RECEIVED
SUPREME COURT
STATE OF WASHINGTON
Jan 13, 2012, 4:19 pm
BY RONALD R. CARPENTER
CLERK

No. 86119-6
161 Wn. App. 727, 255 P.3d 784, (May 9, 2011)
Court of Appeals No. 64466-1-I

RECEIVED BY E-MAIL

IN THE SUPREME COURT
OF THE STATE OF WASHINGTON

STATE OF WASHINGTON,

Respondent,

v.

BRYAN ALLEN,

Petitioner.

2012 JAN 25 A 8:3
BY RONALD R. CARPENTER
CLERK
FILED
SUPREME COURT
STATE OF WASHINGTON

BRIEF OF *AMICUS CURIAE* FRED T. KOREMATSU CENTER FOR LAW AND EQUALITY IN SUPPORT OF PETITIONER

Theodore J. Angelis, WSBA # 30300
Marie E. Quasius, WSBA # 42285
Ryan J. Groshong, WSBA # 44133
K&L GATES LLP
925 Fourth Avenue, Suite 2900
Seattle, WA 98104
(206) 623-7580

Attorneys for *Amicus Curiae*
Fred T. Korematsu Center for
Law and Equality

Robert S. Chang, WSBA # 44083
David A. Perez, WSBA # 43959

Amicus Curiae Fred T.
Korematsu Center for Law and
Equality

ORIGINAL

TABLE OF CONTENTS

Page

I. INTRODUCTION 1

II. STATEMENT OF INTEREST 3

III. ARGUMENT 5

A. Cross-racial Misidentification Is a Leading Cause of False Convictions and Warrants an Exercise of this Court's Rulemaking Authority to Require a Cautionary Instruction to Juries 6

B. A Jury Instruction on Cross-Racial Identification Would Prevent False Convictions and Improve Police and Prosecutorial Practices. 10

C. Issuing an Instruction Regarding the Unreliability of Cross-Racial Identifications Would Preserve the Right to Trial by Jury and Would Not Constitute an Impermissible Comment on the Facts. 14

IV. CONCLUSION 20

APPENDIX A
Effective and Ineffective Jury Instructions on Cross-Racial Identification A-1

APPENDIX B
Social Science Articles B-1

TABLE OF AUTHORITIES

Page

CASES

Ake v. Oklahoma,
470 U.S. 68, 105 S. Ct. 1087, 84 L. Ed. 2d 53 (1985) 10

Brennan v. City of Seattle,
39 Wash. 640, 81 P. 1092 (1905) 10

Duncan v. Louisiana,
391 U.S. 145, 88 S. Ct. 1444, 20 L. Ed. 2d 491 (1968) 14

Gianini v. Cerini,
100 Wash. 687, 171 P. 1007 (1918) 14, 16, 17

Perry v. New Hampshire,
No. 10-8974, 2012 WL 75048 (U.S. Jan. 11, 2012) 7, 10, 12

State ex rel. Foster-Wyman Lumber Co. v. Super. Ct. for King Cnty.,
148 Wash. 1, 267 P. 770 (1928) 6, 8

State v. Allen,
89 Wn.2d 651, 574 P.2d 1182 (1978) 11

State v. Badda,
63 Wn.2d 176, 385 P.2d 859 (1963) 17

State v. Barber,
170 Wn.2d 854, 248 P.3d 494 (2011) 19

State v. Brown,
113 Wn.2d 520, 782 P.2d 1013 (1989) 5, 7

State v. Carothers,
84 Wn.2d 256, 525 P.2d 731 (1974) 15, 16, 17, 18

State v. Clayton,
32 Wn.2d 571, 202 P.2d 922 (1949) 16

State v. Coates,
22 Wash. 601, 61 P. 726 (1900) 17

State v. Denney,
69 Wn.2d 436, 418 P.2d 468 (1966) 17

State v. Edwards,
23 Wn. App. 893, 600 P.2d 566 (1979) 19

State v. Gross,
31 Wn.2d 202, 196 P.2d 297 (1948) 17

State v. Harris,
102 Wn.2d 148, 685 P.2d 584 (1984) 7, 15
State v. Herwitz,
109 Wash. 153, 186 P. 290 (1919) 15
State v. Ito,
129 Wash. 402, 225 P. 63 (1924) 16
State v. Jacobsen,
78 Wn.2d 491, 477 P.2d 1 (1970) 14
State v. Johnson,
77 Wn.2d 423, 462 P.2d 933 (1969) 17
State v. Jordan,
17 Wn. App. 542, 564 P.2d 340 (1977) 19
State v. Lane,
125 Wn.2d 825, 889 P.2d 929 (1995) 14, 15
State v. Laureano,
101 Wn.2d 745, 682 P.2d 889 (1984) 5, 18, 19
State v. Lowenthal,
183 Wash. 14, 48 P.2d 909 (1935) 17
State v. Martin,
101 Wn.2d 713, 684 P.2d 651 (1984) 12
State v. Pearson,
37 Wash. 405, 79 P. 985 (1905) 17
State v. Perkins,
32 Wn.2d 810, 204 P.2d 207 (1949) 15
State v. Renfro,
96 Wn.2d 902, 639 P.2d 737 (1982) 7
State v. Warwick,
105 Wash. 634, 178 P. 977 (1919) 15
State v. Washington,
39 Wn.2d 517, 236 P.2d 1035 (1951) 16, 17

STATE STATUTES

RCW 2.04.190 ..... 6, 8
Washington State Constitution,
Article 4, Section 16 ..... 14, 15, 16, 18

OTHER AUTHORITIES

Brandon Garrett, Convicting the Innocent (2011) 8

Bruce W. Behrman & Sherrie L. Davey, *Eyewitness Identification in Actual Criminal Cases: An Archival Analysis*, 25 L. AND HUMAN BEHAV. 475 (2001) 7

CAROLINE WOLF HARLOW, U.S. DEP'T OF JUSTICE, DEFENSE COUNSEL IN CRIMINAL CASES 1 (2000) 9

Christian A. Meissner & John C. Brigham, *Thirty Years of Investigating the Own-Race Bias in Memory for Faces: A Meta-Analytic Review*, 7 PSYCHOL., PUB. POL'Y, & L. 3 (2001) 4, 7

Derek Simmonsen, *Teach Your Jurors Well: Using Jury Instructions to Educate Jurors about Factors Affecting the Accuracy of Eyewitness Testimony*, 70 MD. L. REV. 1044 (2011) 9

Douglas Vick, *Poorhouse Justice: Underfunded Indigent Defense Servs. and Arbitrary Death Sentences*, 43 BUFFALO L. REV. 329 (1995) 9

Edith Greene, *Judge's Instruction on Eyewitness Testimony: Evaluation and Revision*, 18 J. OF APPLIED SOC. PSYCHOL. 252 (1988) 12, 1

Eyewitness: How Accurate Is Visual Memory?, CBS NEWS, Mar. 8, 2009 ..... 1, 2

Gabriella Ramirez et al., *Judges' Cautionary Instructions on Eyewitness Testimony*, 14 AM. J. OF FORENSIC PSYCHOL. 31 (1996) 11, 1

Heather D. Flowe et al., *The Role of Eyewitness Identification Evidence in Felony Case Dispositions*, 17 PSYCHOL., PUB. POL'Y, AND L. 140 (2011) 5, 8

James M. Doyle, *Discounting the Error Costs: Cross-Racial False Alarms in the Culture of Contemporary Criminal Justice*, 7 PSYCHOL., PUB. POL'Y, & L. 253 (2001) 4

Jennifer E. Dysart et al., *Show-ups: The Critical Issue of Clothing Bias*, 20 APPLIED COGNITIVE PSYCHOL. 1009 (2006) 13

Jennifer Thompson, Op-Ed., *I Was Certain, but I Was Wrong*, N.Y. TIMES, June 18, 2000 1

KING COUNTY OFFICE OF THE PUBLIC DEFENDER, EXPERT SERVICES POLICY 2 (2008) 9

Richard S. Schmechel et al., *Beyond the Ken? Testing Jurors' Understanding of Eyewitness Reliability Evidence*, 46 JURIMETRICS 177 (2006) 7

Saul M. Kassin et al., *On the "General Acceptance" of Eyewitness Testimony Research*, 56 AM. PSYCHOLOGIST 405 (2001) 10

Sheri Lynn Johnson, *Cross-Racial Identification Errors in Criminal Cases*, 69 CORNELL L. REV. 934 (1984) 9

Task Force on Race and the Criminal Justice System, *Preliminary Report on Race and Washington's Criminal Justice System* (2011) 3, 4

Timing of Eyewitness Expert Testimony, Jurors' Need for Cognition, and Case Strength as Determinants of Trial Verdicts, 89 J. OF APPLIED PSYCHOL. 524 (2004) 10

U.S. Census Bureau, *Washington State & County QuickFacts*, CENSUS.GOV (2010) 4

I. INTRODUCTION

She was certain, but she was wrong. In 1984, Jennifer Thompson, a white woman, was raped by a black man. During the ordeal, Ms. Thompson "studied every single detail on the rapist's face" so that she would later be able to identify the perpetrator to the police. Jennifer Thompson, Op-Ed., *I Was Certain, but I Was Wrong*, N.Y. TIMES, June 18, 2000.[1] After providing a composite sketch, Ms. Thompson picked out her attacker from a photo array and also from a lineup. In her own words:

> I knew this was the man. I was completely confident. . . . I had picked the right guy, and he was going to go to jail. If there was the possibility of a death sentence, I wanted him to die. I wanted to flip the switch.

Id. Relying primarily on Ms. Thompson's identification testimony, and after deliberating for only 40 minutes, a jury sentenced 22 year-old Ronald Cotton to life in prison. *Eyewitness: How Accurate Is Visual Memory?*, CBS NEWS, Mar. 8, 2009.[2]

Eventually Mr. Cotton received a retrial, but only after spending years in prison. By sheer coincidence, another man, Bobby Poole, had ended up in the same prison as Mr. Cotton and had bragged to others about being the man who raped Ms. Thompson. Seeking long-awaited

---

[1] *Available at* http://www.nytimes.com/2000/06/18/opinion/i-was-certain-but-i-was-wrong.html (last visited Dec. 17, 2011).

[2] *Available at* http://www.cbsnews.com/stories/2009/03/06/60minutes/main4848039.shtml (last visited Dec. 15, 2011).

vindication, Mr. Cotton watched as officers brought Mr. Poole into the courtroom and stood him before Ms. Thompson. But she insisted that she had never seen Mr. Poole, and she reaffirmed her identification of Mr. Cotton as the man who raped her. Once again, the jury convicted and sentenced Mr. Cotton to life in prison. *Id.*

After 11 long years, DNA evidence confirmed that Bobby Poole had raped Ms. Thompson and that Ronald Cotton was innocent. Without the availability of DNA evidence and modern testing, Ronald Cotton would have spent the rest of his life in prison. Given the limited awareness among police, prosecutors, and jurors of the inherent unreliability of cross-racial eyewitness identifications, Mr. Cotton never stood a chance without conclusive exculpatory evidence.

In a subsequent interview, Mr. Cotton described the helplessness and anguish that he felt, first as a falsely accused defendant, and then as a wrongfully incarcerated man:

> She called my name, pointed a finger. And that's all, that's all it takes, it seemed like . . . . It felt like someone pushing a knife through me. . . . You know they say grown men [don't] cry, but it's a lie you know. I grabbed my pillow many times and hugged it, wishing I was hugging my mom, my dad, sister, brother. Wish it didn't have to be this way . . . .

Id. (internal quotation marks omitted).

Like Mr. Cotton, Petitioner Bryan Allen is a black man accused of a crime by a white victim, without substantial corroborating evidence.

Like Mr. Cotton, Mr. Allen was convicted without an instruction to the jury regarding the inherent unreliability of cross-racial identifications. But unlike Mr. Cotton, Mr. Allen will never be able to obtain DNA evidence that can prove his innocence. Because the unreliability of cross-racial identifications is now well established, Mr. Allen was entitled to a cautionary jury instruction. Because the trial court did not give such an instruction, Mr. Allen was denied his right to a fair trial. This Court should reverse.

II. STATEMENT OF INTEREST

The Fred T. Korematsu Center for Law and Equality ("Korematsu Center"), based at Seattle University School of Law, advances justice through research, advocacy, and education.[3] Currently, the Korematsu Center leads a statewide effort to understand the racial disproportionality that exists within Washington's criminal justice system. *See generally* Task Force on Race and the Criminal Justice System, *Preliminary Report on Race and Washington's Criminal Justice System* (2011) ("Task Force Report" or "TFR").

The Task Force has identified cross-racial misidentification as a factor contributing to disparity in Washington's criminal justice system. As explained in the Task Force Report, scientific research shows that

[3] The Korematsu Center does not, in this brief or otherwise, represent the

eyewitness identifications are particularly unreliable when the eyewitness is identifying a person of another race. *See* TFR at A-20. This phenomenon is particularly threatening to racial minorities. *See, e.g.*, James M. Doyle, *Discounting the Error Costs: Cross-Racial False Alarms in the Culture of Contemporary Criminal Justice*, 7 PSYCHOL., PUB. POL'Y, & L. 253, 253 (2001) (36 percent of all wrongful convictions uncovered by DNA testing nationwide involved white persons misidentifying black persons).[4] Studies have shown that white persons are more susceptible to committing cross-racial misidentification than racial minorities.[5] In addition, each member of a racial minority group faces a disproportionately large cross-racial eyewitness pool, including all white persons and all other racial minority groups.[6] Once misidentified, racial minorities in Washington State also face disparately higher rates of arrest, charging, and conviction, and also receive harsher sentences, even after controlling for legally relevant factors.[7]

---

official views of Seattle University.

[4] All cited social science articles are attached as Appendix B.

[5] *See* Christian A. Meissner & John C. Brigham, *Thirty Years of Investigating the Own-Race Bias in Memory for Faces: A Meta-Analytic Review*, 7 PSYCHOL., PUB. POL'Y, & L. 3, 21 (2001) ("Results indicated that White participants were more likely to demonstrate [own-race bias], especially with regard to false alarm responses.").

[6] *See, e.g.*, U.S. Census Bureau, *Washington State & County QuickFacts*, CENSUS.GOV (2010), *available at* http://quickfacts.census.gov/qfd/states/53000.html (last visited Nov. 20, 2011) (black persons comprise only 3.6 percent of Washington's population).

