Law and Human Behavior, Vol. 21, No. 3, 1997

Social Influence in Eyewitness Recall: A Meta-Analytic Review of Lineup Instruction Effects

Nancy Mehrkens Steblay[1]

Past research has considered the impact of biased police lineup instructions upon eyewitness lineup performance. Biased instructions either suggest to the eyewitness that the perpetrator is in the lineup or otherwise discourage a "no choice" response. A meta-analysis of 18 studies was employed to review the hypothesis that biased instructions lead to greater willingness to choose and less accurate lineup identifications than do unbiased instructions. The role of moderating variables in the instruction procedure was also considered. In support of the hypothesis, a significantly higher level of choosing followed biased instructions. Lineup type moderated performance accuracy, however. For target-absent lineups the increased level of choosing following biased instructions resulted in reduced identification accuracy. Biased instructions within a target-present lineup generated a higher level of confidence, but had minimal impact on accuracy. Implications for police practice are discussed.

The susceptibility of human perceptual processes to outside intrusions and the impact of external cues on reported memory have been investigated and confirmed within many realms of human behavior and decision making. From the early conformity studies of Asch (1951) to current investigations of repressed memory (Loftus, 1993), psychologists have recognized the nontrivial effects of external cues on individuals' behaviors, reported knowledge, and memory itself. Mechanisms of this influence process may differ from one scenario to another, or from one individual to another, but are often founded on an individual's reliance on others to help define an external social situation or to clarify his or her own emotions and beliefs. For example, a *normative* social influence process will render its impact on an individual who is searching for a response that provides a correct, modal, or positive self-presentation. *Informational* social influence will produce its effect through convincing information offered to the uncertain target (Deutsch & Gerard, 1955). In either case, the result is a change in behavior or belief in response to the external agent.

[1]Department of Psychology, Augsburg College, 2211 Riverside Avenue, Minneapolis, MN 55454. e-mail: steblay@augsburg.edu.

The relevance of these most basic social psychological principles can be keenly recognized in the domain of forensic eyewitness reporting. A case in point is a police lineup procedure, in which a witness is asked to identify a suspect. The witness is provided with a scenario in which the task is clearly defined (choose the perpetrator from the lineup), and in which the familiar script is augmented by instructions from an authority figure which may provide informational social influence (the perpetrator is in the lineup) as well as normative social influence (the correct response is to make a choice). Researchers suspect that the effect of such instructions is increased encouragement, and perhaps pressure, for the witness to choose from the lineup. It is also likely that the witness will be swayed to rely on a relative judgment heuristic (selection of the lineup member who most resembles the person previously seen), rather than an absolute judgment (a confirmed match between a lineup member and memory) (Wells, 1984). As suggested by Sporer, Penrod, Read, and Cutler (1995), it is possible that an eyewitness who chooses from a lineup has met two cognitive criteria: A face in the lineup is perceived as familiar and that face is judged as having appeared in the context of the crime. Normative pressure may serve to ease acceptance of a face with familiar features; informational social influence provides the contextual criterion (through a claim that the crime was in fact committed by one of the lineup members). This relaxation of decision criteria should produce more identification errors.

Wells, Seelau, Rydell, and Luus (1994) have confirmed the presence of eyewitness error, citing evidence from three realms. First, analysis of over a thousand actual cases of wrongful convictions has determined that eyewitness recall error is the primary contributor to miscarriages of justice. More precisely, Rattner (1988) reports that 52% of proven cases of wrongful conviction are attributable to false identification. Wells et al. additionally note the large number of research experiments that report a high frequency of false identification by subject eyewitnesses. Finally, Wells et al. refer to the "considerable research evidence indicating that there is a 'sincerity' to most of the false identifications observed in these experiments; the eyewitnesses often seem actually to believe that their false identifications are in fact accurate identifications" (p. 224).

Identification errors may spring from many sources, a number of which have been addressed in research over the past two decades. Findings that specifically link lineup instruction and identification error have led researchers to a tentative conclusion regarding the status of the research literature. Kassin, Ellsworth, and Smith (1989) documented researchers' belief in the power of lineup instructions to alter witness behavior. They surveyed experts in the area of eyewitness accuracy to record their judgments of the status of research literature on each of 21 eyewitness topics. The following research finding was judged reliable by 95% of the 63 respondents: "Police instructions can affect an eyewitness's willingness to make an identification and/or the likelihood that he or she will identify a particular person" (p. 1091).

