above* is an appropriate response ("and we need you to identify him so that this doesn't happen to someone else").

Such biased instructions presumably increase the witness's willingness to make an identification, due either to a lowering of the decision criterion or to a change from a more stringent to a more lenient decision strategy. These shifts should, and do, lead to increases in correct, as well as false, identifications (Clark, 2005).

The change in the innocence risk depends on whether the increase is proportionally larger for correct or for false identifications. As Clark (2005) and Steblay (1997) have noted, there is considerable variation in the patterns of results shown in instruction bias experiments. For the present analysis, the results from 11 studies were chosen, with 14 biased-unbiased comparisons, on the basis of the completeness of the data reported in the original article and are shown in Table 8.

Biased instructions led to higher correct identification rates [$h_C = .129$, $t(13) = 2.489$, $p = .027$], slightly higher false identification rates [$h_F = .091$, $t(13) = 1.695$, $p = .114$], and a slight increase in innocence risk [$h_{IR} = .056$, $t(13) = .830$, $p = .422$]. It is clear from Table 8 that these analyses turn on the results of a single study by Greathouse and Kovera (in press). We must be cautious not to arbitrarily exclude (inconvenient) outlier results. However, the

Table 8
Innocence Risk (IR) With Biased and Unbiased Lineup Instructions

Study	Instructions						Effect Sizes		
	Unbiased			Biased			h_C	h_F	h_{IR}
	Correct	False	IR	Correct	False	IR			
Malpass & Devine (1980)	.455	.067	.128	.636	.100	.136	.366	.120	.024
Malpass & Devine (1981)	.826	.067	.075	.750	.155	.171	-.187	.286	.298
Cutler, Penrod, & Martens (1987a)	.460	.056	.109	.430	.113	.208	-.060	.208	.274
Cutler & Penrod (1988, Experiment 1)	.778	.046	.056	.778	.054	.065	.000	.037	.038
Fleet, Brigham, & Bothwell (1987)	.652	.097	.130	.625	.093	.130	-.056	-.014	.000
Paley & Geiselman (1989)	.400	.067	.143	.530	.150	.221	.261	.272	.203
O'Rourke, Penrod, Cutler, & Stuve (1989)	.360	.040	.100	.360	.100	.217	.000	.241	.326
Foster, Libkuman, Schooler, & Loftus (1994)	.432	.119	.216	.598	.149	.199	.334	.088	-.042
Devenport & Fisher (1997)	.265	.078	.227	.425	.125	.227	.339	.157	.000
Brewer & Wells (2006) ^a	.371	.035	.086	.427	.047	.099	.114	.061	.045
Brewer & Wells (2006) ^b	.300	.030	.091	.380	.053	.122	.169	.116	.101
Brewer & Wells (2006) ^c	.577	.065	.101	.553	.079	.125	-.048	.054	.076
Brewer & Wells (2006) ^d	.627	.043	.064	.693	.087	.112	.140	.181	.171
Greathouse & Kovera (in press)	.429	.188	.305	.643	.033	.049	.433	-.532	-.724
<i>M</i>	.495	.071	.131	.559	.096	.149	.129	.091	.056
<i>SD</i>							.194	.201	.254
<i>t</i> (13)							2.489	1.695	0.830
<i>p</i>							.027	.114	.422

Note—Results were analyzed with and without data from Greathouse and Kovera (in press). See the text for details. The data from Fleet et al. (1987) are for their neutral versus positive bias comparison. The data from Foster et al. (1994) are based on estimates from their figures. The data from Devenport and Fisher (1997) are collapsed over authority and nonauthority conditions. The data from O'Rourke et al. (1989) are estimated proportions based on their regression analysis (see their Table 1). The data from Brewer and Wells (2006) are from (a) thief lineup, high-similarity foils, (b) thief lineup, low-similarity foils, (c) waiter lineup, high-similarity foils, and (d) waiter lineup, low-similarity foils. The false identification rate of .033 for Greathouse and Kovera is estimated; their results actually showed a zero false identification rate. Our estimate of .033 is given by $1/2N$.

study was unusual in that lineup administrators were given incentives to obtain suspect identifications and were also shown various methods for obtaining them. Thus, there may have been more going on in this particular comparison of biased and unbiased instructions than just the instructions. Also, the results were unusual in several respects. The overall identification rates were near ceiling in the unbiased instructions condition; in fact, the overall identification rate was 1.0 in the target-absent condition with unbiased instructions. Even the most biased instructions could not increase the identification rate beyond 100%. Given these aspects of the results, we cautiously set them aside in assessing the effects of biased instructions. With that, the results show higher correct identification rates [$h_C = .106$, $t(12) = 2.114$, $p = .056$], higher false identification rates [$h_F = .139$, $t(12) = 5.303$, $p < .001$], and increased innocence risk [$h_{IR} = .117$, $t(12) = 3.372$, $p = .006$].

Simultaneous and Sequential Lineups

Throughout this article, we have discussed lineups as collections of individuals (or their photographs) presented simultaneously to the witness. In the last few years, a number of state and local jurisdictions have changed from simultaneous to sequential presentation of a lineup. In a sequential lineup, the lineup members are shown one at a time, and the witness makes a decision for each lineup member as he or she is presented. States that have adopted the use of sequential lineups include New Jersey, North Carolina, Virginia, and Wisconsin. Other states, most notably California and Illinois, have considered making the switch from simultaneous to sequential lineups but have not.

At a theoretical level, an argument in favor of sequential lineups is that they minimize the tendency to make identifications on the basis of relative judgments (Lindsay & Wells, 1985; Wells, 1984). To illustrate, in a simultaneous lineup, it may be clear to the witness that the person in Position 3 is the closest match to the perpetrator by far, relative to anyone else in the lineup, and the witness may identify that person on that basis. In a sequential lineup, the witness also may note that number 3 looks more similar to the perpetrator than do lineup members 1 and 2 but may pass on identifying number 3 because an even better match may be yet to come. Some (Flowe & Ebbesen, 2007; Meissner et al., 2005) have argued that sequential lineups may not suppress relative judgments; they may simply induce witnesses to use a higher criterion.

Theoretical mechanisms aside, the argument favoring sequential lineups is that they reduce the likelihood of false identifications (Cutler & Penrod, 1995; Lindsay & Wells, 1985; Steblay, Dysart, Fulero, & Lindsay, 2001). The argument against their use is that they also reduce the likelihood of correct identifications (Malpass, 2006).

Because simultaneous lineups are currently the standard, it seems natural to discuss the effect of sequential lineups relative to simultaneous lineups. Consequently, the effect sizes will be negative to the extent that both correct and false identification rates are lower for sequential lineups, relative to simultaneous lineups. The results are given in Table 9.