[7] *See* TFR at 1-2, 6-7, 14-15, 20, A-18, A-20 (2011) (finding that prosecutors, police, and other actors within the criminal justice system exercise discretion in ways that

Cross-racial misidentification contributes to racial disparity within Washington's criminal justice system and often results in wrongful conviction and false imprisonment. The Korematsu Center urges this Court to address the problem.

III. ARGUMENT

Numerous studies and exonerations have shown that cross-racial misidentification is a leading cause of false convictions. To promote justice, maintain fairness in criminal trials, and clarify the law, this Court should direct trial courts to provide a cautionary jury instruction regarding this unreliable category of evidence. The Court briefly addressed this issue, in *dicta* and without sufficient background, in *State v. Laureano*, 101 Wn.2d 745, 751, 768, 682 P.2d 889 (1984), *overruled on other grounds by State v. Brown*, 113 Wn.2d 520, 529, 782 P.2d 1013 (1989). Here, however, the Court is presented with relevant and compelling eyewitness science exposing the inaccuracy of cross-racial eyewitness testimony and the widespread unawareness of this phenomenon among jurors. Accordingly, the Court should require trial courts to provide a reasonably clear and effective cautionary instruction on cross-racial

---

reflect bias and/or result in disparate treatment of racial minorities); *see also* Heather D. Flowe et al., *The Role of Eyewitness Identification Evidence in Felony Case Dispositions*, 17 PSYCHOL., PUB. POL'Y, AND L. 140, 150 (2011) (study of felony cases from one of the largest District Attorney's Offices in the United States finding that when a stranger provided "eyewitness identification evidence [and there was some] limited corroborating

identification when such evidence is admitted and contested at trial. Such an instruction will prevent false convictions and improve police and prosecutorial practices, without invading the province of juries.[8]

A. CROSS-RACIAL MISIDENTIFICATION IS A LEADING CAUSE OF FALSE CONVICTIONS AND WARRANTS AN EXERCISE OF THIS COURT'S RULEMAKING AUTHORITY TO REQUIRE A CAUTIONARY INSTRUCTION TO JURIES.

This Court has inherent authority and broad discretion over all court rules and procedures—including jury instructions—and must modify those rules and procedures when necessary to promote fairness, and to ensure that defendants are not convicted unless the State proves guilt beyond a reasonable doubt. *See State ex rel. Foster-Wyman Lumber Co. v. Super. Ct. for King Cnty.*, 148 Wash. 1, 4–8, 267 P. 770 (1928) (noting that this Court must continue to analyze trial procedures and ensure proper rules that "promote justice"); *see also* RCW 2.04.190 (providing statutory authority to this Court "to prescribe, from time to time . . . practice and procedure to be used in all suits, actions, appeals, and proceedings of whatever nature"). Pursuant to this mandate, the Court should require a

---

evidence," prosecutors exercised discretion and decided not to press charges 17 percent of the time).

[8] The issue presented for this Court's review is whether the trial court erred by failing to provide a cross-racial identification instruction. That is closely related to whether the cross-racial identification evidence in this case should have been admitted in the first place. *See* ER 403. However, the admissibility of that testimony was not raised at trial or in the petition for review. *See* Pet. for Review at 2-3. Thus, the Korematsu Center will assume that the cross-racial identification evidence in this case was properly admitted and address only whether a cautionary instruction was necessary upon the

cautionary jury instruction on cross-racial identifications.

When confronted with certain dubious categories of evidence, this Court has required cautionary jury instructions to safeguard the fairness and integrity of trials. *See, e.g.*, *State v. Harris*, 102 Wn.2d 148, 155, 685 P.2d 584 (1984) (requiring cautionary instruction if accomplice testimony is to be admitted without sufficient corroboration), *overruled on other grounds by State v. Brown*, 113 Wn.2d 520, 554, 782 P.2d 1013 (1989); *State v. Renfro*, 96 Wn.2d 902, 906–07, 639 P.2d 737 (1982) (requiring cautionary instruction if stipulated polygraph evidence is to be admitted).[9]

Likewise, cross-racial identification testimony is a uniquely unreliable form of evidence.[10] As explained in the Korematsu Center's Memorandum in Support of the Petition for Review ("K.C. Mem."), most jurors are unaware of this cross-racial misidentification phenomenon. *See* K.C. Mem. at 5–6.[11] As the case against Ronald Cotton illustrates, when

---

admission of such evidence.

[9] *See also Perry v. New Hampshire*, No. 10-8974, 2012 WL 75048, at *10 (U.S. Jan. 11, 2012) (noting that the legality of admitting unscreened eyewitness identification evidence rests on "safeguards built into the adversary system that caution juries against placing undue weight on eyewitness testimony of questionable reliability," including "[e]yewitness-specific jury instructions").

[10] *See* Bruce W. Behrman & Sherrie L. Davey, *Eyewitness Identification in Actual Criminal Cases: An Archival Analysis*, 25 L. AND HUMAN BEHAV. 475, 482 (2001) (finding that, in general, witnesses in actual criminal cases falsely identified lineup foils 24 percent of the time); *see also* Meissner, *supra* note 5, at 15 (finding that cross-racial identification is 1.56 times more likely to result in a false identification); *id.* at 25 (finding that "expressed confidence or certainty about an identification is not a strong indicator of accuracy").

[11] *See also, e.g.*, Richard S. Schmechel et al., *Beyond the Ken? Testing Jurors' Understanding of Eyewitness Reliability Evidence*, 46 JURIMETRICS 177, 200 (2006)

jurors have no reason to doubt the reliability of cross-racial eyewitness testimony, they will rely on it to convict an innocent defendant.

The problem is not isolated. One recent study of rape, robbery, and assault cases found that cross-racial identifications were involved at least 30 percent of the time.[12] Worse yet, 11 percent of those cases were based on eyewitness identification with little to no corroborating evidence.[13] In Washington, at least five appellate decisions since 2010 have involved cross-racial identification. *See* K.C. Mem. at 4–5 (citing cases). Statistics also suggest that cross-racial identification may impact the likelihood of wrongful conviction for minorities: of 250 inmates freed by DNA evidence, 49 percent of the 190 disproportionately minority defendants misidentified by eyewitnesses involved a cross-racial identification.[14]

Washington's current trial court procedures, which this Court maintains and oversees, *see Foster-Wyman*, 148 Wash. at 4–8; RCW 2.04.190, do not address the problem of cross-racial misidentification and consequently perpetuate the injustices that result. For example, a

---

(noting that "almost two-thirds of jurors surveyed" were "ill-informed about the inaccuracy of cross-racial identification").

[12] *See* Flowe, *supra* note 7, at 140, 144-46, 149.

[13] *See id.* at 150.

[14] Brandon Garrett, Convicting the Innocent 72–74 (2011) ("Many more DNA exonerees were minorities (70 percent) than is typical even among average and already racially skewed populations of rape and murder convicts [in which] only approximately 30 percent to 40 percent of all rape convicts are minorities.").

defendant's ability to cross-examine witnesses tends to be ineffective against individuals who are making an honest and confident identification that happens to be false.[15] Additionally, because eyewitness experts are both scarce and expensive, expert testimony is rarely available.[16] Most felony defendants in state court are indigent, and the public defense agencies representing them lack sufficient funds to afford expert witness fees.[17] Thus, while cautionary instructions and expert testimony provide complementary mechanisms for protecting against wrongful convictions, jury instructions may be the only means available for effectively educating the jury on the unreliability of cross-racial identification.

With neither an expert nor a jury instruction to rely upon, criminal defendants cannot educate the jury as to the broad consensus within the scientific community regarding the unreliability of cross-racial

---

[15] *See, e.g.*, Sheri Lynn Johnson, *Cross-Racial Identification Errors in Criminal Cases*, 69 CORNELL L. REV. 934, 952-55 (1984).

[16] *See* Douglas Vick, *Poorhouse Justice: Underfunded Indigent Defense Servs. and Arbitrary Death Sentences*, 43 BUFFALO L. REV. 329, 375 & n.199 (1995) ("In the mid-1980s, specialists on eyewitness identification cost approximately $500 per day for courtroom testimony and $100 per hour for consultation . . . .").

[17] *See* CAROLINE WOLF HARLOW, U.S. DEP'T OF JUSTICE, DEFENSE COUNSEL IN CRIMINAL CASES 1 (2000), *available at* http://bjs.ojp.usdoj.gov/content/pub/pdf/dccc.pdf (last visited Dec. 15, 2011) ("At the end of their case approximately . . . 82% of felony defendants in large State courts were represented by public defenders or assigned counsel."); *see also* Derek Simmonsen, *Teach Your Jurors Well: Using Jury Instructions to Educate Jurors about Factors Affecting the Accuracy of Eyewitness Testimony*, 70 MD. L. REV. 1044, 1078 & n.282, 1079 & n.285 (2011) (citing studies finding a substantial lack of resources in public defender offices); KING COUNTY OFFICE OF THE PUBLIC DEFENDER, EXPERT SERVICES POLICY 2, 4 (2008), *available at* http://www.kingcounty.gov/courts/OPD/Partners/~/media/courts/OPD/documents/Expert_Services_FINAL_revised_6_20_11.ashx (last visited Dec. 17, 2011) (award of funds for

identification.[18] As a result, defendants are unable to argue to the jury that they are the victims of an unreliable and faulty cross-racial eyewitness identification, and are thus foreclosed from "the raw materials integral to the building of an effective defense." *Ake v. Oklahoma*, 470 U.S. 68, 77, 105 S. Ct. 1087, 84 L. Ed. 2d 53 (1985); *see also Perry*, 2012 WL 75048, at *5 ("The Constitution . . . protects a defendant against a conviction based on evidence of questionable reliability . . . by affording the defendant means to persuade the jury that the evidence should be discounted as unworthy of credit.").[19] In this case, by refusing to issue an instruction, the trial court denied Mr. Allen his right to mount an effective defense against the only piece of evidence indicting him.

B. A JURY INSTRUCTION ON CROSS-RACIAL IDENTIFICATION WOULD PREVENT FALSE CONVICTIONS AND IMPROVE POLICE AND PROSECUTORIAL PRACTICES.

The underlying purpose of a jury trial "is to ascertain the truth," *Brennan v. City of Seattle*, 39 Wash. 640, 645, 81 P. 1092 (1905), and the "purpose of an instruction is . . . to aid [the jury] in arriving at a proper

---

expert is discretionary, and generally must not exceed $500 for an eyewitness expert).

[18] *See* Saul M. Kassin et al., *On the "General Acceptance" of Eyewitness Testimony Research*, 56 AM. PSYCHOLOGIST 405, 412 (2001) (90 percent agreement among experts regarding the unreliability of cross-racial identification).

[19] Moreover, even if a defendant can afford expert testimony, jury instructions enhance the effectiveness of such testimony. *See* Michael R. Leippe et al., *Timing of Eyewitness Expert Testimony, Jurors' Need for Cognition, and Case Strength as Determinants of Trial Verdicts*, 89 J. OF APPLIED PSYCHOL. 524, 534 (2004) (finding that mock expert testimony "served to significantly lower perceptions of guilt and believability of the eyewitness, but only when . . . the judge . . . summar[ized] the

verdict, so far as it is competent for the court to assist them," *State v. Allen*, 89 Wn.2d 651, 654, 574 P.2d 1182 (1978). The truth-seeking purpose of a jury trial is jeopardized when jurors tend to rely upon a category of evidence that, unbeknownst to them, has been shown to be unreliable.

A jury instruction that properly informs jurors about the problems associated with cross-racial identification would assist juries in arriving at a proper verdict and promote better police and prosecutorial practices.

First, a carefully crafted jury instruction would decrease the likelihood of a false conviction by informing the jury of the unreliability of cross-racial eyewitness testimony.[20] In one study, subjects who received a clearly worded instruction were more likely to be aware of this phenomenon than those who received a less clear instruction or no instruction.[21] Further, in a study using actual Washington jurors, jurors were less likely to convict in a case involving eyewitness identification

---

expert's testimony in [jury] instructions").

[20] The Korematsu Center has provided a proposed jury instruction on cross-racial eyewitness testimony in Appendix A. This proposed instruction is informed by research on the efficacy of such jury instructions, and could serve as a useful model for trial courts and for the drafters of a needed pattern instruction. The Korematsu Center also has provided in Appendix A an example of a jury instruction that the research indicates would *not* be effective, so that trial courts and the drafters of a pattern instruction can avoid the mistakes contained therein.

[21] *See* Gabriella Ramirez et al., *Judges' Cautionary Instructions on Eyewitness Testimony*, 14 AM. J. OF FORENSIC PSYCHOL. 31, 55 (1996).

when given a clearly worded cautionary instruction.[22] In combination, these findings demonstrate that a well-crafted jury instruction can effectively inform juries of cross-racial misidentification and prevent false convictions.

Second, a required instruction on cross-racial eyewitness testimony would have substantial "upstream" effects on police and prosecutors and would go a long way toward mitigating the harms that cross-racial misidentifications cause. Judicial recognition that this category of evidence is problematic would encourage police and prosecutors to approach cross-racial identifications with caution and would emphasize the importance of obtaining corroborating evidence. *Cf. State v. Martin*, 101 Wn.2d 713, 723, 684 P.2d 651 (1984) (noting that a cautionary instruction on testimony involving hypnosis, among other safeguards, would "substantially reduce the likelihood of the use of hypnosis to develop evidence" because "the police officer or prosecutor considering hypnotizing a potential witness [would have to] proceed with caution").[23]

In this case, the police relied entirely on cross-racial identification, and did not conduct any further investigation or obtain any corroborating

---

[22] *See* Edith Greene, *Judge's Instruction on Eyewitness Testimony: Evaluation and Revision*, 18 J. OF APPLIED SOC. PSYCHOL. 252, 261-70 (1988).

[23] The United States Supreme Court also has acknowledged such "upstream effects" by recognizing that subjecting eyewitness identification evidence to special scrutiny may deter improper police conduct. *See Perry*, 2012 WL 75048, at *8.

evidence. In fact, the police even failed to obtain clearly relevant evidence that was readily available, such as Mr. Allen's clothing and sunglasses. *See* 10/21/09 RP 59 ("Q. [Y]ou didn't take pictures of the clothes? A. Unfortunately, no."); *see also* 10/20/09 RP 12 (sunglasses which may or may not have matched victim's original description were not taken into evidence).[24] Mr. Allen's case illustrates the type of low-cost (but effective) investigation and evidence-gathering that police and prosecutors will be more likely to forego if they know that reliance on a cross-racial identification will not be subject to a cautionary instruction to the jury. In this case, we are left with only a single cross-racial identification, devoid of any corroboration or explanation whatsoever.[25]

---

[24] The victim in this case initially reported being harassed by two individuals, one of whom was an "African American . . . wearing [] a black hoodie and jeans and . . . a baseball cap . . . [and] big sunglasses [] with . . . gold on the frames." 10/23/09 RP 3. The only potentially distinct portion of this description was the "gold" on the perpetrator's sunglasses. Some time thereafter, Mr. Allen was apprehended while in possession of a hat and large sunglasses. 10/21/09 RP 52, 70. But no evidence was obtained (testimonial or otherwise) regarding whether the sunglasses actually matched the victim's original description. The police officer who brought the victim to the scene to identify Mr. Allen from the back of a patrol car 15 to 20 feet away, *see id.* at 16-17, 51, later testified that he could not "remember [] the frames," *id.* at 52, describing the glasses as "what's in trend now, these huge sunglasses," 10/20/09 RP 8 (emphasis added). Inexplicably, the sunglasses were not taken into evidence. *See id.* at 12.