The apparent consensus among experts is perhaps not surprising, particularly if one looks at the backdrop of research that demonstrates the impact of question wording on response in settings other than lineups (e.g., Loftus, 1979). Research more specific to the eyewitness experience is that which has tested the effect of suggestive lineup instruction for research subjects who have witnessed an experimenter-controlled prior event. Most common is a hypothesis that predicts lower lineup identification accuracy following "biased" instruction (which exerts pressure on the witness to make an identification) than following a more neutral "unbiased" instruction. A quick tally of research findings that have tested this hypothesis indicates 11 tests with results supportive of the hypothesis, 10 tests that report no significant difference between conditions (at the traditional .05 level), and no studies that report results in the opposite direction. Thus the research shows some support for the hypothesis that biased instructions hinder lineup accuracy, but there exist nonsupportive findings as well. Disagreements also exist regarding the interpretation of results, arguments based on statistical and methodological grounds. For example, instances of nonsignificant findings may be interpreted either as evidence counter to the hypothesis, or rather as a reflection of lack of statistical power due to small sample size (Cutler & Penrod, 1995). It has also been argued that simulation studies, in which subjects recognize their decisions to be relatively inconsequential, may pose external validity problems that eclipse the laboratory effects achieved (Kohnken & Maass, 1988).

Is the confidence of the Kassin et al. experts well placed, i.e., are the experts correct in their assessment of the lineup instruction literature? The current project is an attempt to review the research literature, both to check the accuracy of expert opinion and to provide a statistical summary of this body of research. The usefulness of meta-analysis is throughly reviewed elsewhere (Rosenthal, 1991); however, in the case of eyewitness factors, it should be noted that a summary analysis is essential to application of research findings. Researchers who provide information to the criminal justice system need more than familiarity with isolated studies or a simple tally of results for and against a hypothesis. Rather, the expert must provide a clear and precise synthesis of available evidence, as well as a recognition of the strengths and weaknesses of the evidence in its support of a principle. The probability and effect size indicators generated through meta-analysis provide a substantial step toward a thorough assessment of the literature, particularly allowing for recognition of a pattern of effect across studies, thus helping to overcome the limitations of small sample size.

In addition, the moderator variables examined in the meta-analytic review may produce a more thorough and precise theoretical understanding of the phenomenon under study by examining effect size in relation to method. For example, it is readily apparent when examining lineup literature that both "biased" and "unbiased" instructions are operationally defined in a number of ways by various researchers. In general, biased instructions are expected to suggest to the witness that a lineup choice is the appropriate response, thereby increasing both the likelihood of a choice and the potential for false identification. However, the research indicates three primary means through which researchers have implemented the pressure to choose through the content of instructions. For the purposes of this analysis, instruction content has been grouped and labeled in the following manner: (1) A *Leading Instruction* states or strongly implies that the suspect is in the lineup. (Unbiased instruction in this case states that the suspect "may or may not be present.") (2) A *Pressure Instruction* fails to offer the option of declining to choose, i.e., a "not present" or "no choice" option is missing. (The parallel unbiased instruction provides this option.) (3) A *Task-oriented* instruction stresses the importance of succeeding at the identification task and/or the serious consequences of a failure to identify the perpetrator. (Unbiased instruction cautions the eyewitness to be careful and stresses the importance of not guessing.) This lack of a unitary operational definition has allowed researchers to test the parameters of the instruction effect. As the literature grows, however, it becomes more difficult for the reader to ascertain the effects of subtle changes in instructions from one study to another. This meta-analysis will examine the pattern of effects achieved by such moderator variables and thus will help to define, from a much broader data base than the individual study, the parameters of the instruction effect.

Optimistically, we may hope that increased knowledge of this phenomenon will not simply allow us to reflect with greater understanding upon events that have already occurred or to estimate their likely impact upon the eyewitnesses who struggle to remember events in the courtroom. Rather, this information may be parlayed into a procedural change in the criminal justice system, which might in fact increase the accuracy of eyewitness recall. Numerous factors affect memory for a witnessed event, and a useful distinction has been made between *estimator* and *system* variables (Wells, 1978). Estimator variables, such as stress level of a witness, duration of an event, race of the perpetrator, and presence of a weapon, are integral parts of the event. Thus, their impact on eyewitness recall can only be estimated post hoc. In an attempt to maximize the utility of eyewitness research, some investigators (e.g., Seelau & Wells, 1995) have urged a refocus of attention on system variables. System variables, such as interrogation methods and lineup structure, are potentially under the control of the criminal justice system, thus the research offers direction for changes in police procedures that might enhance the quality of evidence provided by eyewitness recall. One goal of the system variable approach (Seelau & Wells, 1995) is to provide preventative measures to inhibit potential eyewitness errors at a point early in the timeline of criminal justice proceedings. Specific to lineup procedures, instructions that promote accurate decision-making (minimizing the potential for false identification while allowing accurate identification) may be implemented into the police procedures. The task for researchers is to present a clear picture to law enforcement agencies as to the need for procedural changes and the form that such changes should take. A meta-analytic review is a part of the process of developing a summary of data to guide policy changes.

Goals of the current project include: (1) A summary statement of the effects of biased lineup instruction on lineup identification performance. Multiple dependent variables are reported in the research, thus this review will examine the effect of biased lineup instructions on three measures: choosing, accuracy of identification, and confidence in decision; (2) investigation of the role of moderating variables in the instruction process, with emphasis on the conditions under which eyewitness memory is less likely to be accurate; and (3) identification of shortcomings in the body of literature and areas for future investigation.