Sequential lineups showed lower false identification rates [$h_F = -.200$, $t(28) = -3.979$, $p < .001$] and lower correct identification rates [$h_C = -.252$, $t(28) =$

Table 9
Correct and Mistaken Identification Rates and Innocence Risk (IR) in Simultaneous and Sequential Lineups

Study	Simultaneous			Sequential			Effect Sizes		
	Correct	False	IR	Correct	False	IR	h_C	h_F	h_{IR}
Lindsay & Wells (1985)	.583	.433	.426	.500	.167	.250	-.167	-.594	-.375
Cutler & Penrod (1988)	.470	.072	.133	.410	.035	.079	-.121	-.167	-.177
Melara, DeWitt-Rickards, & O'Brien (1989)	.250	.167	.400	.125	.042	.251	-.325	-.429	-.320
Parker & Ryan (1993)	.417	.250	.375	.083	.083	.500	-.820	-.463	.253
Sporer (1993)	.444	.120	.213	.389	.065	.143	-.112	-.192	-.184
Lindsay, Pozzulo, Craig, Lee, & Corber (1997)	.550	.057	.094	.620	.042	.063	.142	-.069	-.116
Lindsay et al. (1991b, Experiment 2)	.567	.333	.370	.467	.067	.125	-.201	-.707	-.585
Lindsay et al. (1991a, Experiment 1)	.567	.200	.261	.467	.054	.104	-.201	-.458	-.416
Kneller, Memon, & Stevenage (2001)	.611	.102	.143	.500	.037	.069	-.224	-.263	-.244
MacLin, Meissner, & Zimmerman (2005, p+p)	.400	.105	.208	.333	.067	.168	-.139	-.136	-.103
MacLin et al. (2005, comp)	.467	.083	.151	.267	.038	.125	-.419	-.192	-.076
Memon & Gabbert (2003)	.467	.089	.160	.167	.017	.092	-.663	-.344	-.207
Haw & Fisher (2004, hc)	.633	.300	.322	.500	.067	.118	-.269	-.636	-.506
Haw & Fisher (2004, lc)	.600	.033	.052	.467	.133	.222	-.267	.381	.521
Carlson, Gronlund, & Clark (2008, Experiment 1)	.717	.021	.028	.571	.047	.076	-.306	.146	.222
Carlson et al. (2008, Experiment 2b)	.706	.644	.477	.462	.326	.414	-.501	-.648	-.127
Carlson et al. (2008, Experiment 2m)	.426	.303	.416	.235	.375	.615	-.410	.152	.401
Carlson et al. (2008, Experiment 2f)	.314	.163	.342	.408	.200	.329	.196	.096	-.028
Smith, Lindsay, Pryke, & Dysart (2001, sr)	.458	.066	.126	.233	.017	.068	-.479	-.258	-.198
Smith et al. (2001, cr)	.453	.154	.254	.302	.148	.329	-.313	-.017	.165
Pozzulo & Marciniak (2006, ac)	.233	.072	.236	.233	.067	.223	.000	-.020	-.031
Pozzulo & Marciniak (2006, nc)	.667	.083	.111	.467	.061	.116	-.407	-.085	.016
Rose, Bull, & Vrij (2005)	.750	.063	.077	.458	.049	.097	-.608	-.061	.071
Wilcock, Bull, & Vrij (2005)	.667	.104	.135	.625	.028	.043	-.088	-.321	-.335
Douglass & McQuiston-Surrett (2006)	.875	.156	.151	.625	.154	.198	-.595	-.006	.124
Greathouse & Kovera (in press, ba-u)	.429	.188	.305	.563	.067	.106	.269	-.373	-.507
Greathouse & Kovera (in press, ba-b)	.643	.033	.049	.500	.067	.118	-.290	.158	.255
Greathouse & Kovera (in press, nba-u)	.467	.143	.234	.786	.133	.145	.675	-.029	-.229
Greathouse & Kovera (in press, nba-b)	.857	.333	.280	.571	.214	.273	-.653	-.268	-.016
<i>M</i>	.541	.168	.225	.425	.099	.188	-.252	-.200	.095
<i>SD</i>							.313	.271	.273
<i>t</i> (27)							-4.330	-3.979	-1.871
<i>p</i>							.000	.000	.072

Note—For MacLin et al. (2005), p+p = paper-and-pencil and comp = computer; for Haw and Fisher (2004), hc and lc denote high contact and low contact between lineup administrator and witness; for Carlson et al. (2008), b, m, and f refer to biased, mixed, and fair lineup composition; for Smith et al. (2001), sr and cr denote same-race and cross-race; for Pozzulo and Marciniak (2006), ac = appearance changed and nc = appearance not changed; for Greathouse and Kovera (in press), ba and nba denote blind and not-blind administration of the lineup, and u and b denote unbiased and biased instructions.

-4.330, $p < .001$]. Overall, innocence risk was lower in sequential lineups than in simultaneous lineups [$h_{IR} = -.095$, $t(28) = -1.871$, $p = .072$]. The pattern of lower correct and lower false identification rates was shown in 20 of the 29 simultaneous-sequential comparisons, and the innocence risk was lower for sequential lineups in 20 of 29 comparisons.

By the index of innocence risk, sequential lineups "win" most of the time, but not all of the time. In nine cases, the innocence risk was higher for sequential lineups than for simultaneous lineups. As some jurisdictions have changed to sequential lineups, whereas others have not, there is currently considerable controversy surrounding sequential lineups (see, e.g., Schacter et al., 2008); thus, it is important to understand what might have produced the nine negative results.

In some cases, the sequential lineup disadvantage may have been due to the fact that the false identification rate for the innocent suspect was very low for simultaneous lineups: Carlson, Gronlund, and Clark (2008), Haw and Fisher (2004), Greathouse and Kovera (in press), and Rose, Bull, and Vrij (2005) obtained false identification

rates from .021 to .063 for the simultaneous lineup condition. This floor effect, combined with the typical reduction in the correct identification rate, is a good recipe for a sequential lineup disadvantage.

In other cases, the sequential disadvantage may arise from a fundamental difference between simultaneous and sequential lineups. In a simultaneous lineup, if the witness lowers his or her decision criterion, the likelihood of both correct and false identifications should increase (as is shown in the biased-unbiased instruction comparisons in Table 8). However, in a sequential lineup, a very low criterion may cause the witness to identify a foil before the suspect is ever presented, thus reducing correct and false identification rates. For example, if the suspect is in Position 4, a witness with a very low criterion may pick a foil and never get to Position 4.

Of course, for this tentative explanation to have any traction, the effect must be larger for guilty suspects than for innocent suspect, and there must be some reason to think that witnesses lowered their criteria in at least some of the studies that showed a sequential lineup disadvantage. Regarding the first component, the effect may be

larger for guilty suspects than for innocent suspects when the correct identification rate is high and the false identification rate is low in the simultaneous lineup condition (as it was in the studies listed above). The correct identification rate would have far to drop, but the false identification rate would not.

Regarding the latter component, there is evidence (which we will not review here) to suggest that witnesses have lower decision criteria (1) when they are given biased instructions or are pressed to make an identification (see Table 8), (2) when there is a cross-racial relationship between a witness and a target (Meissner & Brigham, 2001), and (3) for sequential lineups when there is a mix of good and poor foils (Clark & Davey, 2005). Indeed, the negative cases showed many of these conditions—specifically, biased instructions and pressure to make an identification (Greathouse & Kovera, *in press*; Haw & Fisher, 2004), a mix of good and poor foils (Carlson et al., 2008), and a cross-racial relationship between witness and target (Smith, Lindsay, Pryke, & Dysart, 2001). This analysis of criterion placement in simultaneous and sequential lineups may also apply to other results not included in our analyses. For example, some have suggested that sequential lineups are not advisable for child witnesses (Pozzulo & Lindsay, 1998), who may also tend to have lower decision criteria than do adult witnesses (see Dekle, Beal, Elliot, & Huneycutt, 1996).

We should note that the conclusions drawn from our analyses differ somewhat from those of a previous analysis by Steblay et al. (2001). They also showed decreases in correct and false identifications for sequential lineups but argued that when only more “realistic” studies were considered, the decrease in correct identifications for sequential lineups became sufficiently small as to be “possibly nonexistent” (p. 468) and “not a serious problem” (p. 470). All of the studies considered here used the staged crime procedure, and the decrease in correct identifications was robust and consistent.