[25] The chance of a false identification in situations without corroborative evidence is substantial: In one recent study, 50 percent of subjects who interacted with one experimenter and were shortly thereafter shown a picture of a similar-looking individual wearing similar (but not identical) clothing falsely identified the person in the photograph as the experimenter. *See* Jennifer E. Dysart et al., *Show-ups: The Critical Issue of Clothing Bias*, 20 APPLIED COGNITIVE PSYCHOL. 1009, 1017-18, 1019 (2006) ("The results from these conditions suggest that if a person who resembles the perpetrator is apprehended near the scene of the crime, and is wearing distinct clothing similar to that described by the eyewitness, the likelihood of false identification is considerable."). This effect is magnified, of course, when a cross-racial identification is involved.

C. ISSUING AN INSTRUCTION REGARDING THE UNRELIABILITY OF CROSS-RACIAL IDENTIFICATIONS WOULD PRESERVE THE RIGHT TO TRIAL BY JURY AND WOULD NOT CONSTITUTE AN IMPERMISSIBLE COMMENT ON THE FACTS.

Under article 4, section 16, of the Washington State Constitution, "[j]udges shall not charge juries with respect to matters of fact, nor comment thereon, but shall declare the law." Const. art. 4, § 16 ("Section 16"). Section 16's primary purpose is to protect the right to trial by jury, *Gianini v. Cerini*, 100 Wash. 687, 689–90, 171 P. 1007 (1918), which itself is meant to safeguard against "the corrupt or overzealous prosecutor and the compliant, biased, or eccentric judge," and to ensure "community participation in the determination of guilt or innocence," *Duncan v. Louisiana*, 391 U.S. 145, 156, 88 S. Ct. 1444, 20 L. Ed. 2d 491 (1968). Section 16 protects these interests by "prevent[ing] the jury from being influenced by knowledge conveyed to it by the trial judge as to his opinion of the evidence submitted." *State v. Jacobsen*, 78 Wn.2d 491, 495, 477 P.2d 1 (1970). Accordingly, the "touchstone of error in a trial court's comment on the evidence is whether the feeling of the trial court as to the truth value of the testimony of a witness has been communicated to the jury." *State v. Lane*, 125 Wn.2d 825, 838, 889 P.2d 929 (1995).

Section 16 thus bars only those instructions from a trial court that comment on specific factual evidence within a given case. For example,

in *State v. Herwitz*, 109 Wash. 153, 186 P. 290 (1919), *overruled on other grounds by State v. Perkins*, 32 Wn.2d 810, 867, 204 P.2d 207 (1949), a trial court commented upon the credibility of a particular police officer who was testifying as to a specific conversation between himself and the defendant. 109 Wash. at 157. Although the truth of the testimony was contested, at trial the judge stated to the police officer: "You are a perfectly credible witness before this court, and you have not been impeached." *Id.* Because the officer's credibility was at issue, the judge's comment was "manifest error," requiring reversal. *Id.* at 158–60. Such a specific comment, regarding the credibility of a particular witness, is a violation of Section 16.[26]

In contrast, this Court repeatedly has upheld jury instructions that properly inform the jury about certain general categories of evidence. For example, in *State v. Carothers*, 84 Wn.2d 256, 525 P.2d 731 (1974), *clarified on other grounds by State v. Harris*, 102 Wn.2d 148, 152–53, 685 P.2d 584 (1984), the trial court gave "a standard instruction regarding accomplice testimony," directing that "the testimony of an accomplice . . . should be acted upon with great care and caution . . . ." 84 Wn.2d at 266. This Court held that such a general jury instruction did not run afoul of

[26] *See also, e.g., Lane*, 125 Wn.2d at 839 (trial court erred by declaring to the jury the reasons for a convict's early release, when that was in dispute); *State v. Warwick*, 105 Wash. 634, 636-37, 178 P. 977 (1919) (trial court erred by instructing the jury that a

Section 16:

> While the cautionary instruction may . . . apply only to one witness and the jury will have no doubt about the witness to whom the instruction is referable, the court does not give the jury its evaluation of the particular witness before it. Rather, it instructs the jury about the provisions of a rule of law applicable to the class in which the witness belongs.

84 Wn.2d at 269.[27] Cautionary instructions on the reliability of cross-racial identification fall squarely into this latter category, and thus, do not contravene Section 16.

Although Section 16 is concerned only with comments on specific, factual issues in a given case, and does not preclude general instructions about broad categories of evidence, this Court has recognized the need to deter trial courts from overstepping this boundary and improperly influencing juries through unfettered use of general instructions. *See Gianini*, 100 Wash. at 689–90 ("Such practice, if much indulged in, would . . . make easy the accomplishment of the very evil sought to be guarded against."). Consequently, this Court has restricted the use of cautionary instructions further than is strictly constitutionally necessary, by adhering

---

particular party had "used force" when that was a contested fact).

[27] *See also State v. Washington*, 39 Wn.2d 517, 520, 236 P.2d 1035 (1951) (distinguishing instructions "directed to particular witnesses, rather than special categories such as accomplices, experts, etc."); *State v. Clayton*, 32 Wn.2d 571, 574, 202 P.2d 922 (1949) ("What the court thereby told the jury was not that the uncorroborated testimony . . . in the instant case was sufficient to convict the appellant . . . but rather that in cases of this particular character, a defendant may be convicted based upon such testimony alone."); *State v. Ito*, 129 Wash. 402, 405-06, 225 P. 63 (1924) (noting that "[h]ad the court said that circumstantial evidence in the case is convincing . . . a different

to what amounts to a prophylactic rule.[28] In particular, this Court has limited the use of general cautionary instructions to only those "special categories" of evidence that present a particularly compelling need for an instruction. *State v. Washington*, 39 Wn.2d 517, 520, 236 P.2d 1035 (1951); *see, e.g.*, *Gianini*, 132 Wash. at 689 (refusing to allow a general cautionary instruction on casual statements, in the absence of any compelling need for such an instruction).

In *Carothers*, for example, the Court concluded that the unusual nature of accomplice testimony—incriminating testimony from an individual who admits to his/her own involvement in the crime and who often has ulterior motives—necessitated a cautionary instruction.[29] 84 Wn.2d at 268–69. In particular, this Court recognized that "courts have an expertise upon [the] subject," having observed many "innocent persons ...

---

situation would have been presented" (emphasis added)).

[28] A prophylactic rule is "a judicially-created [] measure designed to deter [violations of constitutional rights]" and "applies only when the benefits of its deterrent effect outweigh the cost to society of impairment to the truth-seeking function of criminal trials." *State v. Chenoweth*, 160 Wn.2d 454, 472 n. 14, 158 P.3d 595 (2007) (describing analogous federal exclusionary rule).

[29] Contrary to the State's assertion, Suppl. Br. of Resp. at 19–20, the accomplice instruction embodied in *Carothers* is not an outlier. For more than a century, Washington courts have required cautionary instructions in cases in which convictions rested on uncorroborated accomplice testimony. *State v. Coates*, 22 Wash. 601, 606, 61 P. 726 (1900); *State v. Pearson*, 37 Wash. 405, 410–15, 79 P. 985 (1905) (reversing conviction where trial court refused to instruct jury that "the testimony of an accomplice comes from a polluted source" and thus should be regarded with "great care and caution"); *State v. Lowenthal*, 183 Wash. 14, 19, 48 P.2d 909 (1935); *State v. Gross*, 31 Wn.2d 202, 216, 196 P.2d 297 (1948); *State v. Badda*, 63 Wn.2d 176, 180–81, 385 P.2d 859 (1963); *State v. Denney*, 69 Wn.2d 436, 440, 418 P.2d 468 (1966); *State v. Johnson*, 77 Wn.2d 423, 439, 462 P.2d 933 (1969).

sent to prison or death upon the testimony of an accomplice," and that "the ordinary citizen cannot be expected to have" the same expertise regarding the unreliability of accomplice testimony. *Id.* at 268.

This reasoning applies with greater force to cross-racial identification testimony. Like accomplice testimony, cross-racial identification testimony is inherently unreliable, and jurors, unlike courts, are largely unaware of this phenomenon. Other factors also set cross-racial identification apart from other types of evidence, meriting a cautionary instruction in this context. As noted, eyewitness experts are unavailable to most criminal defendants, and when they are available, they may be ineffective in the absence of a jury instruction. Furthermore, cross-examination and closing argument alone are entirely insufficient to present a legitimate defense in a case of cross-racial misidentification. For all of these reasons, cautionary instructions on cross-racial identification comport with both Section 16 and this Court's prudential rule confining the use of cautionary instructions to highly compelling cases.

In *Laureano*, this Court briefly addressed a trial court's refusal to provide a jury instruction on cross-racial misidentification. The relevant discussion in *Laureano* was *dicta*, however, because resolution of the cross-racial identification issue had no impact on the disposition of the case, which the Court already had remanded for a new trial due to the

erroneous admission of hypnotic testimony. *See Laureano*, 101 Wn.2d at 751, 768. The Court nonetheless decided to address some of the defendant's thirteen other claims of error, including a brief paragraph on cross-racial identification. *See id.* at 753–69. Notably, the Court in *Laureano* had not been presented with any data on the stark unreliability of cross-racial identification, nor the lack of juror awareness, and indeed, the state of eyewitness science was not nearly as developed then (in 1984) as it is today. The Court in *Laureano* relied instead on two Court of Appeals decisions that assumed cross-racial misidentification could be "more properly tested by examination and cross-examination," *State v. Jordan*, 17 Wn. App. 542, 545, 564 P.2d 340 (1977) (internal quotations omitted); *see also State v. Edwards*, 23 Wn. App. 893, 896–97, 600 P.2d 566 (1979), but that is an assumption we now know to be false. To the extent, if any, that *Laureano* constitutes applicable law, it is both incorrect and harmful and should be overruled. *See State v. Barber*, 170 Wn.2d 854, 864, 248 P.3d 494 (2011).

A carefully crafted jury instruction can address the problem of cross-racial misidentification without violating Section 16's mandate. In light of the advanced state of eyewitness science on cross-racial identification and this Court's ongoing duty to require jury instructions when necessary to ensure fairness and accuracy, this Court should require

a reasonably clear and effective cautionary instruction when cross-racial eyewitness testimony is admitted and contested at trial.

IV. CONCLUSION

This Court oversees and maintains a procedural system that currently fails to address the problem of cross-racial misidentification and the wrongful convictions that result. Defendants like Ronald Cotton and Bryan Allen are left without the ability to explain and argue to juries that cross-racial identifications are unreliable. These defendants have a right to a fair trial. Given the severity of this problem, and the compelling need for a well-crafted general jury instruction on cross-racial identification, this Court should reverse.

RESPECTFULLY SUBMITTED this 13th day of January, 2012.

K&L GATES LLP

By /s/

Theodore J. Angelis, WSBA # 30300
Marie E. Quasius, WSBA # 42285
Ryan J. Groshong, WSBA # 44133
Attorneys for *Amicus Curiae* Fred T. Korematsu Center for Law and Equality

Robert S. Chang, WSBA # 44083
David A. Perez, WSBA # 43959
Amicus Curiae Fred T. Korematsu Center for Law and Equality

APPENDIX A

Effective and Ineffective Jury Instructions on Cross-Racial Identification

A jury instruction on cross-racial identification should adhere to established principles of psycholinguistics, such as the use of logical structuring and the avoidance of uncommon words, passive constructions, and/or compound sentences. *See* Edith Greene, *Judge's Instruction on Eyewitness Testimony: Evaluation and Revision*, 18 J. OF APPLIED SOC. PSYCHOL. 252, 261–62 (1988). Sufficient specificity of information also is critical to inducing meaningful understanding. *Id.* at 262. In addition, recognizing that eyewitness confidence is not strongly indicative of reliability is critical to understanding the phenomenon of cross-racial misidentification. The same study finding substantial juror understanding of cross-racial misidentification after a clearly worded instruction also included a clear instruction on confidence, and found a similar associated effect on juror comprehension. *See* Gabriella Ramirez et al., *Judges' Cautionary Instructions on Eyewitness Testimony*, 14 AM. J. OF FORENSIC PSYCHOL. 31, 55 (1996).

With these principles in mind, the Korematsu Center proposes the following jury instruction on cross-racial identification as an effective model for trial courts and drafters of a needed pattern instruction:

> **The testimony given by an eyewitness is an expression of his or her beliefs, and those beliefs may or may not be accurate. You should take into account whether the defendant and the witness are of the same race, as people tend to have trouble identifying persons of another race, although this is not always the case. You should also take into account that the confidence displayed by an eyewitness does not necessarily indicate that the testimony is accurate. It is possible for eyewitnesses to be confident and still be wrong. Or, eyewitnesses may be unsure and still be correct in their identification. If, after examining the testimony you have reasonable doubt that an accurate identification was made, you must find the defendant not guilty.**

RECEIVED
SUPREME COURT
STATE OF WASHINGTON
12 JAN 17 AM 8:19
BY RONALD R. CARPENTER
CLERK

Cf. Ramirez, 14 AM. J. OF FORENSIC PSYCHOL. at 55, 60–62 (finding that a similar instruction was effective at inducing juror understanding of cross-racial misidentification and insignificance of eyewitness confidence).

In contrast, the Korematsu Center presents the following jury instruction on cross-racial identification as a relatively ineffective instruction, to be avoided:

> **It is not essential that the witness himself be free from doubt as to the correctness of his or her statement, however, you the jury must be satisfied beyond a reasonable doubt of the accuracy of the identification of the defendant before you may convict. Identification testimony is an expression of belief or impression by the witness, and in appraising the identification testimony of an eyewitness, you should scrutinize the identification with great care. You may take into account whether the defendant and the witness are of the same race.**

Cf. id. at 55, 58–60 (finding that an instruction similar to the instruction above was not effective at inducing juror understanding regarding eyewitness identification).