METHOD

Sample

A computer search of the CD-Rom database PsycLIT provided an initial sample of studies relevant to the hypothesis. E-mail and an electronic listserve (PSY-

LAW-L) were used to contact researchers in an attempt to provide leads to additional works, published or unpublished. In order to be included in the sample, the study must have provided a statistical test of the relationship between biased (vs. unbiased) instructions and lineup identification performance (accuracy or willingness to choose.) Only studies with adult subjects were considered. Eighteen articles (15 published papers, 1 convention paper, and 2 unpublished works) were located, providing 22 tests of the hypothesis. The sample included work published between 1975 and 1996, representing a total of 2588 subjects. Sample sizes within studies ranged from 36 to 354 ($M = 129.4$).

Study Characteristics

Both methodological and theoretical variables were coded as part of the data set. These included: researcher, year of publication, source (published, convention presentation, unpublished), number of hypothesis tests per study, sample size, operational definitions of "biased instruction," and "unbiased instruction," lineup type (target present or absent), lineup display mode (live, video, slide), subjects (college student, nonstudent), intensity of the experimental scenario (high—a staged incident; moderate—a video or slide portrayal of a crime; low—incidental event), number of days delay between event and lineup, time of debriefing (before or after the lineup task), and lineup presentation (sequential, simultaneous). Sex of lineup target was found not to vary; all lineups included male targets. Recorded dependent variables were choosing, accuracy (proportion of correct decisions), and confidence in decision.

Statistics

Following the work of Rosenthal (1991), a meta-analytic Z (Z_{ma}) was calculated by combining Z-scores of individual tests of the hypothesis (Stouffer method). This produces an overall probability level associated with the observed pattern of results. A fail-safe N (N_{fs}) was calculated to estimate the number of additional tests averaging null results that would be needed in order to bring the significance level attained through the meta-analysis to a value larger than .05. The effect size h is the coefficient recommended by Cohen (1988) for use when testing differences between proportions. For effect sizes calculated between mean scores, the statistic d was employed. The mean effect size for a group of studies is referred to in subsequent discussion as h (or d).

RESULTS

Dependent Measure: Accuracy

All Tests

Twenty-one tests of the hypothesis were first examined in an attempt to determine the overall status of the hypothesis. Identification accuracy was compared

Table I. Accuracy Effect Sizes (Studies Ordered by Date)

Author	Date	N	Overall (h)	Target-present (h)	Target-absent (h)
Hall & Ostrom	1975	41	.88	1.37	.71
Hilgendorf & Irving[a]	1978	36	.00		.00
Malpass, Devine & Bergen	1980	42	.00	−.38	.40
Warnick & Sanders[b]	1980	115	.37	.37	
Malpass & Devine	1981	100	.46	.20	.92
Hosch et al.	1984	80	.02	.02	
Cutler, Penrod et al.	1986	320	.27	−.42	1.05
Cutler, Penrod, & Martens	1987	165	.50	.00	1.07
Cutler, Penrod, & Martens	1987	290	.16	−.17	.62
Fleet, Brigham, & Bothwell	1987	96	.00	.04	−.04
Cutler & Penrod	1988	175	.05	.00	.13
Cutler & Penrod	1988	94	.34	.14	.58
Kohnken & Maass[c]	1988	52	−.12		−.12
Kohnken & Maass	1988	63	.75		.75
O'Rourke et al.	1989	121	.30	.00	.79
Paley & Geiselman[a]	1989	60	.35	−.26	1.03
Lindsay et al.	1991	120	.38		.38
Foster et al.	1994	354	.02	−.34	.46
Devenport & Fisher	1995	138	.08	−.36	.58

[a]Estimate of effect size; complete data not available.
[b]Calculation includes "Don't Know" (DK) as incorrect responses in target-present lineup; If DKs are excluded, $h = .71$.
[c]Calculation includes DKs as correct responses in target-absent lineup; if DK are excluded, $h = .04$.

for unbiased versus biased instruction conditions (U-B effect). Overall proportion of correct identifications for the unbiased condition was 56%; for the biased condition, 44%. The mean effect size, h, for this set of tests is .26, and the meta-analytic Z (Z_{ma}) is 5.71, $p < .0001$, $N_{fs} = 232$, providing support for the hypothesis. (Positive h and Z values indicate greater accuracy in the unbiased condition.) Only one of the studies (Warnick & Sanders, 1980) provided nonindependent tests of the hypothesis. In subsequent analyses, the data from the three experimental conditions within that study were combined and compared against the single control group to provide one test of the hypothesis.[2] Subsequent analyses attempted to explore moderator variables in the remaining 19 tests, which might account for the variability in effect sizes of individual studies (from −.12 to .88; $s = .27$; see Table I).

Lineup Type

The most dramatic result in the analysis of moderator variables was obtained when tests were separated by lineup type. (See Table II. Actual effect sizes are maintained on this graph; stems represent effect size in increments of .1, leaves relate hundredths digits.) Target-absent lineups ($N = 16$) generated an average effect size $h = .58$ (95% CI: .39 to .78), $Z_{ma} = 9.76, p < .0001, N_{fs} = 547$. Target-present lineups ($N = 15$) produced $h = .02$ (95% CI: −.23 to .26), $Z_{ma} = -1.02$. Thus biased instruc-

[2]Warnick and Sanders report no significant difference between the three experimental conditions. The effect of combining the groups on the preceding analysis, now with 19 studies, was as follows: $Z_{ma} = 5.42$, $h = .25$, $N_{fs} = 188$.