Suspect-Specific Decision Biases

Lineup administrators may influence witnesses not only as to whether to pick, but also as to whom to pick. The influence may be deliberate but may also be inadvertent and quite subtle (i.e., “I noticed you paused on number 3”; Nettles, Nettles, & Wells, 1996). We refer to this as suspect-specific bias because the lineup administrator’s behavior does not simply increase the willingness to pick someone but, rather, guides the witness toward an identification of the suspect. There are only two studies—one by Haw and Fisher (2004) and a recent study by Greathouse and Kovera (*in press*)—that have examined this kind of suspect-specific bias with both target-present and target-absent lineups. The two studies were designed to simulate the real-world situation in which the police officer who administers the lineup knows the position of the suspect and is highly motivated to obtain an identification of that suspect.

The two studies used the same staged crime video and the same lineups but differed in the specifics of the com-

parison conditions. The differences in the studies make for a rather long story, but the bottom line is that each study compared a condition in which suspect-specific bias was more likely to occur with a condition in which suspect-specific bias was less likely to occur. The two studies showed similar patterns of results for simultaneous lineups, in that correct and false identification rates increased when suspect-specific bias was easier to convey and carry out. However, the proportional increases were not consistent across the two studies: Whereas Haw and Fisher (2004) showed an increase in innocence risk due to suspect-specific bias, Greathouse and Kovera (*in press*) showed a slight decrease in innocence risk. The results for sequential lineups were also inconsistent and also showed opposite patterns for innocence risk (increased innocence risk for Greathouse and Kovera [*in press*], but decreased innocence risk for Haw and Fisher [2004]).

Researchers have made strong recommendations that lineups be administered using a blind procedure to minimize administrator influence and suspect-specific bias (Wells, 1988; Wells et al., 1998). This is a reasonable recommendation, given the extensive literature on experimenter bias (see Rosenthal, 2002, for a review); however, more research is needed in order that we may understand more fully how administrator influence affects correct and false identification rates, not only at the level of the data, but also in terms of the underlying mechanisms.

Summary and Synthesis

In this review, we have examined three aspects of police procedure that likely affect witness decision processes. Both biased instructions and simultaneous lineup presentation showed increases in correct and false identification rates, and both showed an increase in innocence risk, relative to unbiased instructions and sequential lineups. The effect sizes for correct and false identification increases for simultaneous–sequential comparisons were twice as large as those for biased–unbiased instructions comparisons. However, the effect sizes for innocence risk were about the same.

We noted several cases in which the sequential lineup advantage was not obtained. In some cases, this appears to have been due to floor effects on the false identification rate. This parallels an argument made by Clark (2005) regarding biased–unbiased instruction experiments. In some cases, the overall identification rates in target-present lineups were quite high, placing a cap on the correct identification rate, giving the appearance that biased instructions do not increase correct identification rates.

The least understood phenomena reviewed in this section concern the effects of suspect-specific biases. To date, there have been only two experiments (Greathouse & Kovera, *in press*; Haw & Fisher, 2004) in which these biases have been examined in both target-present and target-absent lineups, and they have not presented a consistent pattern of results. Such results should present an interesting challenge for models, because the bias is not about whether to identify (which can be modeled as a simple

criterion shift or change in decision strategy) but, rather, is about whom to identify.

Prior Probability That the Suspect is Guilty

Throughout this article, we have calculated innocence risk under the simplified assumption of equal prior probabilities. We asserted earlier in this article that variation in the prior probabilities would not affect the patterns of results. However, it is useful to show how innocence risk varies as a function of the prior probability that the suspect is guilty. Again, these prior probabilities represent the relative frequencies with which police present witnesses with suspects who are guilty versus suspects who are innocent.

Figure 2 shows innocence risk as a function of the prior probability of guilt, comparing better and poorer memory conditions. Innocence risk was calculated using the average correct and false identification rates, varying the prior probability of the suspect's guilt from 0 to 1. The other comparisons showed similar patterns, so only the results for the better–poorer memory comparison are shown in the figure. Clearly, the relationships do not change as the prior probability of the suspect's guilt is varied. The differences are small, of course, when the prior probabilities are near 0 or 1, and the differences tend to be largest when the prior probability of the suspect's guilt is between (roughly) .2 and .5.

IMPLICATIONS FOR THEORY AND LEGAL DECISION MAKING

Four main findings emerge from this review. (1) Limitations in memory were shown to produce a substantial increase in innocence risk. This increase in innocence risk was due primarily to the decrease in correct identification

rates; false identification rates were more variable and did not show a significant increase when averaged over 13 comparisons. (2) The effects of memory limitations may sometimes be hidden by compensatory criterion shifts. This result was particularly evident in studies in which the retention interval was varied. (3) Police procedures that produce increases in false identification rates generally also produce increases in correct identification rates. This pattern was shown for biased instructions, biased lineup composition, and simultaneous–sequential lineup comparisons. In all three cases, the increases in false identifications were proportionally larger than the increases in correct identifications, thus increasing the innocence risk. (4) There was one notable exception to this pattern: Showups, as compared with lineups, showed higher false identification rates and (slightly) lower correct identification rates. These results have important implications for the development of memory and decision models for eyewitness identification and for legal decision making based on eyewitness identification evidence.

Memory and Decision Processes in Eyewitness Identification

Memory is less accurate when exposure durations are shorter, when stress is higher, and when retention intervals are longer. This by itself is not surprising. What is surprising, however, is that the result is carried almost entirely by decreases in correct identification rates. Also, although these results may seem obvious in hindsight, there has been some unresolved controversy, particularly with respect to the effects of high stress (Yuille & Cutshall, 1986) and longer retention intervals (Dysart & Lindsay, 2007; McCloskey & Egeth, 1983). The controversy regarding retention interval, we would argue, may be due in part to differences in results for target-present and target-absent lineups and from the fact that, in some cases, decreases in correct identification rates may have been hidden by criterion shifts. Peering through those complications, there appears to be little controversy: Accuracy decreases and innocence risk increases with the passage of time.

The question remains, however, as to why witnesses should become more willing to make an identification with the passage of time. One speculative answer is that, over time, witnesses not only may lose information about the perpetrator; but also may lose information about their conditions of observation. In other words, they may forget that they did not get a very good look at the perpetrator. The prospect of criterion shifts with variation in memory strength has implications beyond the eyewitness identification literature, of course, as the debate over criterion shifts and differentiation continues (Criss, 2006; Singer & Wixted, 2006; Stretch & Wixted, 1998).

It may seem unsurprising that both correct and false identification rates increase when lineups are poorly constructed and when witnesses are “nudged” through biased instructions to make identifications. The instruction effects may seem unsurprising, of course, because they have been shown countless times in yes–no recognition memory experiments and criterion shifts are a cornerstone of signal detection theory. However, contrary to

Figure 2. Innocence risk as a function of the prior probability of the suspect's guilt for better versus poorer memory conditions.

the results presented here, there has been a widely held view that correct identification rates do not increase with low-similarity foils (Lindsay & Pozzulo, 1999), do not increase with biased instructions (Stebly, 1997), and are not higher for simultaneous lineups than for sequential lineups (Stebly et al., 2001).