In this case, Mr. Allen proposed two alternative instructions that generally adhere to the principles outlined above. The instructions are fairly concise and straightforward, specific, and relatively easy to understand. Unfortunately, neither instruction addresses the issue of eyewitness confidence. But either instruction qualifies as a reasonably clear and effective jury instruction on cross-racial eyewitness identification.

APPENDIX B

Social Science Articles

Bruce W. Behrman & Sherrie L. Davey, *Eyewitness Identification in Actual Criminal Cases: An Archival Analysis*,
25 L. AND HUMAN BEHAV. 475 (2001) B-1

Christian A. Meissner & John C. Brigham, *Thirty Years of Investigating the Own-Race Bias in Memory for Faces: A Meta-Analytic Review*,
7 PSYCHOL., PUB. POL'Y, & L. 3 (2001) B-18

Edith Greene, *Judge's Instruction on Eyewitness Testimony: Evaluation and Revision*,
18 J. OF APPLIED SOC. PSYCHOL. 252 (1988) B-51

Gabriella Ramirez et al., *Judges' Cautionary Instructions on Eyewitness Testimony*,
14 AM. J. OF FORENSIC PSYCHOL. 31 (1996) B-64

Heather D. Flowe et al., *The Role of Eyewitness Identification Evidence in Felony Case Dispositions*,
17 PSYCHOL., PUB. POL'Y, AND L. 140 (2011) B-100

James M. Doyle, *Discounting the Error Costs: Cross-Racial False Alarms in the Culture of Contemporary Criminal Justice*,
7 PSYCHOL., PUB. POL'Y, & L. 253 (2001) B-120

Jennifer E. Dysart et al., *Show-ups: The Critical Issue of Clothing Bias*,
20 APPLIED COGNITIVE PSYCHOL. 1009 (2006) B-130

Michael R. Leippe et al., *Timing of Eyewitness Expert Testimony, Jurors' Need for Cognition, and Case Strength as Determinants of Trial Verdicts*,
89 J. OF APPLIED PSYCHOL. 524 (2004) B-145

Richard S. Schmechel et al., *Beyond the Ken? Testing Jurors' Understanding of Eyewitness Reliability Evidence*,
46 JURIMETRICS 177 (2006) B-163

Saul M. Kassin et al., *On the "General Acceptance" of Eyewitness Testimony Research*,
56 AM. PSYCHOLOGIST 405 (2001) B-201

Law and Human Behavior, Vol. 25, No. 5, October 2001 (© 2001)

Eyewitness Identification in Actual Criminal Cases: An Archival Analysis[1]

Bruce W. Behrman[2,3] and Sherrie L. Davey[2]

This study analyzed 271 actual police cases in order to address several prevalent issues in the eyewitness literature. Suspect identification (SI) rates were obtained for 289 photographic lineups, 258 field showups, 58 live lineups, and 66 lineup identifications preceded by earlier identifications. SI rates were assessed for 3 levels of extrinsic evidence: no extrinsic evidence, evidence of minimal probative value, and evidence of substantial probative value. The SI rates for the photographic lineups were assessed as a function of delay, same vs. cross-race conditions, witness type, and weapon presence. SI rates declined significantly over time; SI rates were significantly greater for the same-race condition. SI rates were much greater for field showups than photographic lineups, 76% vs. 48%. The SI rates for the field showups did not vary as a function of eyewitness conditions. The relation between confidence and suspect/foil identifications for the live lineups was significant and moderately high. The utility of archival identification studies for eyewitness testimony research is discussed.

The research base in the field eyewitness testimony has been growing steadily for years, but there have been few studies that delve into real police cases. Because research in eyewitness memory has scarcely been tested against actual cases, it has been criticized for lacking external validity (Tollestrup, Turtle, & Yuille, 1994; Yuille, 1993). The scarcity of archival studies has led to the recognition of a need for a diversity of approaches to the problems of eyewitness memory, and in particular a need for studies that investigate real crimes. Archival work provides a complexity of variables not possible to simulate in the laboratory, but of course laboratory experiments afford specific controls that cannot be duplicated in real life. Thus, archival work should not be considered a standard against which experimental studies should be evaluated,

[1] Portions of this research were presented at the meetings of the American Psychological Society, June 1999, Denver, CO; and June 2000, Miami, FL.

[2] Department of Psychology, California State University, Sacramento, California.

[3] To whom correspondence should be addressed at Department of Psychology, California State University, 6000 J Street, Sacramento, California 95819.

0147-7307/01/1000-0475$19.50/1 © 2001 American Psychology-Law Society/Division 41 of the American Psychology Association

but rather as a valuable additional approach to the study of eyewitness issues. Several archival studies have been published that assess descriptions provided by witnesses in criminal cases: Kuehn (1974), Sporer (1992), and Tollestrup et al. (1994). Gonzalez, Ellsworth, and Pembroke (1994) assessed identifications made by witnesses to actual crimes as an adjunct to an experimental study on the suggestiveness of showups, but to date only one large-scale investigation has examined real-life identifications made under a variety of witnessing conditions (Tollestrup et al.). The present study investigates identification data from actual police cases in order to build upon the few archival studies already done.

A review of the literature indicates that several of the eyewitness issues examined here have been studied in great detail by means of experimental methods: the effect of delay, cross versus own-race effects, weapon focus effects, and the relation between confidence and accuracy. A large number of laboratory studies indicate that delay is an important factor in determining identification accuracy, and that the number of correct identifications declines as the interval between the crime and the identification procedure increases. (Deffenbacher, Carr, & Leu, 1981; Egan, Pittner, & Goldstein, 1977; Malpass & Devine, 1981a, 1981b; Shepherd & Ellis, 1973). The cross-race effect has been replicated in a large number of studies. These studies indicate that recognition memory is better for culprits of one's own race than for those of a different race (Anthony, Cooper, & Mullen, 1992; Meissner & Brigham, in press). One of the most widely studied eyewitness phenomena is that of weapon focus. Several studies have found that the presence of a weapon negatively affects a witness's ability to recognize the perpetrator of a crime. (Kramer, Buckhout, & Eugenio, 1990; Loftus, Loftus, & Messo, 1987). Summarizing this research, Steblay (1992) conducted a meta-analysis of various experimental studies of the weapon focus effect and found a difference between weapon-present and weapon-absent cases, with the weapon-present cases resulting in fewer accurate identifications than the weapon-absent cases. The literature in general indicates that the correlation between confidence and accuracy is low (Bothwell, Deffenbacher, & Brigham, 1987; Cutler & Penrod, 1989). However, more recent studies have found that substantial correlations may occur under certain conditions. For instance, Sporer, Penrod, Read, and Cutler (1995) have shown that moderate correlations may be obtained when only choosers (those witnesses who choose either the suspect or a foil) are included in the analysis. Some recent work by Lindsay, Read, and Sharma (1998) indicated that accuracy–confidence correlations were substantial when they were based on data collapsed across a variety of witnessing conditions. To date only the two issues of retention interval and weapon presence have been assessed by archival work and that too in only one study: Tollestrup et al. (1994). Tollestrup et al. examined both the effect of delay and weapon focus on the rate of suspect identification and obtained data consistent with the general thrust of the laboratory findings. In the present study we extended the Tollestrup et al. study and assessed suspect identification rates as a function of delay, and weapon focus; in addition, we examined same-race bias and the relation between suspect/foil identifications and confidence.

Tollestrup et al. (1994) have argued that the results of most laboratory studies pertain primarily to unaffected witnesses to a crime. These researchers argue

that victims of crimes and some witnesses to a crime have different personal and emotional levels of involvement in the crime than subjects in a laboratory experiment. Consequently, although laboratory results may apply to unaffected witnesses to a crime, it may be that crime victims are more accurate in their descriptions and also more likely to identify the suspect. Tollestrup et al. found that victims of robbery identified the suspect more often (46.5%) than did witnesses of robbery (33.3%). In a 1982 experimental study, Hosch and Cooper researched the relative accuracy of witnesses versus victims. However, the findings of their laboratory work indicated that being the victim of a crime does not significantly improve or decrease the ability to accurately identify a perpetrator. In light of the Tollestrup et al. arguments that unaffected witnesses may process information about the crime differently than victims who are more directly involved, we examined identification outcomes for both victims and witnesses to the crimes.

An important variable affecting eyewitness behavior focuses on the type of procedure used to identify suspects. Two basic types of identification procedures can be found in the literature, lineups, and field showups. A showup refers to the observation of a single suspect by a witness in the field, typically at the crime scene, whereas a lineup refers to the presentation of the suspect and several foils, either live or via photographs. Most of the controversy in this area of research has focused upon the suggestiveness of field showups. Some researchers have found that single suspect identification procedures (showups) result in more false identifications than lineups (Wagenaar & Veefkind, 1992; Yarmey, Yarmey, & Yarmey, 1996). Wagenaar suggests that witnesses in real life may trust the police to have arrested the correct person, resulting in an increased false alarm rate. The author warns against one-person identification procedures. A third study echoes the concerns of Wagenaar and Veefkind regarding showups (Malpass and Devine, 1983). Although Malpass and Devine do warn against biasing practices in lineups and photospreads, the authors believe that if a lineup is constructed fairly, with several foils that are similar looking to the suspect, it may be the preferred method of identification. The authors suggest that these procedures "... distribute the probability of identification of an innocent suspect across the lineup foils, reducing the risk of an identification error" (p. 83). In contrast to the previously discussed findings, Gonzalez et al. (1993) show that witnesses at a lineup are less likely to say "not there" than are witnesses at a showup. The study found that police suggestion is no greater for showups than for lineups. The study by Gonzalez and his colleagues is an important one in that currently showups are the more typical and preferred method of identification used by police departments (Gonzalez et al.). The current study distinguished between several types of identification procedures, and in particular assessed the differences between field showups and photographic lineups. In addition, we investigated the SI rates for single-photo presentations and live (physical) lineups.

Finally, we compared identification rates of suspects at lineups conducted after a previous identification procedure, either another lineup or a showup, with lineups administered with no prior identification. Courts have acknowledged that prior identifications may affect later identifications (*Simmons v. United States,* 390 U.S. 386 n. 6, 1968). Generally, if an earlier identification is considered to be biasing, the later identification will be considered tainted unless the witness states that he can

distinguish between his memory of the perpetrator and that of the suspect in the earlier identification (*People v. Orozco*, 1981). One commentator has noted that it is difficult to determine whether the viewer is identifying the perpetrator or the person observed in the earlier identification procedure (Sobel, 1971). Sobel is referring to the influence of photographs on later identifications, but the same reasoning applies to the earlier observation of a live person. Sobel has suggested that "there will be few instances where the witness will fail to identify the person most recently viewed..." (p. 296). However, no real-life evidence has been gathered that indicates that earlier identifications do indeed affect later ones.

A major problem with archival work is that a certain percentage of real world lineups do not include the perpetrator of the crime. Tollestrup et al. (1994) generated various evidentiary levels and asserted that such a procedure provided a solution to the problem of perpetrator-absent lineups. Like Tollestrup et al. we have attempted to alleviate the problem through the use of various categories of extrinsic incriminating evidence.[4] In this study we used three categories of evidence. In a large proportion of cases no evidence was recorded. A number of cases contained evidence that was incriminating but not particularly strong (e.g., the suspect has been implicated in prior cases with similar methods of operation). Finally, a number of cases included evidence that strongly implicated the suspect (e.g., a confession). Of course, one cannot be sure that the culprit is in the lineup or showup, even when the case includes highly incriminating evidence. Consequently, suspect identification rates should not be interpreted as measures of accuracy. Inferences regarding the convergence of archival and laboratory work should be drawn with caution. Archival identification research may be used to provide various measures of the choosing behavior of real witnesses and, viewed as such, may be used to explore areas of convergence with and divergence from laboratory results.

An additional problem may limit archival work. Tollestrup et al. (1994) noted that the police (the Royal Canadian Mounted Police in this case) did not distinguish between misidentifications and rejections of lineups. Either response was noted as a failure to identify anyone. We were able to gather only a small number of misidentifications for our photographic lineups, probably for the same reason. The Sacramento police officers generally note in their reports either that a witness identified the suspect, that the witness rejected the lineup (or showup) or was simply unable to make an identification. In several cases involving photographic lineups, detectives did note that a witness had chosen a foil and which one he had chosen, but these cases were comparatively rare. We were able to tabulate and record all false identifications for live lineups. Physical lineups are more formal procedures. Here the police recorded each pick made whether it was a suspect or one of the foils. In sum, most of the analyses reported in this paper were based on suspect identification rates; the analyses of live lineups included the calculation of misidentifications.

[4]Extrinsic incriminating evidence in this paper refers to incriminating evidence extrinsic to the identification of the person. In some cases suspects were identified not only by facial features but also by voice or body type, such identification evidence was not considered extrinsic. The identification of clothing was considered extrinsic to the identification, as were other more obvious forms of incriminating evidence, for instance, confessions or fingerprints.

METHOD

Descriptive Information

This archival analysis used files from the Sacramento City Police Department and a group of police reports from crimes committed in several counties in Northern California, including Sacramento County. A total of 271 cases were analyzed.[5] The total number of crimes was 349; the vast majority were armed robberies ($n = 261$). The remaining crime types included a variety of felonies: residential burglary, assault with a deadly weapon, car jacking, and attempted homicide. The crimes were committed between 1987 and 1998. Several types of eyewitness identification procedures were analyzed including 258 field showups, 289 photographic lineups, 58 live lineups, and 18 single photo showups. Most photographic lineups in our sample contained five photographs. All of the live lineups consisted of six people. We also analyzed 66 identifications that had been preceded by an earlier identification procedure. In addition to gathering demographic data on both the witnesses and the perpetrator, including gender, race, and birth date, each file was also assessed for date, time, and type of crime committed, weapon used, and whether the identifier was a victim or witness to the crime. Information on the identification type (including time and date), the number of suspect identifications, and confidence level (for the live lineups only) of the identification was gathered.

As already noted, the data for this study is based on 271 police cases. There were 374 perpetrators ($\mu = 1.38$; range: 1–7 per case), 282 victims ($\mu = 1.09$, range: 1–2), and 224 witnesses ($\mu = 0.83$; range: 0–6) . The average age of the male perpetrators was 27.24 years; the range was 18–68 years. Only three perpetrators were female; their ages were 23, 34, and 43 years. There were 161 female and 91 male victims.[6] The average age of the female and male victims was 33.15 and 35.36 years respectively; the range was 16–74 years for males and 15 to 70 years for females. There were 102 male witnesses and 84 female witnesses. The average age of male witnesses was 28.56; the range was 16–51 years. The average age of female witnesses was 34.10 years; the range was 15–67 years.