Table II. Stem and Leaf Display of Target-Absent and Target-Present Lineup Effect Sizes

Target absent		Target present	
Stem	Leaf	Stem	Leaf
1.3		1.3	7
1.2		1.2	
1.1		1.1	
1.0	3,5,7	1.0	
.9	2	.9	
.8		.8	
.7	1,5,9	.7	
.6	2	.6	
.5	8,8	.5	
.4	0,6	.4	
.3	8	.3	7
.2		.2	0
.1	3	.1	4
.0	0	.0	0,0,0,2,4
−.0	4	−.0	
−.1	2	−.1	7
−.2		−.2	6
−.3		−.3	4,6,8
−.4		−.4	2

tions produced a moderate effect on accuracy in target-absent lineups (biased instructions producing significantly less accurate identifications), but minimal effect in target-present lineups. The mean accuracy rate for unbiased versus biased lineups in the target-absent lineup condition was 60% versus 35%; In the target-present lineups, the identification rates were 54% (unbiased) and 53% (biased).[3]

Instruction Content

The operational definition of "biased" was used to partition the tests into subsets (see Table III). Eleven tests (52%) included the combination of leading and pressure instructions, producing $h = .27$ (95% CI: .10 to .43), $Z_{ma} = 4.93$, $p < .0001$, $N_{fs} = 88$. Four tests used leading instruction alone (with a *Not present* or *Don't know* option available), $h = .28$ (95% CI: −.43 to .99), $Z_{ma} = 1.81$, $p = .035$, $N_{fs} < 1$. One test used pressure instruction alone, $h = .34$, $Z = 1.70$. Two studies defined bias by failing to offer a "don't know" option yet with a "not present" option included, $h = .21$, $Z_{ma} = 1.40$. Only one study (Hilgendorf & Irving, 1978) used a task-oriented instruction alone, reporting no effect.[4] Thus a signifi-

[3]Cohen (1988) proposes a convention for interpretation of effect size h: .20 is considered a small effect; .50, a medium effect; .80, a large effect. However, as noted by G. L. Wells (personal communication, May 22, 1996), the natural metric of percent accuracy offers a more meaningful indicator of effect size in this case than does Cohen's generic interpretation. Consider, for example, 1000 target-absent lineups conducted over some period of time. Given the accuracy data in this study, we would expect unbiased instructions to result in 250 fewer identification errors. Alternatively, the chances of error in a target-absent lineup increase by 63% (35% error to 60% error) when biased instruction is used.

[4]This study (Hilgendorf & Irving, 1978) does not report the data necessary for precise calculations. It is also the only study that defines bias using only the task-oriented instruction. For these reasons, the study will not be included in subsequent analyses.

Table III. Accuracy—Unbiased vs. Biased Instructions

Moderator variable	h	Z_{ma}	N
Lineup type (target)			
Present	.02	-1.02	15
Absent	.58	9.76	16
Instruction content			
Leading + pressure	.27	4.93	11
Leading alone	.28	1.81	4
Pressure alone	.34	1.70	1
Task-oriented	.00	0.00	1
Other	.21	1.40	2
Lineup display			
Photo	.18	2.95	8
Live	.39	3.20	5
Video	.33	3.92	4
Lineup presentation			
Sequential	.36	4.17	3
Simultaneous	.25	3.59	12
Delay			
No delay	.15	2.57	8
2-day	.46	3.79	5
3-day	.46	2.28	1
7-day	.23	2.46	3
14-day	.16	1.36	1
Intensity			
High	.20	2.56	8
Moderate	.26	4.62	9
Low	.88	2.54	1
Time of debriefing			
Before	.50	3.67	4
After	.04	.35	6
Subject population			
Student	.26	5.00	16
Nonstudent	.34	2.59	2
Published/not published			
Published	.26	5.33	15
Convention paper	.88	2.54	1
Unpublished	.04	.35	2

cant outcome occurred for the combination of leading plus pressure instructions, but also for each of these options used alone. Task-oriented instructions appeared to have a minimal effect.

Additional Moderating Variables

As indicated in Table III, the remaining moderator variables produced a relatively consistent range of effect sizes across levels within each variable. All effects were positive, and most effects were small in size (.16 to .39). An exception to this pattern of small effects was the effect size for six tests with a 2- to 3-day delay between event and lineup, h = .46. One convention paper also rated as a low-intensity scenario (incidental event) produced h =.88. There were very small effect sizes for the two unpublished reports (.00 and .08).

Analysis of time of debriefing was limited to eight studies in which subjects were exposed to what they believed were actual events rather than experimental tasks (and in which data were available). The analysis produced intriguing, although tentative, results. In four tests, subjects were debriefed (told that the event was not real) before the lineup task, $h = .50$. In six tests, subjects were debriefed after the lineup task, $h = .04$.