What is the basis of such conclusions? First, there may be a historical primacy effect in the literature. Early experiments did, in fact, show that low-similarity foils (Lindsay & Wells, 1980), biased instructions (Malpass & Devine, 1981), and simultaneous lineups (Lindsay & Wells, 1985) had only slightly higher correct identification rates but much higher false identification rates, as compared with high-similarity foils, unbiased instructions, and sequential lineups. The no-increase view may have become so well accepted on the basis of these initial results that subsequent results showing increases in the correct identification rate may simply have been underappreciated. Also, at the level of individual experiments, the increases in correct identification rates are indeed often statistically nonsignificant, and these nonsignificant results may have come to be viewed as nonexistent. Eyewitness identification experiments, which typically generate one observation per participant, have notoriously low statistical power, which can produce many statistically nonsignificant results; only through meta-analytic aggregation over several studies does the trend become clear.

We emphasize, however, that the increases in correct identification rates are generally proportionally smaller than the increases in false identification rates. Thus, although correct and false identification rates increase together, innocence risk does increase due to the inclusion of low-similarity foils, biased lineup instructions, and simultaneous lineups. We emphasize also that in some studies, the correct identification rate clearly did not increase, although the overall identification rate did increase. Such results are not easily explained by simple criterion shifts (see Clark, 2005, for details).⁹

One last finding from our analyses is of particular interest to the controversy concerning simultaneous and sequential lineups. Generally, sequential lineups show lower innocence risk than do simultaneous lineups. The present analyses suggest conditions under which this advantage for sequential lineups may not be found: If the false identification rates are already quite low or if witnesses are pressed or highly motivated to make an identification, the innocence risk for sequential lineups may be higher than that for simultaneous lineups.

Legal Decision Making

What are the implications of our analyses for decision making within the criminal justice system? Here, the focus will not be on the decisions made by witnesses but, rather, on the decisions made by those whose task is to interpret or evaluate the decisions made by witnesses. This would include police officers, prosecuting and defense attorneys, judges, and jurors. We also will briefly consider decision making by those who would establish policy relevant to eyewitness identification evidence.

Direct versus indirect causes of false identification and wrongful conviction. Our analyses suggest that the risk of wrongful conviction increases due to factors that have a clear causal link to false identifications, but also due to factors that appear to have little causal connection to false identifications. The first case describes a clear causal pathway in that A (police pressure) causes B (a false identification) that causes C (a wrongful conviction). The second pathway is less direct. Stress, the limited opportunity to observe, and the loss of information over time do not cause substantial increases in the false identification rate. However, because they are associated with large decreases in the correct identification rate, the proportion of suspect identifications that are false identifications increases, thus increasing the innocence risk.

This suggests a pathway from false identification to wrongful conviction without any clear cause of the false identification. Because there is no identifiable cause for the false identification, this pathway may be largely invisible to decision-makers within the criminal justice system—that is, lawyers, judges, and juries. Wells and Loftus (2003) made this point, noting that limitations in perception and memory “might not matter much to the jury when they deliberate because they fail to explain why the witness picked the *suspect* (emphasis added) out of the lineup (rather than) a filler.” The bad memory explanation of false identification may appear unsatisfying to the jury because it fails to provide a cause of the false identification, and its role in wrongful convictions requires jurors to consider not only the likelihood of a false identification, but also the likelihood of a correct identification. We will expand upon this issue for jurors in the next section, but the issues likely apply for other legal decision makers as well.

Juror reasoning. Our analysis of innocence risk is closely tied to Bayes’s theorem. Jurors’ deviation from Bayesian reasoning is, of course, well known (Schum & Martin, 1982). Much of the research on jurors’ understanding about eyewitness identification evidence focuses on juror knowledge (see, e.g., Benton, Ross, Bradshaw, Thomas, & Bradshaw, 2006; Deffenbacher & Loftus, 1982; Kassin & Barndollar, 1992; Schmechel, O’Toole, Easterly, & Loftus, 2006). We suggest that in addition to gaps in knowledge, there may also be errors in reasoning and hypothesis testing.

Jurors may evaluate the identification evidence as a test of a single hypothesis: specifically, *If the defendant is guilty, the witness should identify him. The witness did identify him. Therefore, the defendant is guilty.* This reasoning is, of course, based on an invalid deductive syllogism (affirming the consequent) and fails to consider the alternative hypothesis that the defendant is innocent. People often fail to see the error in affirming the consequent (Rumain, Connell, & Braine, 1983) and have difficulty considering alternative hypotheses (Nickerson, 1998; Wason, 1966).

Jurors may consider the alternative hypothesis that the defendant is innocent only if there is a clear cause for his false identification. The reasoning here is, *If the defendant*

didn't do it, then why did the witness identify him? Poor lineup composition and suggestive police procedures have a clear causal connection to misidentification, whereas memory limitations do not. A juror may reason that without some causal agent, the likelihood of a false identification is quite small—for example, one in six for a six-person lineup. However, if the conditions of observation were very poor, the probability of a correct identification would also be about one in six, so that the identification would have very little probative value and the innocence risk would be quite high.

The flip side of this is that many variables that *do* have a clear causal relationship to false identification also produce an increase in the likelihood of a correct identification. Jurors may readily accept the fact that both likelihoods increase when the lineup is biased or when police nudge the witness to make an identification. However, unless they know that false identification rates increase more than correct identification rates, they may ignore evidence about police bias, assuming it to be a “wash.” Indeed, the results of studies in which people have rendered verdicts in mock trials show the rate of guilty verdicts to be unaffected by evidence of biased police procedures (Cutler, Penrod, & Dexter, 1990; Devenport, Stinson, Cutler, & Kravitz, 2002).

Policy and reform of identification procedures. The DNA exoneration cases have prompted an effort to reform the procedures police use to obtain identification evidence. Such reforms require a comparison between Procedure *X* and Procedure *Y*. The simplest cases are those for which *X* dominates *Y*, either by decreasing mistaken identification rates without decreasing correct identifications or by increasing correct identification rates without increasing mistaken identification rates. Such cases are “no-brainers” for policy decisions, because they produce only benefits and no costs.

Recommendations for police procedures have, at times, rested on the appearance of dominated alternatives. Steblay (1997) concluded that biased instructions serve only to increase false identification rates, with no change in correct identification rates. Steblay et al. (2001) argued that the decrease in correct identifications due to the use of sequential lineups was so small that it posed no problem. Some of the early experiments comparing description-based and suspect-based procedures for selecting foils showed increases in correct identification rates, with no change in false identification rates (Juslin, Olsson, & Winman, 1996; Wells, Rydell, & Seelau, 1993). Such clear-cut results would make the adoption of unbiased instructions, sequential lineups, and description-based foil selection easy decisions. However, the analyses presented here show that such policy decisions are not so clear-cut. The only analysis that showed dominance of one procedure over another was for showup–lineup comparisons. Even in the case of showups, the policy implications are not straightforward. The argument in favor of the showup procedure is that a showup can be conducted right away (Calandra & Carey, 2006), whereas the presentation of a lineup may impose a delay. Data supporting this argument

are shown in Table 7. Yarmey et al.'s (1996) results showed lower innocence risk for a showup conducted immediately (.202) than for a lineup conducted the next day (.302). This trade-off between time and the better procedure may disappear, however, as new procedures are developed that allow police officers to compose and present a photographic lineup on laptop computers (MacLin, Meissner, & Zimmerman, 2005).¹⁰

Unbiased instructions, sequential lineups, and the use of plausible foils were all shown to reduce innocence risk, because reductions in false identification rates were proportionally larger than reductions in correct identification rates, and may still be recommended procedures based on that index. However, since there are few cases in which one procedure dominates another, policy decisions must address the question of how much cost (decrease in correct identifications) is acceptable in exchange for how much benefit (decrease in false identifications). The answer to this question goes beyond the numbers presented here and requires a careful analysis of costs, benefits, and social values (see Malpass, 2006).