Evidence Categories and Booking Rates

We classified the cases in terms of extrinsic evidence levels in order to evaluate the validity of the lineup and showup identifications. These categories of evidence are listed in Table 1. We have divided the cases into two major groups: cases for which there is incriminating extrinsic evidence and cases for which no extrinsic evidence was recorded in the police reports. The cases containing extrinsic incriminating evidence were divided into two groups: those that included evidence of substantial

[5]The first author was consulted in 130 of these cases (These cases included cases from Sacramento County and several other counties in Northern California). The remaining 141 cases were researched at the Sacramento Police Department. The SI rates for these two groups of cases were not significantly different, either for field showups or photographic lineups. Seventy-three of the cases on which the first author was consulted proceeded to trial. The police reports themselves did not contain information regarding the further disposition of the cases. Conviction rates of the cases that proceeded to trial are not available.

[6]Gender data for 68 eyewitnesses was not recorded in the files. The age of 121 eyewitnesses was not recorded in the files.

Table 1. Extrinsic Incriminating Evidence Categories

Minimal probative value (MPV)	Substantial probative value (SPV)
Concurrent identification of clothing, weapon, or Vehicle[a]	Suspect on surveillance tape
Anonymously provided information; Information provided by cofelon[b]	Independent identification of weapon, clothing, or vehicle
Similarities of MO[c]	Stolen items, weapon(s), drugs found in suspects possession (on person; in residence or vehicle)
Previous familiarity with suspect[d]	Physical evidence: Ballistic evidence, fingerprints, electronically marked bills
	Confessions
	Vehicle at scene registered to suspect
	Identification of unique characteristics (tattoos)

Note: Reasons for classifying evidence as MPV are as follows: [a]Evidence identified at the same time and place as the suspect; subject to suggestion, [b]Subject to bias, [c]Relevance, [d]Subject to expectations.

probative value (SPV) and those that contained only evidence of minimal probative value (MPV). The first category (SPV) contains evidence that is of substantial probative value and is highly incriminating (e.g., confessions or fingerprint evidence). The second category (MPV) contains evidence that is minimally probative of guilt. We consider this a weaker form of evidence because it either suffers from relevance problems (e.g., similar MOs) or is subject to bias, suggestion, (e.g., identification of the clothing at a field showup while the clothing is still on the suspect), or the expectations of the witness (e.g., familiarity with the suspect). If a case contained both MPV and SPV types of evidence, we placed it in the SPV category. An MPV + SPV group was created by combining the MPV and SPV cases. The largest group of cases, the no extrinsic evidence (NEE) group, contained no evidence extrinsic to identifications of the suspect. We included in this group cases containing single or multiple identifications of the suspect, but excluded cases with extrinsic evidence as defined in Table 1.

We also noted the percentage of suspects booked for each class of evidence, as stated in the police reports. In California, arrest and booking for a felony is based on probable cause (California Penal Code, 1999), thus booking rates provided an independent measure of evidence level. We found that 92% of the suspects who were identified from photographic lineups and whose cases fit into the first category of evidence (SPV) were in fact booked into jail, whereas 71% of such suspects in the second category (MPV) were booked. The difference disappears for field-showups: 96% of the suspects whose cases fall into either class of evidence, SPV or MPV, were booked. Booking rates for all evidence categories for the photographic and field identification procedures are shown in Table 2.

ANALYSIS AND RESULTS

A total of 689 identification attempts from police cases ranging from homicide to nonviolent theft were analyzed. These included 258 field showups, 284 photographic lineups and 58 live lineups. A breakdown of suspect identification (SI) rates for various evidence and identification categories are presented in Tables 3 and 4.

Table 2. Booking Rates by Identification Procedure and Evidence Category

	NEE		MPV		SPV		MPV + SPV		Total	
ID type	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*
Photo	57	188	71	41	92	55	83	96	66	284
Field	77	184	96	23	96	51	96	74	83	258

Note: NEE denotes cases with no extrinsic evidence, MPV indicates cases with evidence that is of minimal probative value, and SPV indicates cases with evidence that is of substantial probative value.

Approximately 48% of the witnesses who observed a photographic lineup identified the suspect as the perpetrator. Delay of identification was a major factor affecting SI rates for the photographic lineups. Overall, 55% of the witnesses picked the suspect from a photo lineup if the identification attempt was within a 0–7-day delay period, whereas only 45% did so after a delay of greater than 7-days. The disparity between delay groups was particularly evident when the cases included hard (SPV) extrinsic evidence, 64 and 33% for the 0–7-day and 8+-day delay groups $\chi^2(1, N = 55) = 4.89$, $p < .05$, $\phi = .30$. When the two extrinsic evidence categories (MPV and SPV) were combined, the differences between the delay groups were significant, $\chi^2(1, N = 96), = 5.48$, $p < .05$, $\phi = .24$.

Two hundred and thirty-one witnesses participated in cross-race versus same-race identifications from the photographic lineups. The overall SI rate for the cross-race cases was 45%. In same-race identifications the SI rate was 60%. This difference was significant, $\chi^2(1, N = 231) = 4.50$, $p < .05$, $\phi = .14$. The extrinsic evidence categories (with the exception of the SPV category) exhibited similar trends, but the differences were not significant. The presence of a weapon did not reduce identification rates. Overall, the percentage of witnesses who selected the suspect in

Table 3. SI Rates and Evidence Categories for Photographic Lineups

	NEE		MPV		SPV		MPV + SPV		Total	
Condition	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*
Days delay										
0–7	49	55	75	8	64[a]	22	66[a]	30	55	85
8+	47	133	48	33	33	33	41	66	45	199
Race										
Cross-racial	42	117	64	14	46	28	52	42	45[b]	159
Intraracial	61	41	80	15	37	16	58	31	60	72
Victim vs. witness										
Victim	48	85	57	21	50	18	54	39	50	124
Witness	44	79	62	16	38	29	47	45	45	124
Weapon focus										
Weapon used	45	170	61	36	44	34	53	70	48	240
No weapon used	61	23	20	5	48	21	42	26	51	49

Note: NEE denotes cases with no extrinsic evidence, MPV indicates cases with evidence that is of minimal probative value, and SPV indicates cases with evidence that is of substantial probative value.

[a]For SPV, $\chi^2(1, N = 55) = 4.89$, $p < .05$; for SPV + MPV, $\chi^2(1, N = 96) = 5.47$, $p < .05$.

[b]For Total, $\chi^2(1, N = 231) = 4.50$, $p < .05$.

Table 4. SI Rates and Evidence Categories for Field Showups

	NEE		MPV		SPV		MPV + SPV		Total	
Condition	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*
Across conditions	78	184	71	51	78	23	73	74	76	258
Race										
Cross-racial	81	129	66	30	76	77	70	47	78	176
Intraracial	71	49	72	18	100	3	76	21	73	70
Victim v. witness										
Victim	73	86	76	29	67	9	74	38	73	124
Witness	80	89	60	20	82	11	68	31	75	120
Weapon focus										
Weapon used	74	155	74	38	81	16	76	54	75	20
No weapon used	96	27	67	15	71	7	68	22	82	49

Note: NEE denotes cases with no extrinsic evidence, MPV indicates cases with evidence that is of minimal probative value, and SPV indicates cases with evidence that is of substantial probative value.

crimes in which a weapon was involved was 48%, but in crimes without a weapon the percentage was 51%. None of the weapon/no-weapon differences for the various evidence categories was significant.

Victims performed slightly better than witnesses when the identification attempts were assessed as a whole, 50% versus 45%, but this difference was not statistically significant; nor were the differences for the various evidence categories.

We collected a total of 258 field showups. The average delay was less than a day. Ninety-three percent of the field showups were administered within a day after the crime. The SI rates are given in Table 4. Overall the SI rate was 76%. None of the differences in the race, weapon, or victim/witness categories were significant. The difference between the SI rates for the photographic lineups and the field showups (76% vs. 48%) was significant, $\chi^2(1, N = 542) = 45.19$, $p < .001$, $\phi = .29$. The difference between the SI rate for the photographic lineups administered 0–7 days after the crime and the field showups (76% vs. 55%) was significant, $\chi^2(1, N = 343) = 13.81$, $p < .001$, $\phi = .20$.

We collected data for a total of 58 live lineup identification attempts. All of the identification attempts for the live lineups reported here were administered by the Sacramento Police Department. Witnesses to live lineups administered by the Sacramento Police Department, fill out a standard identification form. The witness was asked to check one of three response options: (1) "I am sure that number ___ was the person who ... [committed the crime]," (2) "Although I am not positive, I think number ___ was the person who ...," and (3) "I did not recognize anyone in the lineup as being the person who..." Consequently, for the live lineups we were able to count the SI rate, the false alarm rate, and the percentage of witness who marked "I did not recognize anyone ..." The SI rate was 50%, the false alarm rate was 24%; 26% of the witnesses were not able to make a choice. We computed χ^2 statistics and ϕ coefficients for the relationships between two dichotomous variables: (1) the choice of suspect or the choice of a foil versus (2) the statement that the witness was sure that the person was the culprit or the statement that the witness thought that the person was the culprit. A relatively high correlation was found between the degree of

confidence and whether a witness selected the suspect or a foil when all categories of evidence were combined, $\chi^2(1, N = 43) = 11.55$, $p < .001$, $\phi = .52$. Of the witnesses who expressed a moderate degree of confidence ("I think . . ."), the number of those who selected the suspect was 24, the number who picked a foil was 13. However, all but one of the witnesses who expressed a very high degree of confidence ("I am sure . . .") picked the suspect (19 picked the suspect, 1 picked a foil). We also computed the correlation between confidence and suspect/foil identification using the data from the combined extrinsic evidence categories (MPV + SPV). Here, the correlation was moderately high, $\chi^2(1, N = 34) = 8.09$, $p < 01$, $\phi = .49$. In this case, 15 of those who were "sure" of their pick chose the suspect, 1 chose a foil. Of those who expressed a moderate degree of confidence ("think") 19 chose the suspect, 10 chose a foil. Finally, we computed the correlation between confidence and suspect/foil picks using the data only from the highly incriminating evidence category (SPV). For this data, the correlation was again moderately high, $\chi^2(1, N = 27) = 6.07$, $p < .05$, $\phi = .47$. A very similar pattern of responses was evident: Twelve of the "sure" witnesses picked a suspect, only 1 chose a foil; 15 of the "think" witnesses chose the suspect, 8 chose a foil.

Our data contained 18 single-photograph identifications. These are analogous to field showups, yet we thought them to be sufficiently different to warrant a separate analysis. The SI rate for the single photos was .83. The average delay was less than a day. The SI rate is consistent with that of the field showups.

We computed the SI rates for 66 identifications that were preceded by earlier identification procedures. The later lineups were either photographic or live lineups. The earlier identifications were based on either field showups or photographic lineups. The average delay period for the earlier identifications was 11.43 days. The average delay for the later identification was 28.18 days. The SI rate for the later identifications was 62%. This rate was compared with an SI rate computed from a group of live ($n = 26$) and photographic ($n = 199$) lineups that had not been preceded by earlier identifications, 225 identification attempts in all, and that had occurred 8 or more days after the crime. The SI rate for this control group was 45%. The difference between the control group and group that had observed the suspect at an earlier identification procedure was significant, $\chi^2(1, N = 291) = 6.06$, $p < .05$, $\phi = .14$. We also assessed the number and type of responses for both the earlier and later identifications. Of the 33 witnesses who identified the suspect at the earlier procedure, 24 again identified him at the later one and 9 were not able to identify anyone at the later identification.[7] Of the 31 witnesses who were not able to make an identification at the first identification procedure, 17 were not able to make an identification later, but 14 did identify the suspect at the later identification. This means that 27% of the witnesses who made a positive earlier identification did not make a later identification, whereas 45% of the witnesses who did not identify the suspect at the earlier identification procedure did so at the later one. Finally, we computed the

[7]Some police reports did indicate that a witness had picked a foil from a photographic lineup. However, only a few reports indicated that a foil had been chosen. Consequently, we placed foil choices in a "no identification" category for the purpose of this analysis. Thus, responses in the "no identification" category were of three types: lineup rejections, choosing of foils, and responses indicating that the witness was not sufficiently certain to make a choice.

correlation between the SI rate and the length of delay between the first and second identifications.[8] The delay was divided into two intervals, delays from 0 to 30 days, and delays greater than 30 days. Of the 43 later responses that occurred 30 or fewer days after the first identification, 60% were positive identifications of the suspect; of the 23 later responses that followed the first identification by 31 or more days, 65% were positive identifications of the suspect. The correlation between the SI rate and delay was not significant, $\chi^2(1, N = 66) = .14$, $p > .05$, $\phi = 05$.

DISCUSSION

As expected, delay of identification was a major factor affecting SI rates for the photographic lineups. Our results indicate that the SI rate drops substantially after 7 days. This is true of all extrinsic evidence categories, MPV, SPV, and MPV + SPV. Thus, our findings converge with the laboratory evidence, which has, with few exceptions, indicated that the rate of identifying the culprit declines with time (Shapiro & Penrod, 1986). The decline in SI rate over time may invite the conclusion that recognition memory degraded over time, but it would be wrong to draw such a blanket conclusion from these data. A memory component is likely responsible for a portion of the decline in the SI rate. However, a host of variables influences witnesses' decisions in real cases. For example, witnesses to real crimes may simply become more cautious as the delay between the crime and the identification increases. Our results are also consistent with those of Tollestrup et al. (1994). We did not find a decline over the first few (0–2) days as did Tollestrup et al. We computed SI rates for each day for delays up to 30 days. The rates were similar, hovering around 60–70% (with the exception of the NEE category) for each of the days within the first week, and bottomed out after the first week at about 40%.