Instruction Bias Effects Within Lineup Type

The previous analyses attempted to uncover moderator variables to account for the variability in U-B accuracy effects across studies. The most notable was the comparison of lineup type, with target-absent lineups generating a moderate (.55) U-B effect and target-present lineups yielding minimal (.06) effect. It appears that the analyses above (which like many of the individual studies combined data of target-present and target-absent lineups) masked this important moderator, the combined data averaging to produce significant but small effects. To better partition the variability within U-B effects, the following analyses examine the U-B effect for each potential moderator variable, within each of the two lineup types. The results (see Table IV) indicate that target-absent lineups present a robust pattern of moderate to large U-B effect sizes regardless of the additional variables considered. Target-present lineups generate a pattern of small and often negative effect sizes.

In target-absent lineups, instructions that include leading plus pressure biases ($N = 10$) produced a moderate effect size, $h = .65$, (95% CI: .41 to .90), and the highest Z value, $Z_{ma} = 9.39$, $N_{fs} = 316$. Similarly, effects are strong for both sequential ($h = .97$) and simultaneous ($h = .53$) lineups, and for all levels of time delay (h's from .43 to .92). Moderate to large effect sizes were generated across all methods of lineup display and in all levels of scenario intensity. Effects were smaller, and/or Z_{ma}'s not significant, for the few tests in which instructional bias involved a lack of *Don't Know* response option, for low-intensity scenarios, and in the four tests in which debriefing was after the lineup task.[5]

Effect sizes above .80 are noted for video lineup display and for sequential lineup presentation. The studies represented in the analysis for video lineup displays ($N = 3$) and sequential presentations ($N = 3$) are small in number and overlapping, all coming from the Cutler et al. laboratory. An examination of combinations of biased and unbiased instructions also shows that these four studies use a unique pairing of instructions: The biased instructions did not directly state that the perpetrator was in the lineup, but rather requested that subjects "Write the number of the suspect." Unbiased instructions included an option to indicate in writing that the robber is not present (with no allowance for a "Don't Know"

[5]An alternative means of looking at instructions is to compare the content between biased and unbiased instructions within each study. It is assumed that the U-B effect achieved in a study should reflect the difference between the content of instructions in the two conditions. Six tests which clearly compare leading plus pressure biased instructions to an unbiased instruction in which this bias is remedied produced $h = .50$, $Z_{ma} = 4.50$. Pressure instructions alone ($N = 5$) produced $h = .82$; leading instructions alone generated $h = .35$ ($N = 3$). A study in which the only change in instructions for the unbiased condition is that a "don't know" is added produced $h = .13$.

Table IV. Accuracy for Unbiased vs. Biased Instructions Target Present/Target Absent Lineups

	Target present			Target absent		
Moderator variable	h	Z_{ma}	N	h	Z_{ma}	N
Instruction content						
Leading + pressure	−.12	−1.71	9	.65	9.39	10
Leading alone	.26	−.55	3	.52	3.37	4
Pressure alone	.14	.47	1	.58	1.99	1
Other	.19	.99	2	.13	.63	1
Lineup display						
Photo	−.17	−2.42	7	.58	6.79	7
Live	.40	1.03	3	.53	3.54	5
Video	.05	.35	4	.82	7.21	3
Lineup presentation						
Sequential	−.14	−1.53	3	.97	8.68	3
Simultaneous	.05	−.56	10	.53	6.14	10
Delay						
No delay	−.10	−1.70	7	.43	5.07	6
2-day	.22	.43	4	.70	4.50	5
3-day	.20	.69	1	.92	3.07	1
7-day	.00	.00	2	.58	4.79	3
14-day	−.17	−.98	1	.62	3.64	1
Intensity						
High	−.10	−.65	5	.41	3.95	7
Moderate	−.08	−1.48	9	.72	9.52	8
Low	1.37	2.0	1	.71	1.62	1
Debriefing						
Before	.20	.69	1	.61	4.06	4
After	−.17	−1.09	4	.21	1.47	5

response). These studies were designed to test for the effects of subtle changes in instructions. The four produce consistently strong U-B effects for the target-absent lineups, $h = .88$, $Z_{ma} = 9.33$, and no or reversed effects in the target-present lineups, $h = -.15$, $Z_{ma} = -1.81$. Thus, it is difficult to pinpoint one specific correlate of strong effect size in this group of studies.

Within target-present tests, the overall U-B effect is limited ($h = .06$), although as indicated on Table IV, the pattern of results is not consistent across the variables considered. Nonsignificant Z_{ma}'s and/or very small effect sizes are common in the target-present tests. However, a significant $Z_{ma} = -1.71$, $d = -.12$ (95% CI: −.29 to .05) was found for the 9 tests, which included leading plus pressure instructions. The negative effect suggests that *biased* instructions may produce higher accuracy when instructions include leading plus pressure content; however, N_{fs} is < 1. Photo lineup displays in seven tests generated $h = -.17$ and a significant $Z_{ma} = -2.42$; again, however, N_{fs} is small, 8. The variability in effect sizes within target-present lineups (from −.17 to 1.37) suggests noise in the data, which is not yet explained.