We should be clear in noting the limitations of innocence risk as an index of eyewitness identification evidence, particularly with regard to policy decisions. We calculated innocence risk as the probability that the identification of a suspect will be the false identification of an innocent person. We chose the conditional probability because it is fairly intuitive and because it is based on proportional, rather than absolute, change. Thus, a .10 increase added to a .10 false identification rate is quite different from a .10 increase added to a .40 correct identification rate. However, one problem is that innocence risk can be made to be very small by simply driving the overall identification rate toward zero. For example, with a false identification rate of .001 and a correct identification rate of .10, the innocence risk is .01. Although the innocent are almost never misidentified, the vast majority of bad guys walk away as well.

Other measures of innocence risk could also be calculated. For example, innocence risk could be calculated as an odds ratio (or log odds ratio) or as a simple difference. Wells and Olson (2002) developed a statistical test for target-present–target-absent comparisons that calculates a *z* score from Cohen's *h*. As Nelson (2005) has pointed out, there are many ways to ask useful questions within a Bayesian framework.

FUTURE RESEARCH DIRECTIONS

Our analyses suggest several directions for future research. First, general memory limitations, and their relevance to wrongful convictions, may be understudied and underappreciated. The effects of variables such as short exposure durations, stress, and the passage of time may seem so obvious that they are not worthy of study. Not only are they not obvious; they have, in fact, been a source of controversy. Also, even if these variables have little effect on false identification rates, their effect on correct identification rates and innocence risk makes them critical to legal decision making based on identification evidence.

One area of future research concerns the role of criterion shifts and differentiation as a function of memory strength in eyewitness identification.

Our analyses also examined several aspects of police procedure. In general, procedures that increased false identification rates also increased correct identification rates. On the one hand, this general pattern would seem to be easily accounted for in terms of relative similarity or simple criterion shifts. On the other hand, simple accounts often fall short when models are fit to data. There has been some progress in terms of fitting decision models to data (Clare & Lewandowsky, 2004; Clark, 2003), but this work has just scratched the surface.

Finally, our review has revealed some gaps in the research literature. In particular, there is a need for more research in which suspect-specific memory alterations due to witnesses exposure to postevent information are examined and for more research in which suspect-specific biases that bias not only the decision of whether to pick, but also the decision of whom to pick are examined.

AUTHOR NOTE

This research was supported by National Science Foundation Grant SES 0214373. The authors thank Bobbie Spellman, Robert Nosofofsky, and Julieta Stepanyan for insightful comments and discussion, and the anonymous reviewers for their thoughtful and helpful reviews. Correspondence concerning this article should be addressed to S. E. Clark, Psychology Department, University of California, Riverside, CA 92521 (e-mail: clark@ucr.edu).

REFERENCES

- AMERICAN LAW INSTITUTE (1975). *A model code of pre-arraignment procedure*. Philadelphia: Author.
- BENTON, T. R., ROSS, D. F., BRADSHAW, E., THOMAS, W. N., & BRADSHAW, G. S. (2006). Eyewitness memory is still not common sense: Comparing jurors, judges and law enforcement to eyewitness experts. *Applied Cognitive Psychology*, **20**, 115-129.
- BOYCE, M. A., LINDSAY, D. S., & BRIMACOMBE, C. A. E. (2008). Investigating investigators: Examining the impact of eyewitness identification evidence on student-investigators. *Law & Human Behavior*, **32**, 439-453.
- BREWER, N., CAON, A., TODD, C., & WEBER, N. (2006). Eyewitness identification accuracy and response latency. *Law & Human Behavior*, **30**, 31-50.
- BREWER, N., & WELLS, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, **12**, 11-30.
- BROWN, E., DEFFENBACHER, K., & STURGILL, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology*, **62**, 311-318.
- CALANDRA, D., & CAREY, J. E. (2006). *Field guide for the California peace officers legal sourcebook*. Sacramento: California District Attorneys Association.
- CARLSON, C., GRONLUND, S. D., & CLARK, S. E. (2008). Lineup composition, suspect position, and the sequential lineup advantage. *Journal of Experimental Psychology: Applied*, **14**, 118-128.
- CHALLIS, B. H., & ROEDIGER, H. L., III (1993). The effect of proportion overlap and repeated testing on primed word fragment completion. *Canadian Journal of Experimental Psychology*, **47**, 113-123.
- CHAN, J. C. K., & McDERMOTT, K. B. (2007). The testing effect in recognition memory: A dual process account. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **33**, 431-437.
- CLARE, J., & LEWANDOWSKY, S. (2004). Verbalizing facial memory: Criterion effects in verbal overshadowing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **4**, 739-755.
- CLARK, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology*, **17**, 629-654.
- CLARK, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law & Human Behavior*, **29**, 395-424.
- CLARK, S. E., & DAVEY, S. L. (2005). The target-to-foils shift in simultaneous and sequential lineups. *Law & Human Behavior*, **29**, 151-172.
- CLARK, S. E., & GRONLUND, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, **3**, 37-60.
- CLARK, S. E., HOWELL, R. T., & DAVEY, S. L. (2008). Regularities in eyewitness identification. *Law & Human Behavior*, **32**, 187-218.
- CLARK, S. E., & TUNNICLIFF, J. L. (2001). Selecting lineup foils in eyewitness identification experiments: Experimental control and real-world simulation. *Law & Human Behavior*, **25**, 199-216.
- CLARK, S. E., & WELLS, G. L. (2008). On the diagnosticity of multiple-witness identifications. *Law & Human Behavior*, **32**, 406-422.
- COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- CONN, V. S., VALENTINE, J. C., COOPER, H. M., & RANTZ, M. J. (2003). Grey literature in meta-analyses. *Nursing Research*, **52**, 256-261.
- COOPER, H. (2003). Editorial. *Psychological Bulletin*, **129**, 3-9.
- CRISS, A. H. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory & Language*, **55**, 461-478.
- CUTLER, B. L., & PENROD, S. D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. *Journal of Applied Psychology*, **73**, 281-290.
- CUTLER, B. L., & PENROD, S. D. (1995). *Mistaken identification: The eyewitness, psychology, and the law*. New York: Cambridge University Press.
- CUTLER, B. L., PENROD, S. D., & DEXTER, H. R. (1990). Juror sensitivity to eyewitness identification evidence. *Law & Human Behavior*, **14**, 185-191.
- CUTLER, B. L., PENROD, S. D., & MARTENS, T. K. (1987a). Improving the reliability of eyewitness identification: Putting context into context. *Journal of Applied Psychology*, **72**, 629-637.
- CUTLER, B. L., PENROD, S. D., & MARTENS, T. K. (1987b). The reliability of eyewitness identification: The role of system and estimator variables. *Law & Human Behavior*, **11**, 233-258.
- CUTLER, B. L., PENROD, S. D., O'ROURKE, T. E., & MARTENS, T. K. (1986). Unconfounding the effects of contextual cues on eyewitness identification accuracy. *Social Behaviour*, **1**, 113-134.
- DEFFENBACHER, K. A., BORNSTEIN, B. H., & PENROD, S. D. (2006). Mugshot exposure effects: Retroactive interference, source confusion, and unconscious transference. *Law & Human Behavior*, **30**, 287-307.
- DEFFENBACHER, K. A., & LOFTUS, E. F. (1982). Do jurors share a common understanding concerning eyewitness behavior? *Law & Human Behavior*, **6**, 15-30.
- DEKLE, D. J. (1997). Testing delays resulting in increased identification accuracy in line-ups and show-ups. *Journal of Offender Rehabilitation*, **25**, 35-49.
- DEKLE, D. J., BEAL, C. R., ELLIOT, R., HUNEYCUTT, D. (1996). Children as witnesses: A comparison of lineup versus showup identification methods. *Applied Cognitive Psychology*, **10**, 1-12.
- DEVENPORT, J. L., & FISHER, R. P. (1997). The effects of authority and social influence on eyewitness suggestibility and person recognition. *Journal of Police & Criminal Psychology*, **11**, 35-40.
- DEVENPORT, J. L., STINSON, V., CUTLER, B. L., & KRAVITZ, D. A. (2002). How effective are the cross-examination and expert testimony safeguards? Jurors' perceptions of the suggestiveness and fairness of biased lineup procedures. *Journal of Applied Psychology*, **87**, 1042-1054.
- DOUGLASS, A. B., & McQUISTON-SURRETT, D. (2006). Post-identification feedback: Exploring the effects of sequential photospreads and eyewitnesses' awareness on the identification task. *Applied Cognitive Psychology*, **20**, 991-1007.
- DYSART, J. E., & LINDSAY, R. C. L. (2007). Show-up identifications: Suggestive technique or reliable method? In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology: Vol II. Memory for people* (pp. 361-376). Mahwah, NJ: Erlbaum.