The tendency to choose the cross-racial suspect from the photographic lineups is lower than the tendency to pick the intrarace suspect in all evidence categories, except the SPV category. These results are in line with the effects found by many in the field (Malpass & Kravitz, 1969; Malpass, Laviqueur, & Weldon, 1973; Platz & Hosch, 1988; Shepherd, Deregowski, & Ellis, 1974). Overall, our findings are consistent with the meta-analytic work on cross-race effects indicative of a small-to-moderate decrease in the rate of identifying the culprit when the identification is cross-racial (Anthony et al., 1992; Meissner & Brigham, in press). A reversal occurs in the SPV group; this is puzzling and at odds with the literature in general. As already noted, the perpetrator is more likely to be present in the SPV group. It is possible that witnesses tend to pick a suspect of a different race when there is even a vague memory of the perpetrator available to them and when the suspect is in fact the perpetrator. It may be that a witness's decision criterion is somewhat less strict for individuals of a different race (Slone, Brigham, & Meissner, 2000), and that this tendency to choose other-race faces manifests itself when the perpetrator is present in the lineup. Of course the number of witnesses in the SPV group is small, and the difference between the same and cross-race SPV groups are not statistically significant. In general our data indicate

[8]We thank an anonymous reviewer for suggesting that we compute this correlation.

that witnesses do not pick suspects from a different race more often than those of their own race.

A host of studies have supported the phenomenon of weapon focus, including meta-analytic studies (Kramer et al., 1990; Loftus et al., 1987; Steblay, 1992); however, support for the weapon focus effect was not evident in the present study. Further, the suspect was chosen more often in the weapon-present group than in the weapon-absent group for both the MPV and MPV + SPV conditions. One possible explanation for our finding is that weapon-focus is simply not a real-life phenomenon. Perhaps the arousal level during a crime is simply too high in real life crimes, whether a weapon is present or not. The emotions of fear or anger, associated with weapon-present crimes, are not present to the same extent in a laboratory as in a real crime (Yuille, 1993). Still, this explanation does not explain the difference between the results of our archival study and the archival study of Tollestrup et al. (1994) that corroborated the laboratory findings. It should be pointed out, however, that when an analysis of covariance was performed on the Tollestrup et al. data the presence of a weapon had only a marginally significant effect on the suspect identification rate.

Our findings do not indicate that victims respond differently than witnesses in identifying suspects. Tollestrup et al. (1994) and Yuille (1993) have suggested the emotional arousal experienced by victims might result in different response levels for victims than for unaffected bystanders or even victims of nonviolent crimes, for example, fraud cases. Tollestrup et al. did find a significant association between witness types (robbery victims, robbery witnesses, and fraud victims) across evidence levels. Tollestrup et al. found that SI rates were related to witness type. Across evidence categories the SI rates were 46.5, 33.3, and 25.5 for robbery victims, robbery witnesses, and fraud victims respectively. We found the SI rate only slightly higher for victims than witnesses overall. The SI rate was higher for victims than for witnesses in both the SPV and the combined extrinsic evidence categories, MPV + SPV, although the percentages were reversed for the MPV group. None of these differences were significant. We agree with Tollestrup et al. that a victim should be defined as one who "directly interacts" (p. 146) with the perpetrator. We evaluated the victims and witnesses as designated in the police reports in terms of Tollstrup's "direct interaction" criterion and found that victims did indeed experience a more direct interaction with the perpetrator than the witnesses. However, in many of our cases a witness may have experienced a high degree of emotion even though they had not been directly involved (e.g., a person in line during bank robbery as compared with the teller who had been held up by the robber). Thus our sample of witnesses may have included witnesses who were quite emotionally affected by the crime even though they had not directly interacted with the perpetrator. The fact that our witness group may have contained witnesses who were highly emotional may account for the fact that our data do not clearly support the Tollstrup et al. results. In addition, it should be noted that Tollestrup et al. found a significant association between witness types and identification outcomes when comparing robbery victims, robbery witnesses, and fraud victims, but a subsequent post hoc multiple comparison test did not reveal significant differences between these groups.

The relationship between confidence and suspect/foil identification for the live lineups is a solid one; indications are that highly confident witnesses choose fewer distracters than moderately confident ones. We obtained significant correlations for the highly incriminating evidence category (SPV), the inclusive evidence category (MPV + SPV), and for the data computed across all categories of evidence. Only one highly confident witness in the sample of witnesses at live lineups picked a foil. Wells and Bradfield's (Wells & Bradfield, 1998) observation that detectives may inadvertently cue witness when they know who the suspect is may be correct in certain cases; such inadvertent cuing may have happened in some of our cases. It must be noted however that the Sacramento Police Department is aware of the effects of bias on the part of detectives. According to the Identification Administrator for the SPD (H. Ayers, personal communication, September 10, 1997) detectives are instructed to be careful concerning any statements or actions they might make during or after the identification, which might be considered suggestive. The SPD detectives are instructed in the contents of *The California Police Officers Legal Sourcebook* (California Department of Justice, 1993), including the chapter on lineups and showups. Section 8C of that chapter states: "... if you tell a witness that he has picked the 'right' or ('wrong') person, it may jeopardize the admissibility of later in-court identifications." (p. 229). Of course the detectives present at the lineup may still unconsciously cue witnesses, but at least their awareness of the problem may mitigate some of the suggestiveness inherent in the post-lineup situation.

The bulk of the experimental literature indicates that the confidence–accuracy relation is a weak one (Sporer et al., 1995). However, a recent trend in the experimental literature indicates that under some circumstances confidence may be a good predictor of accuracy. Lindsay et al. (1998) obtained confidence–accuracy correlations of .51 and .68 for two different videos when the witnessing conditions varied widely. Sporer et al. (1995) found a substantial confidence–accuracy correlation when the analysis was limited to witnesses who made positive identifications. Our results parallel those of Lindsay et al. (1998) and Sporer et al. (1995); however, it would be wrong to infer that our data indicate that confidence predicts accuracy. Our findings do indicate that highly confident witnesses are much more prone to choose the suspect in a criminal proceeding than are moderately confident ones. Further, it is more likely that the witnesses' choices are correct in the SPV cases than other cases. It also should be emphasized that our confident witnesses (except for one) did not make misidentifications, and in this sense were seldom inaccurate.

The data from the field showups is inconsistent with those obtained by Gonzalez et al. (1994), which indicated that field showups produced fewer suspect identifications than lineups. Two of the studies reported in the Gonzalez et al. article were laboratory studies; the SI rates, in both the perpetrator-present lineups and the perpetrator-absent lineups were greater for lineups than the showups. Study 3 of the Gonzalez et al. report was an archival study. The percentage differences were striking. Only 22% of the witnesses positively identified the suspect at the showups, whereas 75% did so when photographic lineups were used. It should be noted that the sample used to compute these lineups figures was small ($n = 24$). Our findings show that the suspect identification rate is highest for field showups. Our results indicate that field showups produced a high rate of suspect identification, a finding that

is consistent with the hypothesis that showups are suggestive. The SI rates were 76% for field showups and 48% for photographic lineups, $n = 258$ and 284 respectively. These differences carry throughout all of the evidence levels. Of course, some of the differences reported in our study may be confounded by the amount of delay. However, we found that the SI rate for lineups administered within the first week were significantly smaller than those for showups. The relatively high SI rate for the lineups reported by Gonzalez et al. may in part be due to the small sample, but it is the low SI rate (22%) for showups, which is troubling. The Gonzalez et al. sample of showups is large ($n = 172$). At this point we have no explanation for the differences between the two studies. At any rate, our results are consistent with the laboratory findings that showups are biasing (Yarmey et al., 1996). In addition, our study confirms the intuition of the United States Supreme Court that the risks of bias are greater in showups than lineups (*Stovall v. Denno*, 1967). Several other courts have expressed the opinion that showups are suggestive (*People v. Johnson*, 1989; *People v. Orozco*, 1981). In fact, this opinion is expressed in the National Police Foundation report, *Model Rules: Eyewitness Identification* (Project on Law Enforcement Policy and Rulemaking, 1974), and in the U.S Department of Justice document, *Eyewitness Evidence: A Guide for Law Enforcement* (Technical Working Group for Eyewitness Evidence, 1999), both of which indicate that showups are biasing, and suggest methods to mitigate the bias. It might be noted that the booking rate for the showups in our study for which there is no extrinsic evidence is 77%, whereas the rate for the extrinsic evidence categories is about 96%; nevertheless, the rate of identification for the no-evidence category is 78%, equal to or slightly higher than it is for those cases that include extrinsic incriminating evidence. Certainly, booking rates are not a failsafe method of assessing perpetrator-absent identifications, but it seems reasonable to argue that a showup category with a 23% nonbooking rate contains more perpetrator-absent showups than does a category with a 4% nonbooking rate. Consequently, the fact that the NEE category has a *higher* SI rate than the SPV or the MPV + SPV extrinsic evidence categories provides additional support for the view that showups are indeed biasing.

The field showup data was assessed in terms of both evidence categories and the typical eyewitness categories of race, weapon presence, and victim/witness differences. In general, the SI rates were high for the various groups. The cross-race groups produced lower SI rates than the own-race groups for all incriminating evidence categories, a trend consistent with the experimental data. The weapon-present groups in the MPV, SPV, and the MPV + SPV evidence categories produced higher SI rates than the weapon-absent groups, a trend not expected from the laboratory data. None of these categories produced significant differences. It may be that the images generated in field showups are too fresh and vivid to allow for the variables of race, weapon presence, or witness type to produce large, statistically significant effects.

The current study clearly indicates that witnesses at field showups react differently than those assessing a photographic lineup. We have found that witnesses demonstrate high SI rates at field confrontations. But perhaps of greater interest is the fact that none of the classic eyewitness factors, race, weapon presence, or witness type, produced significant or even marginally significant effects when the identifications were made at field showups. If most jurisdictions identify suspects primarily

through the use of showups, (Gonzalez et al., 1993) many of the factors found to be influential in the laboratory are irrelevant in a large proportion of real cases. Certainly later identifications might be based on photographic or live lineups, but these would likely be tainted by the earlier field identification. The finding that SI rates do not vary with many of the eyewitness factors that influence witness behavior at lineups does not mitigate the biasing effect of field showups. However, it does underline the point that the eyewitness literature may have to take into account the methods of identification to a far greater extent than in the past when assessing the effects of situational or witness variables on identification.

The data comparing the identification rates of suspects at lineups conducted after witnesses have observed the suspect at an earlier procedure with lineups conducted with no prior identification procedure, are consistent with the hypothesis that earlier identifications may contaminate later ones. Further, the data tracking the changes in "yes" and "no" responses (which included rejections, "I don't know" responses and false alarms) indicate that a process similar to the concept underlying unconscious transference is operating here (Ross, Ceci, Dunning, & Toglia, 1994). A large number of witnesses who say "no" or "I don't know" in the earlier identification say "yes" later. Apparently, these witnesses did distinguish, at least to some degree, the perpetrator and the suspect at the earlier lineup. Either the witness recalled that the person in the preliminary identification procedure did not fit their image of the perpetrator or their image of the perpetrator was simply too vague to say "yes." Evidently, whatever image the witnesses had of the suspect at an earlier time became more dominant at the later identification. The witness's image of the perpetrator could have been overwritten by the image of the suspect in the earlier identification procedure; in other cases the witness could have confused the sources of the images (Ross et al., 1994). If a witness rejected the earlier identification because they had two distinct images, that of the perpetrator and that of the identification picture, then at the later identification procedure he may have recalled the image in prior lineup as the one at the crime scene. Whether the process involved is one of overwriting or one of source exchange, the data indicate that the image observed at the earlier identification has become the one associated with the perpetrator. It would then be difficult for a witness to any longer distinguish his memory of the perpetrator from that obtained at the preliminary identification. In either case these data are consistent with the laboratory findings on bystander misidentification. These findings do argue against putting a suspect in a later lineup, when he has been observed at an earlier one. Chances of identifying the suspect are greatly increased and this is the case even when the later identification is administered some months after the crime. Any argument that a live lineup could be used for elimination purposes after the suspect has once been identified is highly questionable in light of this data.

There are several limitations to this study in need of mention. The cases cover a broad range of crimes. Our set of crimes was similar to those of the Tollestrup et al. (1994) study in that our cases included a wide variety of robberies; however, our set was slightly broader than Tollestrup et al.'s in that it contained additional types of crimes, for example, residential burglaries. The essential aspects of the crimes in our study were that the victims were aware that a crime was taking place, and that the observation times were relatively short. The victims in our cases were quite

likely attentive to the perpetrator, unlike many victims of fraud (Tollestrup et al.). Consequently, our results should not be applied to cases involving fraudulent activity in which the victims may not realize the culprit is a criminal until long after the event. There were certainly a wide range of viewing conditions and viewing times in our study; still, the crimes did not provide lengthy opportunities for rehearsal as did, say, the homicide assessed in Yuille and Cutshall's seminal field study (Yuille & Cutshall, 1986). In the current study, the crimes occurred relatively quickly, but future work might vary viewing conditions such as time to observe, in order to assess the interaction of viewing times with other situational factors.

It was difficult to obtain large samples for a number of analyses. We found it difficult to achieve large samples of cases in either the MPV or the SPV categories. Weapon-absent cases were particularly difficult to come by. Another problem was the difficulty in obtaining live lineups. These are relatively infrequently administered in California and, we gather, in other jurisdictions (Tollestrup et al., 1994). Thus, our sample of live lineups was restricted. This in turn limited our sample of confidence ratings, because a standard scale of confidence is administered only for live lineups by the Sacramento Police Department. Many expressions of confidence were recorded in the police reports for the photographic lineups, a very large category of identifications; however, as noted previously, the police reports contained only a few instances of recorded misidentifications, making it impossible to obtain a large sample of confidence–suspect/foil relations.

The current study clearly demonstrates the validity of comments by Tollestrup et-al. (1994) to the effect that forensic research is multifaceted and that factors affecting eyewitness behavior may vary from one context to another. One of the problems with the literature to date has been the assumption that eyewitness factors have the same effect and that witnesses respond similarly regardless of the forensic situation (Tollestrup et al., 1994). A diversity of methods is needed if we are to provide the legal profession with practical advice regarding eyewitness memory. It is from a combination of methods, controlled experiments, field studies, and archival studies that conclusions should be drawn. An examination of the similarities and differences between the identification outcomes generated by the different approaches may determine when one approach is applicable in contexts other than its own. The results of the study reported here converge with several eyewitness findings from the laboratory: the effect of delay and of cross-race identification, and the biasing effect of showups; however, our results diverge from the experimental literature with regard to weapon focus effects, and to some extent the relationship between confidence and suspect identification. Thus, the present study highlights the need for several approaches to eyewitness research.

ACKNOWLEDGMENTS

We thank the Sacramento Police Department (SPD) for providing access to its police reports on major crimes and to the Records Department of the SPD for assisting us in researching the files. We especially thank Capt. M. Busch and Harold Ayers of the SPD for their invaluable assistance with this research. We thank Carole

W. Barnes for providing the initial liaison with the SPD. We thank Susan Kollenberg and Shannon Corey for their time and effort in categorizing and collating the massive amount of data collected. We thank Nadia Dauod, Rita Mechill, and Mona Stilson for their extraordinary contributions to the collection of data for this research. Critical comments of Betsy Feldman on an earlier version of the article are gratefully acknowledged.