The intriguing pattern of results discussed earlier regarding time of debriefing (which showed a significant outcome when debriefing was before but not after the lineup task) is also affected by lineup type. The "before task" debriefing in four tests generates a significant outcome and moderate effect size in the target-absent condition, $h = .61$, $Z_{ma} = 4.06$ ($N_{fs} = 20$) and a small effect in one target-present condition, $h = .20$, $Z = .69$. When debriefing is after the lineup task, the effect size is small in the target-absent condition ($N = 5$) $h = .21$ and $Z_{ma} = 1.47$ and negative in the target-present condition ($N = 4$) $h = -.17$, $Z_{ma} = -1.09$. Again, the earlier combined data masked the impact of lineup type.

Dependent Measure: Choosing

Thirteen studies reported the numbers of subjects who made a positive identification (whether correct or incorrect) and those who did not make a lineup choice (rejected the lineup). The prediction for this dependent variable is that biased instructions will produce higher levels of choosing than will unbiased instructions. These 13 studies produced $h = -.48$ (95% CI: –.69 to –.27) and $Z_{ma} = -9.69$, $p < .0001$, $N_{fs} = 440$, indicating the predicted higher levels of choosing for subjects in the biased instruction condition. (Negative h and Z values indicate higher levels of choosing in the biased condition.)

Given the small number of studies in this set, few moderator variables could be effectively analyzed. However, three variables that allowed analysis—lineup type, scenario intensity and delay—produced consistently strong results across levels of the variables. Separated by lineup type, the data produced $h = -.46$ (95% CI: –.80 to –.12), $Z_{ma} = -5.06$, $p < .0001$, for eight studies with target-present lineups. A similar outcome occurred for the seven studies with target-absent lineup data, $h = -.48$ (CI: –.79 to –.18), $Z_{ma} = -5.37$. Both moderate- and high-intensity scenarios generated significant Z_{ma}'s, – 7.99 and –6.04, respectively, with a greater effect size for the moderate-intensity scenario, $h = -.60$ ($N = 5$) versus $h = -.40$ ($N = 8$). Delay produced varying effect sizes: Zero delay, $h = -.32$, $Z_{ma} = -4.87$, $N = 5$; 2-day delay, $h = -.72$, $Z_{ma} = -5.70$, $N = 4$; 3-day delay, $h = -.77$, $Z_{ma} = -7.89$, $N = 2$; 7-day delay, $h = -.12$, $Z = -.55$, $N = 1$.

Within target-present and target-absent lineups, the U-B effect on choosing was also examined as a function of debriefing (within the 8 high-realism studies examined earlier). Biased instructions produced higher levels of choosing for subjects debriefed both prior to and after a lineup task for both target-present and absent lineups. However, the effect sizes were larger for those subjects debriefed prior to the task. Within the target-present lineups, the one study that debriefed prior to the lineup task produced $h = -.85$, $Z = -2.4$; the four studies that debriefed after the lineup generated $h = -.29$, $Z_{ma} = -2.26$. In the target-absent lineups, a similar pattern emerged: Prior debriefing ($N = 4$) produced $h = -.62$, $Z_{ma} = -4.13$; later debriefing ($N = 5$), $h = -.21$, $Z_{ma} = -1.50$.

Dependent Measure: Confidence

Ten studies included a postlineup measure of confidence. Four reported no significant effect of lineup instructions on subjects' confidence ratings; these data are entered into the analysis as conservative estimates, $z = 0$, $d = 0$. The effect size based on nine tests with available data, d, is –.08. Z_{ma} for the entire group of 10 tests is –1.96, $p = .03$, $N_{fs} = 4$, with biased instructions producing more confidence than unbiased instructions. The effect size is small in target-present lineups

($N = 7$), $d = -.16$, $Z_{ma} = -2.07$, $p = .02$, $N_{fs} = 4$, and minimal in target-absent lineups ($N = 7$), $d = .02$, $Z_{ma} = -.40$.

DISCUSSION

The available data support the hypothesis that biased instructions significantly affect eyewitness lineup identification performance. The most reliable effect is achieved through a combination of instruction content that suggests to the witness that the perpetrator is in the lineup and which provides no explicit option to reject the lineup. Eyewitness vulnerability to normative and informational influences is apparent in the significantly higher level of choosing following biased instruction. It may be assumed that any errors due to inherent limitations in human memory will be compounded by increased levels of choosing. In fact, accuracy is moderated by lineup type. Instructions that included the combination of leading and pressure components produced significantly more correct identifications in a target-present lineup than did the unbiased instruction. In these controlled studies, the target-present lineup always in fact includes the perpetrator, thus increased choosing and use of a relative judgment strategy will likely lead to higher levels of accuracy. Conclusions about the effects of biased instruction on target-present lineups must be tempered, however, by the presence of high variability in effect sizes, along with the fact that either leading or pressure instruction alone did not produce a positive outcome in target-present conditions. There remains much noise in the data yet to be accounted for. Another consideration is that in real lineups the risk of error is higher, since the target may be an innocent suspect who resembles the true culprit.