- EBBESEN, E. B., & KONECNI, V. J. (1996). Eyewitness memory research: Probative versus prejudicial value. *Expert Evidence*, *5*, 2-28.
- ESTES, W. K. (1997). Processes of memory loss, recovery, and distortion. *Psychological Review*, *104*, 148-169.
- FEDERAL RULES OF EVIDENCE (2004). Committee on the Judiciary, 108th Congress, House of Representatives. Washington, DC: U.S. Government Printing Office.
- FERKENHOFF, E. (2002, July 9). Man held mistakenly is dropped as suspect. *Chicago Tribune*, p. 3.
- FLEET, M. L., BRIGHAM, J. C., & BOTHWELL, R. K. (1987). The confidence-accuracy relationship: The effects of confidence assessment and choosing. *Journal of Applied Social Psychology*, *17*, 171-187.
- FLOWE, H. D., & EBBESEN, E. B. (2007). The effect of lineup member similarity on recognition accuracy in simultaneous and sequential lineups. *Law & Human Behavior*, *31*, 33-52.
- FOSTER, R. A., LIBKUMAN, T. M., SCHOOLER, J. W., & LOFTUS, E. W. (1994). Consequentiality and eyewitness person identification. *Applied Cognitive Psychology*, *8*, 107-121.
- GLANZER, M., & ADAMS, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *11*, 8-20.
- GLANZER, M., & ADAMS, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *1*, 5-16.
- GONZALEZ, R., ELLSWORTH, P. C., & PEMBROKE, M. (1993). Response biases in lineups and showups. *Journal of Personality & Social Psychology*, *64*, 525-537.
- GREATHOUSE, S. M., & KOVERA, M. B. (in press). Instruction bias and lineup presentation moderate the effects of administrator knowledge on eyewitness identification. *Law & Human Behavior*.
- GRONLUND, S. D. (2005). Sequential lineup advantage: Contributions of distinctiveness and recollection. *Applied Cognitive Psychology*, *19*, 23-37.
- GROSS, S. R., JACOBY, K., MATHESON, D. J., MONTGOMERY, N., & PATIL, S. (2005). Exonerations in the United States 1989 through 2003. *Journal of Criminal Law & Criminology*, *95*, 523-560.
- HAW, R. M., & FISHER, R. P. (2004). Effects of administrator-witness contact on eyewitness identification accuracy. *Journal of Applied Psychology*, *89*, 1106-1112.
- HUFF, C. R., RATTNER, A., & SAGARIN, E. (1996). *Convicted but innocent*. Thousand Oaks, CA: Sage.
- INNOCENCE PROJECT. Available at www.innocenceproject.org.
- JUSLIN, P., OLSSON, N., & WINMAN, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 1304-1316.
- KASSIN, S. M., & BARNDOLLAR, K. A. (1992). The psychology of eyewitness testimony: A comparison of experts and prospective jurors. *Journal of Applied Social Psychology*, *22*, 1241-1249.
- KNELLER, W., MEMON, A., & STEVENAGE, S. (2001). Simultaneous and sequential lineups: Decision processes of accurate and inaccurate eyewitnesses. *Applied Cognitive Psychology*, *15*, 659-671.
- KRAFFKA, C., & PENROD, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality & Social Psychology*, *49*, 58-69.
- LAUGHERY, R. L., FESSLER, P. K., LENOROVITZ, D. R., & YOBlick, D. A. (1974). Time delay and similarity effects in facial recognition. *Journal of Applied Psychology*, *59*, 490-496.
- LINDSAY, R. C. L., LEA, J. A., & FULFORD, J. A. (1991). Sequential lineup presentation: Technique matters. *Journal of Applied Psychology*, *76*, 741-745.
- LINDSAY, R. C. L., LEA, J. A., NOSWORTHY, G. J., FULFORD, J. A., HECTOR, J., LEVAN, V., & SEABROOK, C. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology*, *76*, 796-802.
- LINDSAY, R. C. L., MARTIN, R., & WEBBER, L. (1994). Default values in eyewitness descriptions: A problem for the match-to-description lineup foil selection strategy. *Law & Human Behavior*, *18*, 527-541.
- LINDSAY, R. C. L., & POZZULO, J. D. (1999). Sources of eyewitness identification error. *International Journal of Law & Psychiatry*, *22*, 347-360.
- LINDSAY, R. C. L., POZZULO, J. D., CRAIG, W., LEE, K., & CORBER, S. (1997). Simultaneous lineups, sequential lineups, and showups: Eyewitness identification decisions of adults and children. *Law & Human Behavior*, *21*, 391-404.
- LINDSAY, R. C. L., ROSS, D. F., READ, J. D., TOGLIA, M. P. (EDS.) (2007). *Handbook of eyewitness psychology: Vol. II. Memory for people*. Mahwah, NJ: Erlbaum.
- LINDSAY, R. C. L., WALLBRIDGE, H., & DRENNAN, D. (1987). Do the clothes make the man? An exploration of the effect of lineup attire on eyewitness identification accuracy. *Canadian Journal of Behavioural Science*, *19*, 463-478.
- LINDSAY, R. C. L., & WELLS, G. L. (1980). What price justice? Exploring the relationship of lineup fairness to identification accuracy. *Law & Human Behavior*, *4*, 303-313.
- LINDSAY, R. C. L., & WELLS, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, *70*, 556-564.
- LOFTUS, E. F. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.
- LOFTUS, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*, 361-366.
- LOFTUS, E. F., MILLER, D. G., & BURNS, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning & Memory*, *4*, 19-31.
- LUUS, C. A. E., & WELLS, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law & Human Behavior*, *15*, 43-57.
- MACLIN, O. H., MEISSNER, C. A., & ZIMMERMAN, L. A. (2005). PC_Eyewitness: A computerized framework for the administration and practical application of research in eyewitness psychology. *Behavior Research Methods*, *37*, 324-334.
- MALPASS, R. S. (2006). A policy evaluation of simultaneous and sequential lineups. *Psychology, Public Policy, & Law*, *12*, 394-418.
- MALPASS, R. S., & DEVINE, P. G. (1980). Realism and eyewitness identification research. *Law & Human Behavior*, *4*, 347-358.
- MALPASS, R. S., & DEVINE, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology*, *66*, 482-489.
- MAULDIN, M. A., & LAUGHERY, K. R. (1981). Composite production effects on subsequent facial recognition. *Journal of Applied Psychology*, *66*, 351-357.
- MCCLELLAND, J. L., & CHAPPELL, M. (1998). Familiarity breeds differentiation: A subjective-likelihood model approach to the effects of experience in recognition memory. *Psychological Review*, *105*, 724-760.
- MCCLOSKEY, M., & EGETH, H. E. (1983). What can a psychologist tell a jury? *American Psychologist*, *38*, 550-563.
- MEISSNER, C. A., & BRIGHAM, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, & Law*, *7*, 3-35.
- MEISSNER, C. A., TREDoux, C. G., PARKER, J. F., & MACLIN, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. *Memory & Cognition*, *33*, 783-792.
- MELARA, R. D., DEWITT-RICKARDS, T. S., & O'BRIEN, T. P. (1989). Enhancing lineup identification accuracy: Two codes are better than one. *Journal of Applied Psychology*, *74*, 706-713.
- MEMON, A., BARTLETT, J., ROSE, R., & GRAY, C. (2003). The aging eyewitness: Effects of age on face, delay, and source-memory ability. *Journals of Gerontology*, *58B*, P338-P345.
- MEMON, A., & GABBERT, F. (2003). Unravelling the effects of sequential presentation in culprit-present lineups. *Applied Cognitive Psychology*, *17*, 703-714.
- MEMON, A., HOPE, L., & BULL, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology*, *94*, 339-354.
- MORGAN, C. A., HAZLETT, G., DORAN, A., GARRETT, S., HOYT, G., THOMAS, P., ET AL. (2004). Accuracy of eyewitness memory for persons encountered during exposure to highly intense stress. *International Journal of Law & Psychiatry*, *3*, 265-279.
- MUNSTERBERG, H. (1908). *On the witness stand*. New York: McClure.
- NATIONAL INSTITUTE OF JUSTICE (1999). *Eyewitness evidence: A guide for law enforcement*. Washington, DC: U.S. Department of Justice.