REFERENCES

Anthony, T., Cooper, C., & Mullen, B. (1992). Cross-racial facial identification: A social cognitive integration. *Personality and Social Psychology Bulletin, 18,* 296–301.

Bothwell, R., Brigham, J., & Malpass, R. (1989). Cross-racial identification. *Personality and Social Psychology Bulletin, 15,* 19–25.

Bothwell, R., Deffenbacher, K., & Brigham, J. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology, 72,* 691–695.

California Department of Justice. (1993). *California Peace Officer's legal sourcebook.* Sacramento: Author.

California Penal Code, West's Annotated California Codes: Penal Code S 836 West Group 1999.

Cutler, B., & Penrod, S. (1989). Forensically relevant moderators of the relation between eyewitness identification accuracy and confidence. *Journal of Applied Psychology, 74,* 650–652.

Deffenbacher, K., Carr, T., & Leu, J. (1981). Memory for words, pictures and faces: Retroactive inference. *Journal of Experimental Psychology, Human Learning and Memory, 7,* 299–305.

Egan, D., Pittner, M., & Goldstein, A. (1977). Eyewitness identification: Photographs vs. live models. *Law and Human Behavior, 1,* 199–206.

Gonzalez, R., Ellsworth, P. C., & Pembroke, M. (1993). Response biases in lineup and showups. *Journal of personality and Social Psychology, 64,* 525–537.

Hosch, H. M., & Cooper, D. S. (1982). Victimization as a determinant of eyewitness accuracy. *Journal of Applied Cognitive Psychology, 67,* 649–652.

Kramer, T., Buckhout, R., & Eugenio, P. (1990). Weapon focus, arousal and eyewitness memory. *Law and Human Behavior, 14,* 167–184.

Kuehn, L. L. (1974). Looking down a gun barrel: Person perception and violent crime, *Perceptual and Motor Skills, 39,* 1159–1164.

Lindsay, S. D., Read, J. D., & Sharma, S. (1998). Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. *Psychological Science, 9,* 215–219.

Loftus, E., Loftus, G., & Messo, J. (1987). Some facts about "weapon focus." *Law and Human Behavior, 11,* 55–62.

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

Malpass, R., & Devine, P. (1981b). Guided memory in eyewitness identification. *Journal of Applied Psychology, 66,* 343–350.

Malpass, R. S., & Devine, P. G. (1983). Measuring the fairness of eyewitness identification lineups. In S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating witness evidence* (pp. 81–102). New York: Wiley.

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology, 13,* 330–334.

Malpass, R., Laviqueur, H., & Weldon, D. (1973). Verbal and visual training in face recognition. *Perception and Psychophysics, 14,* 285–292.

Meissner, C. A., & Brigham, J. C. (in press). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law.*

People v. Johnson, 210 Cal. App. 3d. 316 (1989).

People v. Orozco, 114 Cal. App. 3d. 435 (1981).

Platz, S., & Hosch, H. (1988). Cross-racial/ethnic eyewitness identification: A field study. *Journal of Applied Social Psychology, 18,* 972–984.

Project on Law Enforcement Policy and Rulemaking. (1974). Model rules: Eyewitness identification. Washington, DC: Police Foundation.

Ross, D. F., Ceci, S. J., Dunning, D., & Toglia, M. P. (1994). Unconscious transference and mistaken identity: When a witness identifies a familiar but innocent person. *Journal of Applied Psychology, 79,* 918–930.

Shapiro, P., & Penrod, S. (1986). Meta-Analysis of facial identification studies. *Psychological Bulletin, 100,* 139–156.

Sheperd, J. W., Deregowski, J. B., & Ellis, H. D. (1974). A cross-cultural study of recognition memory for faces. *International Journal of Psychology, 9,* 205–211.
Shepherd, J., & Ellis, H. (1973). The effect of attractiveness on recognition and memory for faces. *American Journal of Psychology, 86,* 627–633.
Simmons v. United States, 390 U.S. 377 (1968).
Slone, A. E., Brigham, J. C., & Meissner, C. C. (2000). Social and cognitive factors affecting the own-race bias in whites. *Basic and Applied Social Psychology, 22,* 71–84.
Sobel, L . (1971). Assailing the impermissible suggestion, *38 Brooklyn Law Review*, 263–271.
Sporer, S. L. (1992). *Person descriptions in an archival analysis of criminal cases.* Unpublished manuscript, University of Marburg, Germany.
Sporer, S. L., Penrod, D., Read, S., & Cutler, B. (1995). Choosing, confidence and accuracy: A meta-analysis of the confidence relation in eyewitness identification studies. *Psychological Bulletin, 118,* 315–327.
Steblay, N. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior, 16,* 413–424.
Stovall v. Denno, 388 U.S. 293 (1967).
Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement.* Washington DC: U.S. Department of Justice, Office of Justice.
Tollestrup, P., Turtle, J., & Yuille, J. (1994). Actual victims and witnesses to robbery and fraud: An archival analysis. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 144–160). New York, NY: Cambridge University Press.
Wagenaar, W. A., & Veefkind, N. (1992). *Comparison of one-person and many-person lineups: A warning against unsafe practices.* Unpublished manuscript, Leiden University, Leiden, The Netherlands.
Wells, G. L., & Bradfield, A. L. (1998). "Good, you identified the suspect" Feedback to the witness distorts their reports of the witnessing experience. *Journal of Applied Psychology, 83,* 360–376.
Yarmey, D. A., Yarmey, M. J. , & Yarmey, A. L. (1996). Accuracy identifications in showups and lineups. *Law and Human Behavior, 20,* 459–477.
Yuille, J. C., (1993). We must study forensic eyewitnesses to know about them. *American Psychologist, 48,* 572–573.
Yuille, J. C., & Cutshall, J. L. (1986). A case study of eyewitness memory for a crime. *Journal of Applied Psychology, 71,* 291–301.

Psychology, Public Policy, and Law
2001, Vol. 7, No. 1, 3–35

Copyright 2001 by the American Psychological Association, Inc.
1076-8971/01/$5.00 DOI: 10.1037//1076-8971.7.1.3

THIRTY YEARS OF INVESTIGATING THE OWN-RACE BIAS IN MEMORY FOR FACES
A Meta-Analytic Review

Christian A. Meissner and John C. Brigham
Florida State University

The current article reviews the own-race bias (ORB) phenomenon in memory for human faces, the finding that own-race faces are better remembered when compared with memory for faces of another, less familiar race. Data were analyzed from 39 research articles, involving 91 independent samples and nearly 5,000 participants. Measures of hit and false alarm rates, and aggregate measures of discrimination accuracy and response criterion were examined, including an analysis of 8 study moderators. Several theoretical relationships were also assessed (i.e., the influence of racial attitudes and interracial contact). Overall, results indicated a "mirror effect" pattern in which own-race faces yielded a higher proportion of hits and a lower proportion of false alarms compared with other-race faces. Consistent with this effect, a significant ORB was also found in aggregate measures of discrimination accuracy and response criterion. The influence of perceptual learning and differentiation processes in the ORB are discussed, in addition to the practical implications of this phenomenon.

> She based her identification on Smith's eyes, which she said were greenish-blue and upon his hands which she said were "light and slender" like the holdup man's. Mrs. McCormick testified that Smith's eyes were "different from most colored people . . . bright and piercing." Smith's defense attorneys then attempted to parry the state's first thrust in the trial. Mrs. McCormick was handed a picture of a man she couldn't identify. It was a picture of David Charles, with shorter hair, taken while he was in Vietnam. Assistant defense attorney Kitchen asked Mrs. McCormick if she had ever made the statement that all Black people look alike. "Yes, I made that statement," Mrs. McCormick said, "and they do to a certain extent, but there's a difference here" (Lickson, 1974, p. 66).

In 1971, five Black men, who became known as the "Quincy Five," were wrongfully indicted for the murder of Khomas Revels during a robbery in Tallahassee, Florida. Although no forensic evidence obtained from the crime scene was ever linked to the men, five White eyewitnesses positively identified them as among the perpetrators. In each of three trials the state argued, "What better evidence can there be than, 'I saw him,' from unprejudiced witnesses? This has been used since time immemorial. This is proof beyond a reasonable doubt. Five eyewitnesses!" (Lickson, 1974, p. 87). Despite the lack of physical evidence against these men, two of the defendants, Dave Roby Keaton and Johnny Frederick, were found guilty on the basis of eyewitness testimony and coerced confessions obtained by investigators. During the third trial involving David

---

Christian A. Meissner and John C. Brigham, Department of Psychology, Florida State University.

Correspondence concerning this article should be addressed to Christian A. Meissner or to John C. Brigham, Department of Psychology, Florida State University, Tallahassee, Florida 32306-1270. Electronic mail may be sent to meissner@psy.fsu.edu.

Charles Smith, hired investigators on the defense team located the three actual perpetrators of the robbery and murder, who became known as the "Jacksonville Three." The Jacksonville men were later brought to trial and convicted based on latent fingerprint evidence and identification of the automobile used in the murder. The Quincy Five were finally exonerated.

At the trial of David Charles Smith, social psychologist Dr. William Haythorn of Florida State University (a colleague of John C. Brigham) was called as an expert witness in rebuttal of the eyewitness misidentifications. Because the only evidence against the Quincy Five was in the form of cross-racial identifications, Haythorn and Brigham set out to locate empirical evidence on the often purported claim that "they [other-race persons] all look alike." However, at the time of this case (c. 1971) only a handful of studies had examined the phenomenon (Berger, 1969; Horowitz & Horowitz, 1938), and only one study had been published in the previous decade (Malpass & Kravitz, 1969). Due in part to this lack of scientific evidence on cross-racial identification, the court prohibited the expert testimony of Haythorn.

Today, three decades later, a plethora of researchers have studied the own-race bias (ORB) in memory for human faces (also referred to as the cross-race effect or other-race effect). Although most now agree that the phenomenon is reliable across cultural and racial groups (Kassin, Ellsworth, & Smith, 1989), there is less consensus about the social and cognitive mechanisms that may govern the effect. Furthermore, little is known regarding variables that might moderate the effect, including those applicable to the eyewitness scenario, such as study time and retention interval. Thus, the goal of the current review and meta-analysis is not only to reconsider the reliability and generalizability of the ORB, but also to evaluate the validity of various theoretical mechanisms previously discussed in the literature and to propose a framework that might best account for the pattern of results across studies. Finally, we discuss the various practical implications of our findings for the legal and criminal justice systems.

Reliability of the ORB Effect

Literature reviews of the ORB have noted the robustness of the phenomenon (Brigham & Malpass, 1985; Chance & Goldstein, 1996), and researchers have endorsed the importance and reliability of the effect in several surveys (Kassin et al., 1989; Yarmey & Jones, 1983). Furthermore, expert witnesses have cited the effect in cases involving disputed cross-race identification (Brigham, Wasserman, & Meissner, 1999; Leippe, 1995), and attorneys have acknowledged the importance of racial interactions in eyewitness identifications (Brigham, 1981; Brigham & WolfsKeil, 1983). Given the source of such endorsements, one might be quick to concede the robust and generalizable nature of the ORB effect. However, it is important to further investigate the particular levels at which reliability might be assessed. For example, (a) Is the effect generally replicable across studies? (b) Is the effect consistent across various racial/ethnic groups? (c) Is the effect significant across different types of memory tasks? and (d) Is the effect reliable across individuals and testing occasions?

Replicability Across Studies

The first issue has been explored in several previous meta-analytic reviews of the effect. Bothwell, Brigham, and Malpass (1989) found that roughly 80% of the samples they reviewed demonstrated a significant ORB effect. Overall, effect size estimates from several previous meta-analyses (Anthony, Copper, & Mullen, 1992; Bothwell et al., 1989; Shapiro & Penrod, 1986) have indicated a significant weak-to-moderate effect, accounting for 6% to 11% of the variability across studies. R. C. Lindsay and Wells (1983) also examined the reliability of the effect across 13 studies by way of a vote-counting procedure. Although they asserted that fewer than half of the studies (6 of 13) demonstrated a true ORB effect, their criterion requiring a complete crossover interaction of White and Black participants may have been overly stringent. As Chance and Goldstein (1996) later noted, a large majority of the studies (11 of 13) reviewed by R. C. Lindsay and Wells showed at least some evidence of the effect.

Consistency Across Racial/Ethnic Groups

Several of these reviews have also examined the consistency of the ORB effect across racial/ethnic groups. Whereas Bothwell et al. (1989) found relatively equivalent estimates for both White and Black individuals, Anthony et al. (1992) found that the ORB effect among White participants accounted for 2.5 times the variance than that among Black participants. These inconsistencies could be due to the analysis of slightly different groups of studies. Moreover, both reviews relied on moderately small samples, in meta-analytic terms (number of independent samples: *k*s = 28 and 44; number of participants: *n*s = 1,445 and 1,725, respectively), increasing the likelihood of significant fluctuations in moderator effects due to the influence of one or more studies.

Generalizability Across Memory Tasks

Most studies documenting the ORB effect have used a standard recognition paradigm in which participants are tested on their ability to discriminate between a subset of faces shown previously (targets) and a subset of novel faces (distractors). Although a handful of studies have utilized some variant of this basic task (Cross, Cross, & Daly, 1971; D. S. Lindsay, Jack, & Christian, 1991; Luce, 1974; Malpass, 1974), some reviewers, such as R. C. Lindsay and Wells (1983), have criticized the literature for not examining performance on other memory tasks, including more applied identification tasks. More recently, however, researchers have responded to this criticism by documenting the effect across a variety of paradigms, including matching tasks (Malpass, Erskine, & Vaughn, 1988) and lineup identification paradigms (Berger, 1969; Brigham, Maass, Snyder, & Spaulding, 1982; Doty, 1998; Fallshore & Schooler, 1995; Platz & Hosch, 1988). In addition, researchers have shown the presence of the effect across other measures of performance such as reaction time (Chance & Goldstein, 1987; Valentine, 1991) and other tasks of forensic relevance including facial reconstruction tasks (Ellis, Davies, & McMurran, 1979) and photo lineup construction by law enforcement officers (Brigham & Ready, 1985).