Results more clearly demonstrate the significant negative impact of biased instructions on identification accuracy in target-absent lineups. It is here that social influence produces a lineup choice and, of course, an error. In the target-absent lineup, accuracy is significantly reduced by a combination of leading plus pressure instructional content, or either of these two components alone. Variations in display (photo, live, video), time delay, or lineup presentation (simultaneous vs. sequential) do not appear to counteract the negative impact of the biased instruction. Recent evidence supports the use of sequential lineups as a means to reduce false identifications through reduced reliance on a relative judgment strategy (Lindsay, Lea, & Fulford, 1991; Lindsay et al., 1991; Lindsay & Wells, 1985; Sporer, 1992). Future investigation of the interaction between lineup presentation and instruction would be beneficial. Additionally, future research could address the impact of task-oriented instruction. The current meta-analysis, due to limitations of available data, could not assess this form of instruction.

Also affected by lineup type is confidence of the eyewitness in his or her choice. Confidence is significantly higher following biased instruction, but only for the target-present lineup. Target-absent lineups produce as much choosing as target-present, but not with the certainty that may exist for target-present lineup choices. Perhaps social influence has relaxed the decision criteria to a point where eyewitnesses are not entirely comfortable with their choice, but choose nevertheless. This interpretation makes intuitive sense, given that the target-present lineup at least presents a familiar target; the target-absent lineup less so. This result must be viewed as tentative, however, given the low fail-safe N connected with the Z-score.

An important consideration for eyewitness research is the external validity of the research. The instruction effect exists—but what are the implications for the real world? May we expect the U-B effect to exist at the same level of magnitude across situations? One aspect of this validity question relates to mundane realism: Does the experimental scenario provide a realistic stimulus event for the subjects, the results of which can then be generalized to real-life lineup episodes? The lineup scenarios of these studies present the subjects with a moderately difficult task (as judged by the control accuracy rate of 56%). Eleven of the 20 studies involved realistic staged events, ranging from incidental interactions with confederates in a classroom to higher-intensity theft or vandalism scenes (typically with uniformed officers presenting the lineup). While effect sizes are larger for studies with less realism, the high-realism paradigms also generate a significant U-B effect. The simulation studies thus seem to generate the experimental realism necessary to assume reasonable external validity. Additionally, studies from nonstudent populations also produce the U-B effect, providing evidence of generalizability across population.

Of more concern, perhaps, are the results from a smaller set of high-realism studies, which demonstrated that debriefing of the subjects after the lineup task produced a smaller effect than that generated by subjects debriefed before the lineup. This narrowing of effect size in the "debriefing-after" condition suggests that the greater realism of the scenario for these subjects made them more cautious in their behavior, attenuating the U-B effect for both choosing and accuracy. Even so, these data show a pattern of effects consistent with that of the other moderator variables (larger effects in the target-absent condition) although reduced in magnitude. A difficulty in reaching a firm conclusion about this external validity issue stems from the small number of studies in this subset (eight total; fewer when separated by target-present vs. target-absent lineups) and the noise in the data (effect sizes ranging from –.38 to .04 in target-present lineups; –.19 to –.58 in the target-absent data). A task for future research is to continue to expand the parameters of the experimental scenario, particularly to address how lineup decisions made under perceived real-world conditions (and consequentiality) compare to simulation studies. In other words, will social influence be resisted when the stakes are higher? Past research in social psychology would not support this prediction (e.g., Milgram, 1965). Interestingly, one study, which tested directly for consequentiality in lineup decisions (Malpass & Devine, 1981), found higher levels of choosing when the consequences for the perpetrator were severe (pressing of charges, felony arrest, jail time) than when consequences were trivial ("catching hell"). Additionally, in real crimes, risks to the witness may be more salient—a court appearance, prospect for retaliation, pangs of conscience, the burden of society's welfare should the perpetrator continue to offend. The interactive effect of these personal consequences on choosing behavior remains untested. However, this need for realistic research must be balanced with the ethical and logistical issues inherent to such an involved paradigm. The available data may be the best we can have; the data do suggest that effect size may be reduced in size for simulation studies, but the U-B effect is still an empirically and theoretically sound principle.

A final issue is the overall relevance of the instruction procedures tested to those that actually occur in police lineups. It is assumed that police lineup procedures vary widely, most likely encompassing the wide range of instructions used in these studies. At the least, we now know that it is possible, if not likely, that problems can occur to the extent that police procedures reflect the instructional bias exemplified in the research. Even the very subtle instructional bias of omitting an option to reject the lineup produced a significant impact on accuracy in perpetrator-absent lineups. The need for education of court practitioners on this topic and implementation of standard lineup procedures was highlighted recently by an investigation of attorney sensitivity to lineup suggestiveness (Stinson, Devenport, Cutler, & Kravitz, 1996). This research suggested that the difficulties of attorneys in detecting and correcting lineup instructional bias may well limit the effectiveness of the current presence-of-counsel safeguard. Consideration of costs and benefits reveals that a rather simple change in lineup instruction could easily provide a missing safeguard against tainting of the lineup decision. This safeguard provides the potential to decrease false identifications while allowing for correct identifications. Even if the U-B effect is only a small component of the mistaken identification problem, its simple remedy should ensure use in police procedures.