- NAVON, D. (1992). Selection of lineup foils by similarity to the suspect is likely to misfire. *Law & Human Behavior*, **16**, 575-593.
- NELSON, J. D. (2005). Finding useful questions: On Bayesian diagnosticity, probability, impact, and information gain. *Psychological Review*, **112**, 979-999.
- NETTLES, W., NETTLES, Z., & WELLS, G. L. (1996, November). "I noticed you paused on number three": Biased testing in eyewitness identification. *Champion*, pp. 10-12, 57-58.
- NICKERSON, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, **2**, 175-220.
- NOSOFSKY, R. M. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In F. G. Ashby (Ed.), *Multi-dimensional models of perception and cognition* (pp. 363-393). Hillsdale, NJ: Erlbaum.
- NOSWORTHY, G. J., & LINDSAY, R. C. L. (1990). Does nominal lineup size matter? *Journal of Applied Psychology*, **75**, 358-361.
- O'ROURKE, T. E., PENROD, S. D., CUTLER, B. L., & STUVE, T. E. (1989). The external validity of eyewitness identification research: Generalizing across subject populations. *Law & Human Behavior*, **13**, 385-395.
- PALEY, B., & GEISELMAN, R. E. (1989). The effects of alternative photo-spread instructions on suspect identification performance. *American Journal of Forensic Psychology*, **7**, 3-13.
- PARKER, J. F., & RYAN, V. (1993). An attempt to reduce guessing behavior in children's and adults' eyewitness identifications. *Law & Human Behavior*, **17**, 11-26.
- PAYNE, D. G., & ROEDIGER, H. L., III (1987). Hypermnnesia occurs in recall but not in recognition. *American Journal of Psychology*, **100**, 145-165.
- POZZULO, J. D., & LINDSAY, R. C. L. (1998). Identification accuracy of children versus adults: A meta-analysis. *Law & Human Behavior*, **22**, 549-570.
- POZZULO, J. D., & MARCINIAK, S. (2006). Comparing identification procedures when the perpetrator has changed appearance. *Psychology, Crime & Law*, **12**, 429-438.
- READ, J. D. (1995). The availability heuristic in person identification: The sometimes misleading consequences of enhanced contextual information. *Applied Cognitive Psychology*, **9**, 91-121.
- READ, J. D., HAMMERSLEY, R., CROSS-CALVERT, S., & MCFADZEN, E. (1989). Rehearsal of faces and details in action events. *Applied Cognitive Psychology*, **3**, 295-311.
- ROSE, R. A., BULL, R., VRIJ, A. (2005). Non-biased lineup instructions do matter: A problem for older witnesses. *Psychology, Crime & Law*, **11**, 147-159.
- ROSENTHAL, R. (2002). Covert communication in classrooms, clinics, courtrooms, and cubicles. *American Psychologist*, **57**, 839-849.
- RUMAIN, B., CONNELL, J., & BRAINE, M. D. S. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: If is not the biconditional. *Developmental Psychology*, **19**, 471-481.
- SCHACTER, D. L., DAWES, R., JACOBY, L. L., KAHNEMAN, D., LEMPERT, R., ROEDIGER, H. L., III, & ROSENTHAL, R. (2008). Policy forum: Studying eyewitness identifications in the field. *Law & Human Behavior*, **32**, 3-5.
- SHECK, B., NEUFELD, P., & DWYER, J. (2000). *Actual innocence: Five days to execution and other dispatches from the wrongly convicted*. New York: Doubleday.
- SCHMECHEL, R. S., O'TOOLE, T. P., EASTERLY, C., & LOFTUS, E. F. (2006). Beyond the ken? Testing jurors' understanding of eyewitness reliability evidence. *Jurimetrics*, **46**, 177-214.
- SCHUM, D. A., & MARTIN, A. W. (1982). Formal and empirical research on cascaded inference in jurisprudence. *Law & Society Review*, **17**, 105-152.
- SHEPHERD, J. W., ELLIS, H., & DAVIES, G. (1982). *Identification evidence: A psychological evaluation*. Aberdeen: Aberdeen University Press.
- SHIFFRIN, R. M., & STEYVERS, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, **4**, 145-166.
- SINGER, M., & WIXTED, J. T. (2006). Effect of delay on recognition decisions: Evidence for a criterion shift. *Memory & Cognition*, **34**, 125-137.
- SMITH, S. M., LINDSAY, R. C. L., PRYKE, S., & DYSART, J. E. (2001). Postdictors of eyewitness errors: Can false identifications be diagnosed in the cross-race situation? *Psychology, Public Policy, & Law*, **7**, 153-169.
- SPORER, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology*, **78**, 22-33.
- STEBLAY, N. M. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law & Human Behavior*, **21**, 283-297.
- STEBLAY, N. [M.], DYSART, J., FULERO, S., & LINDSAY, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law & Human Behavior*, **25**, 459-473.
- STEBLAY, N. [M.], DYSART, J., FULERO, S., & LINDSAY, R. C. L. (2003). Eyewitness accuracy rates in police showup and lineup presentations: A meta-analytic comparison. *Law & Human Behavior*, **27**, 523-540.
- STRETCH, V., & WIXTED, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1379-1396.
- TUNNICLIFF, J. L., & CLARK, S. E. (2000). Selecting foils for identification lineups: Matching suspects or descriptions? *Law & Human Behavior*, **24**, 231-258.
- TVERSKY, A. (1977). Features of similarity. *Psychological Review*, **84**, 327-352.
- WAGENAAR, W. A., & VEEFKIND, N. (1992). Comparison of one-person and many-person lineups: A warning against unsafe practices. In F. Losel, D. Bendre, & T. Bliesner (Eds.), *Psychology and law: International perspectives* (pp. 275-285). Berlin: De Gruyter.
- WASON, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology* (Vol. 1, pp. 135-151). Harmondsworth, U.K.: Penguin.
- WELLS, G. L. (1978). Applied eyewitness testimony research: System variables and estimator variables. *Journal of Personality & Social Psychology*, **36**, 1546-1557.
- WELLS, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology*, **14**, 89-103.
- WELLS, G. L. (1988). *Eyewitness identification: A system handbook*. Toronto: Carswell Legal Publications.
- WELLS, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, **48**, 553-571.
- WELLS, G. L., & LOFTUS, E. F. (2003). Eyewitness memory for people and events. In A. M. Goldstein (Ed.), *Handbook of Psychology: Forensic psychology* (pp. 149-160). New York: Wiley.
- WELLS, G. L., MALPASS, R. S., LINDSAY, R. C. L., FISHER, R. P., TURTLE, J. W., & FULERO, S. M. (2000). From the lab to the police station: A successful application of eyewitness research. *American Psychologist*, **55**, 581-958.
- WELLS, G. L., MEMON, A., & PENROD, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, **7**, 45-75.
- WELLS, G. L., & OLSON, E. A. (2002). Eyewitness identification: Information gain from incriminating and exonerating behaviors. *Journal of Experimental Psychology: Applied*, **3**, 155-167.
- WELLS, G. L., & OLSON, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, **54**, 277-295.
- WELLS, G. L., RYDELL, S. M., & SEELAU, E. P. (1993). The selection of distractors for eyewitness lineups. *Journal of Applied Psychology*, **78**, 835-844.
- WELLS, G. L., SMALL, M., PENROD, S., MALPASS, R. S., FULERO, S. M., & BRIMACOMBE, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law & Human Behavior*, **22**, 603-647.
- WILCOCK, R. A., BULL, R., & VRIJ, A. (2005). Aiding the performance of older eyewitnesses: Enhanced non-biased line-up instructions and line-up presentation. *Psychiatry, Psychology & Law*, **12**, 129-140.
- WOGALTER, M. S., MALPASS, R. S., & MCQUISTON, D. E. (2004). A national survey of US police on preparation and conduct of identification lineups. *Psychology, Crime & Law*, **10**, 69-82.
- YARMEY, A. D. (2004). Eyewitness recall and photo identification: A field experiment. *Psychology, Crime & Law*, **10**, 53-68.
- YARMEY, A. D., YARMEY, M. J., & YARMEY, A. L. (1996). Accuracy of eyewitness identification in showups and lineups. *Law & Human Behavior*, **20**, 459-477.

- YUILLE, J. C., & CUTSHALL, J. L. (1986). A case study of eyewitness memory to a crime. *Journal of Applied Psychology*, 71, 291-301.
- YUILLE, J. C., & TOLLESTRUP, P. A. (1992). A model of the diverse effects of emotion on eyewitness memory. In S. Christianson (Ed.), *The handbook of emotion and memory: Research and theory* (pp. 201-215). Hillsdale, NJ: Erlbaum.

NOTES

1. There are exceptions. Nicolas Mobley, a filler in a photographic lineup, was held for 29 days in a murder investigation by Chicago police after five witnesses identified him (Ferkenhoff, 2002). We thank Gary Wells for bringing this case to our attention.
2. Computational models for eyewitness identification have started to appear (Clark, 2003; Clare & Lewandowsky, 2004; Gronlund, 2005). However, for breadth, and in order to keep the focus on phenomena, our review will not go into the specific details of particular models.
3. Our assessment of Dekle's (1997) results differs from Dekle's, who concluded that accuracy for showups was maintained over the delay and that accuracy for showups was higher than that for lineups. In reaching this conclusion, however, Dekle excluded witnesses who gave *don't know* responses and then summed over correct identifications and correct rejections, a calculation that will tend to favor showups but has no bearing on the probative value or risk due to a suspect identification.
4. Our analysis excluded four well-known studies that manipulated delay, because they reported results collapsed over target-present and target-absent lineups, making it impossible to evaluate correct and false identification effects separately or to calculate innocence risk. Of these, two reported decreases in overall accuracy (Cutler, Penrod, O'Rourke, & Martens, 1986; Read, Hammersley, Cross-Calvert, & McFadzen, 1989), and one showed an increase in overall accuracy (Cutler, Penrod, & Martens, 1987b). Two other studies (Cutler, Penrod, & Martens, 1987a; Yarmey, 2004) did not report the relevant data.
5. For example, the lineups were not purely target present or target absent and, in some cases, contained multiple targets and multiple innocent suspects. Also, all the witnesses were presented with both correct (guilty suspects) and incorrect (innocent suspects) information during the postevent phase.
6. We should note, however, that there is some new evidence that repeated recognition testing can produce facilitation if the testing procedures result in the recollection of information from memory (Chan & McDermott, 2007).
7. Calculation of innocence risk requires some assumptions about misidentifications in the target-absent lineup condition. At issue is whether the false identification rate should assume that the appearance of the innocent suspect is linked to that of the target. In other words, if the perpetrator shaves his moustache, should the target-absent lineup also consist of people without moustaches? Adherence to a factorial design would seem to dictate that if the target-present lineup has both changed and unchanged appearance conditions, then so too should the target-absent lineup condition. This design simulates suspect change, rather than perpetrator change. Read's (1995) results showed an increase in innocence risk for both longer and shorter exposure durations. Innocence risk increased from .05 to .32 in the longer exposure condition and from .13 to .22 in the shorter exposure condition. Pozzulo and Marciniak (2006) showed that innocence risk increases in both simultaneous (.11 to .24) and sequential (.12 to .19) lineups. If innocence risk calculations are made under the assumption of an unchanged innocent suspect, the numbers change (somewhat inconsistently) but show the same general pattern.
8. What if some of the foils are "good," but a few easily rejected? Can the lineup be robust to a few poorly chosen foils? The results of a handful of experiments have been mixed. Brewer and Wells (2006) used eight-person lineups consisting entirely of higher similarity foils or consisting of a mix of higher and lower similarity foils, and Nosworthy and Lindsay (1990) added four or seven low-similarity foils to a base lineup with three higher similarity foils (so that the lineup sizes were 4, 7, and 10). These mixed-similarity lineups produced small and inconsistent results. On the basis of their results, Nosworthy and Lindsay suggested that the identification task may be robust to some limited compositional bias, provided that at least some of the foils are plausible choice alternatives. However, a recent study by Carlson et al. (2008, Experiment 2) showed substantial increases in both correct (from .314 to .426) and false (from .163 to .303) identification rates comparing lineups with all "good" foils to lineups with a mix of three good and two easy-to-reject foils.
9. The problem is that the additional identifications are directed to lineup members other than the perpetrator. One way to account for such results is to assume that biased instructions lead witnesses to ignore or give less weight to mismatching features and more weight to matching features. This model, which is similar to Tversky's (1977) feature contrast theory, should (in principle) allow the increase in the identification rate due to biased instructions to be directed toward foils, rather than toward the perpetrator.
10. One may argue, however, that a *live* showup is still better than a *photo* lineup, when conducted right away. At this point, there are no data bearing on this comparison.

(Manuscript received September 17, 2006;
revision accepted for publication July 16, 2008.)