ACKNOWLEDGMENTS

Portions of these data were originally presented at the AP-LS 1996 Biennial Conference, Hilton Head, South Carolina. I thank Gina Samsel and Stacy Keding for their help in data retrieval and analysis in the early stages of this project, and Gary Wells and three anonymous reviewers for their helpful comments regarding the manuscript. I particularly wish to thank those authors who, by sending their data or additional information, helped to make the data set more complete and precise.

REFERENCES

Asch, S. (1951). Effects of group pressure upon the modification and distortion of judgment. In H. Guetzkow (Ed.), *Groups, leadership, and men* (pp. 177–190). Pittsburgh, PA: Carnegie Press.

*Buckhout, R., Figueroa, D., & Hoff, E. (1975). Eyewitness identification: Effects of suggestion and bias in identification from photographs. *Journal of the Psychonomic Society, 6,* 71–74.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.) Hillsdale, NJ: Erlbaum.

*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., & Martens, T. K. (1987a). The reliability of eyewitness identification. *Law and Human Behavior, 11,* 233–258.

*Cutler, B. L., Penrod, S. D., O'Rourck, T. E., & Martens, T. K. (1986). Unconfounding the effects of contextual cues on eyewitness identification accuracy. *Social Behaviour, 1,* 113–134.

*Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987b). Improving the reliability of eyewitness identification: Putting context into context. *Journal of Applied Psychology, 72,* 629–637.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influence upon individual judgment. *Journal of Abnormal and Social Psychology, 51,* 629–636.

*Devenport, J. L., & Fisher, R. P. (1995). The effect of authority and social influence on eyewitness suggestibility and person recognition. Unpublished manuscript.

*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.

*Foster, R. A., Libkuman, T. M., Schooler, J. W., & Loftus, E. F. (1994). Consequentiality and eyewitness person identification. *Journal of Applied Cognitive Psychology, 8,* 107–121.

*Hall, D. F., & Ostrom, T. M. (1975, August). Accuracy of eyewitness identification after biasing or unbiased instructions. Paper presented at the meeting of the American Psychological Association, Chicago.

*Hilgendorf, E. L., & Irving, B. L. (1978). False positive identification. *Medicine, Science, and the Law, 18,* 255–262.

*Hosch, H. M., Leippe, M. R., Marchioni, P. M., & Cooper, D. S. (1984). Victimization, self-monitoring, and eyewitness identification. *Journal of Applied Psychology, 69,* 280–288.

Kassin, S. M., Ellsworth, P. C., & Smith, V. L. (1989). The "general acceptance" of psychological research on eyewitness testimony: A survey of the experts. *American Psychologist, 44,* 1089–1098.

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

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

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

Loftus, E. F. (1993). The reality of repressed memories. *American Psychologist, 48,* 518–537.

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

*Malpass, R. S., Devine, P. G., & Bergen, G. T. (1980). *Eyewitness identification: Realism vs. the laboratory.* Unpublished manuscript. State University of New York, Plattsburgh.

Milgram, S. L. (1965). Some conditions of obedience and disabedience to authority. *Human Relations, 18,* 57–76.

*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 and Human Behavior, 13,* 385–395.

*Paley, B., & Geiselman, R. E. (1989). The effects of alternative photospread instructions on suspect identification performance. *American Journal of Forensic Psychology, 7,* 3–13.

Rattner, A. (1988). Convicted but innocent: Wrongful conviction and the criminal justice system. *Law and Human Behavior, 12,* 283–293.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* Newbury Park, CA: Sage.

Seelau, S. M., & Wells, G. L. (1995). Applied eyewitness research: The other mission. *Law and Human Behavior, 19,* 319–324.

Sporer, S. L. (1992). Eyewitness identification accuracy, confidence, and decision-times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78,* 32–46.

Sporer, S. L. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118,* 315–327.

Stinson, V., Devenport, J. L., Cutler, B. L., & Kravitz, D. A. (1996). How effective is the presence-of-counsel safeguard? Attorney perceptions of suggestiveness, fairness, and correctability of biased lineup procedures. *Journal of Applied Psychology, 81,* 64–75.

*Warnick, D. H., & Sanders, G. S. (1980). Why do eyewitnesses make so many mistakes? *Journal of Applied Social Psychology, 10,* 362–366.

Wells, G. L. (1978). Applied eyewitness testimony research: System variables and estimator variables. *Journal of Personality and 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., Seelau, E. P., Rydell, S. M., & Luus, C. A. E. (1994). Recommendations for properly conducted lineup identification tasks. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 223–244). New York: Cambridge University Press.

*Study included in meta-analysis.