Reliability Across Individuals and Testing Occasions

A fourth level of reliability, namely consistency of the ORB effect across individual participants, has only recently been examined. In general, memory for human faces has been shown to demonstrate reliable properties when assessed by tools designed to investigate cognitive maturation and/or neurological impairment. For example, Malina, Bowers, Millis, and Uekert (1998) found that the Faces subtest of the Recognition Memory Test (Warrington, 1984) had sufficient internal consistency and reliability (Cronbach's $\alpha = .77$) for clinical use (see also Soukop, Bimbela, & Schiess, 1999). Similarly, the Benton Facial Recognition Test (Benton, Hamsher, Varney, & Spreen, 1983), the Faces subtest of the Wechsler Memory Scale—III (1997), and the Face Recognition subtest of the Kaufman Assessment Battery for Children (Kamphaus, Beres, Kaufman, & Kaufman, 1996) have all produced sizeable reliability estimates ($rs > .75$). Interestingly, and pertinent to the current investigation, several laboratory efforts at demonstrating reliability in a face-recognition task have yielded only moderate reliability estimates (Chance & Goldstein, 1979; Goldstein & Chance, 1980; Malpass et al., 1998; Prospero, Corey, Malpass, Parada, & Schreiber, 1996). Although researchers had taken care to randomly assign faces to recognition sets in controlling for item effects (see Chance & Goldstein, 1979), more deliberate standardization in controlling the memorability of materials and test sets may provide for better estimates of facial memory reliability in future studies.

Although it has largely been assumed that the ORB effect would follow a similar pattern of reliability across testing occasions (namely, moderate-to-large reliability estimates), little research has been available to test this assumption. In a recent study, we (Slone, Brigham, & Meissner, 2000) sought to test the reliability of the ORB effect across an immediate and (2-day) delayed testing occasion. Our results indicated that although participants performed reliably on both own-race and other-race faces, rs (127) = .56 and .44, $ps < .001$, respectively, the magnitude of the difference between own-race and other-race performance (i.e., the ORB) was only somewhat reliable across the delay, $r(127) = .21$, $p < .05$). Malpass et al. (1998) also recently investigated the reliability of other-race face recognition across two separate testing occasions. Although they found no reliability in performance on other-race faces, $r(11) = .08$, *ns,* this may have been due, in part, to the small sample of participants ($n = 13$). Their estimate of reliability for same-race recognition was significant, but of moderate size, $r(59) = .36$, $p < .01$. Although left unaddressed by the current meta-analysis, the issue of test–retest reliability in the ORB merits further investigation. Once again, greater care in the standardization of materials across race of face may provide more reasonable estimates of reliability in future studies.

The Search for Social–Cognitive Mechanisms

Thus far, theoretical notions for the ORB have spanned the realms of both social and cognitive mechanisms. Whereas early candidates included the effect of social attitudes and the notion of physiognomic differences between races, more recent hypotheses have involved the potential influence of interracial contact and the notion of a perceptual learning mechanism. Unfortunately, inconsistency has often plagued the literature seeking to verify each theory. Because previous

reviews of the ORB effect have given much attention to the various theoretical positions (Brigham & Malpass, 1985; Chance & Goldstein, 1996; Shepherd, 1981), we provide only a cursory updated description of each approach.

Racial Attitudes

One initial explanation for the ORB effect was that individuals with less prejudiced racial attitudes would be more motivated to differentiate other-race members, when compared with more prejudiced persons. Early research indicated that racial attitudes appeared to influence the degree of stereotypic likeness assigned to other-race members (Secord, Bevan, & Katz, 1956). In addition, early studies examining participants' performance on identification of race/ethnicity (e.g., Jewish vs. non-Jewish) demonstrated that more-prejudiced individuals often performed better than less-prejudiced individuals (Allport & Kramer, 1946; Lindzey & Rogolsky, 1950). However, other studies were not always supportive of the findings (Carter, 1948), and subsequent researchers noted that high-prejudiced performance was likely influenced by a response bias to label more faces as out-group members (Elliott & Wittenberg, 1955).

Within the ORB literature, several early studies demonstrated a small relationship between attitudes toward other-race persons and recognition memory performance (Berger, 1969; Galper, 1973). However, when response bias was taken into account, Dowdle and Settler (cited in Yarmey, 1979) found that racial attitudes were unrelated to memory performance. Similarly, more recent studies have consistently failed to find a relationship between racial attitudes and memory for other-race faces (Brigham & Barkowitz, 1978; Lavrakas, Buri, & Mayzner, 1976; Platz & Hosch, 1988; Slone et al., 2000; Swope, 1994). However, racial attitudes are related to another factor thought relevant to recognition of other-race faces, namely, amount of interracial contact. A number of studies have found that those with more prejudiced attitudes report less contact with other-race members (Brigham, 1993; Brigham & Barkowitz, 1978; Brigham & Meissner, 2000; Brigham & Ready, 1985; Slone et al., 2000; Swope, 1994).

Physiognomic Homogeneity

A second possibility for the ORB effect involves possible group differences in the inherent memorability of faces, such that faces of some races might show less physiognomic variability among group members when compared with other races. However, researchers examining this hypothesis have generally found little support for its validity. For example, Goldstein (1979) found no differences in physiognomic variability among Japanese, Black, and White faces. Additionally, several studies have demonstrated that latency and accuracy of same–different judgments do not differ across race of participant or race of face (Goldstein & Chance, 1976, 1978). Finally, within-race rated similarity has shown, at best, only an inconsistent relationship to perception by own-race and other-race individuals, leading Goldstein and Chance (1979) to conclude that, overall, there is little "compelling evidence for the homogeneity hypothesis" (p. 111). We should note that although physiognomic homogeneity may not be responsible for the ORB memory effect, a number of studies have indicated that different physiognomic facial features may be more appropriate for discriminating between faces of

certain races (Ellis, Deregowski, & Shepherd, 1975; Shepherd, 1981; Shepherd & Deregowski, 1981).

Interracial Contact

A number of researchers have posited that the quality or quantity of interracial contact may play a vital role in the degree of ORB demonstrated by any particular individual. For example, researchers have proposed that increased contact with other-race individuals may increase memory performance by (a) reducing the likelihood of stereotypic responses and increasing the likelihood that individuals may look for more individuating information (Malpass, 1981; Shepherd, 1981), (b) influencing individuals' motivation to accurately recognize other-race persons through associated social rewards and punishments (Malpass, 1990), or (c) reducing the perceived complexity of unfamiliar other-race faces (Goldstein & Chance, 1971). Two major approaches to investigating contact are to examine groups of individuals differing in their degree of other-race contact or to assess individuals' self-reported contact with other-race persons.

With regard to the former approach, several early studies demonstrated that adolescents and children living in integrated neighborhoods better recognized novel other-race faces than did those living in segregated neighborhoods (Cross et al., 1971; Feinman & Entwisle, 1976). Other more recent studies have also shown evidence of the influence of contact in samples of White and Black individuals from Great Britain and Africa (Carroo, 1986; Chiroro & Valentine, 1995; Wright, Boyd, & Tredoux, 1999). Finally, a novel application of the contact hypothesis was recently conducted by Li, Dunning, and Malpass (1998) who demonstrated that White "basketball fans" were superior to White "basketball novices" in recognizing Black faces. Given that the majority of professional basketball players are Black, this effect was predicted on the basis of the fans' experience in differentiating individual players. It is interesting to note that not all studies have found the predicted relationship between high-contact and low-contact groups. Burgess (1997) found only a small effect of contact on the performance of Southern (Florida) and Northern (Maine) American samples of White individuals. Similarly, Ng and Lindsay (1994) found little support for the influence of contact on the performance of Canadian and Singapore samples.

In a number of other studies, researchers have assessed the relationship between memory for other-race faces and individuals' self-reported experience with other-race persons. Whereas early studies generally failed to find a significant relationship (Berger, 1969; Brigham & Barkowitz, 1978; Cross et al., 1971; Malpass & Kravitz, 1969), numerous studies over the past several decades have found at least some evidence of the relationship in both recognition tasks (Byatt & Rhodes, 1998; Carroo, 1986, 1987; Lavrakas et al., 1976; Li et al., 1998; D. S. Lindsay et al., 1991; Slone et al., 2000; Swope, 1994; Wright et al., 1999) and more applied lineup identification paradigms (Brigham et al., 1982; Platz & Hosch, 1988). This curious pattern of results over time will be further examined in the current meta-analysis. It is possible that the precision and validity of measures used to assess interracial contact have improved over the years. Alternatively, as Chance and Goldstein (1996) posited, a cohort effect may exist such that opportunities for interracial contact have increased following the desegrega-

tion and civil rights movements of the 1960s and 1970s, allowing for a greater range in the degree of interracial contact in recent years.

Perceptual Learning

As reviewed in the previous section, a fair degree of empirical support exists for the notion that interracial contact has some influence on the magnitude of the ORB. However, researchers are still attempting to elucidate the specific cognitive mechanisms through which contact might actuate this influence, and to model their effects in more formal ways. The most popular general approach is likely that of *perceptual learning.* As historically defined by Gibson (1969), perceptual learning involves "an increase in the ability to extract information from the environment, as a result of practice and experience with stimulation coming from it" (p. 3). Numerous reviews have been written concerning the various mechanisms likely to underlie the phenomenon (Ahissar & Hochstein, 1998; Proctor & Dutta, 1995; Walk, 1978), and most note the important role of Gibson's notion of *differentiation,* defined as focused attention directed toward invariant cues that provide the best bases for discriminations within a given stimulus set. More recent work by Haider and Frensch (1996, 1999) has furthered Gibson's notion by demonstrating that perceptual skill involves learning to distinguish between "task-relevant" and "task-redundant" information. Thus, increases in accuracy and speed of processing appear to reflect the extent to which individuals have knowledge of, and provide attention to, the appropriate (invariant) features of the stimulus.

Such an encoding-based effect has been documented in a variety of perceptual skill domains, including chess (Reingold, Charness, Pomplun, & Stampe, in press), bird watching (K. E. Johnson & Mervis, 1997, 1998), sports (Helsen & Pauwels, 1993; Shea & Paull, 1996), radiology (Christensen et al., 1981; Lesgold et al., 1988; Myles-Worsley, Johnston, & Simons, 1988), and even chicken sexing (Biederman & Shiffrar, 1987). It is possible that perceptual learning might also be responsible for the ORB phenomenon. For example, individuals may be able to discriminate own-race faces more accurately due to their use of appropriate (invariant) aspects of the face. On the other hand, cues used for own-race faces may not be appropriate when attempting to remember other-race faces, and thus performance would worsen when attempting to discriminate such unfamiliar stimuli. A handful of studies have investigated this notion of perceptual learning from a discrimination training perspective. Other research within this general framework has attempted to identify various aspects of the face that might be deemed "task-relevant" when recognizing own-race versus other-race faces and to provide evidence in support of more formal models of the ORB.

Discrimination training. Some researchers in the face memory domain have directly investigated the perceptual learning hypothesis by providing individuals with discrimination training on own-race and other-race faces. Although training seems to have no effect on improving own-race recognition (Malpass, 1981), there is some evidence that training may reduce the ORB, at least in the short run. For example, Malpass, Lavigueur, and Weldon (1973) attempted to improve recognition memory for own-race and other-race faces by either verbal or visual training tasks. Although verbal training showed no effect on recognition, a

relatively short visual training task (1 hr) produced a significant reduction in the magnitude of the ORB. Lavrakas et al. (1976) also investigated the effects of training by presenting participants with a concept learning task. Post-training recognition performance demonstrated significant improvement on other-race faces for individuals in the concept learning conditions compared with the unchanged performance of individuals in a control condition. However, when participants in all conditions were tested again 1 week later, the performance of trained and untrained participants on other-race faces was no different. Finally, E. S. Elliott, Wills, and Goldstein (1973) investigated the influence of paired associate discrimination training in reducing the magnitude of the ORB. Whereas participants in the no-training and own-race training conditions displayed the typical ORB effect, those in the other-race training condition demonstrated significant improvement in recognition accuracy for other-race faces.

Configural–featural hypothesis. Although relatively short-lived effects of discrimination training have been found, other researchers have sought to identify the various cognitive processes that might differentiate own-race and other-race face recognition. One notable advance in the face memory literature has involved work on the *face inversion effect,* the finding that inverted (upside-down) photos of faces are identified more poorly than inverted photos of other objects. In early work on this effect, Yin (1969) concluded that face recognition was the product of a unique system, different from systems responsible for recognizing other kinds of visual stimuli. In contrast to this "neural specialization" hypothesis, Diamond and Carey (1986) proposed that perceptual learning might be operating in face recognition. In several experiments they showed that the inversion effect was not unique to faces, but rather occurred when participants had a great deal of experience with the stimulus materials. Inversion appeared to disrupt the effectiveness with which individuals were able to encode stimuli that were highly familiar to them. This, they claimed, stemmed from experienced participants' reliance on configural (or relational) properties of the stimulus. Novice participants, on the other hand, relied on only the featural (or isolated) aspects of the face that were less influenced by inversion. A number of subsequent studies have supported this general configural–featural hypothesis (see Farah, Wilson, Drain, & Tanaka, 1998).

The notion of expertise and configural processing has also been applied to the ORB effect. In particular, Rhodes, Brake, Taylor, and Tan (1989) proposed that greater experience with own-race faces would lead to a larger inversion effect, due to an increased reliance on configural information. The encoding of other-race faces, on the other hand, should not be as influenced by inversion due to the featural aspects that are relied on. As hypothesized, Rhodes et al. observed that own-race faces were significantly more susceptible to inversion than other-race faces for measures of both reaction time and accuracy. However, several other studies have observed either no interaction of inversion with the ORB (Buckhout & Regan, 1988) or larger inversion effects on other-race faces (Valentine & Bruce, 1986). Given the various methodological differences across studies, further empirical and theoretical work on the significance of inversion effects in the ORB would be valuable.

Finally, Fallshore and Schooler (1995) examined whether such perceptual expertise might also be involved in the *verbal overshadowing effect,* the finding

that generating a verbal description of a face significantly impairs subsequent identification accuracy (see Meissner & Brigham, in press, for a meta-analytic review). Specifically, they hypothesized that requesting participants to provide a description of a same-race face might cause significant declines in recognition performance by (a) forcing participants to rely on the featural (more verbalizable) aspects of the face and (b) disrupting the configural (less verbalizable) memory trace that was originally encoded. Performance in cross-race identification, however, was predicted not to show the overshadowing effect due to individuals' reliance on featural aspects when encoding other-race faces. Consistent with their hypotheses, Fallshore and Schooler found that although participants' recognition performance on same-race faces demonstrated the overshadowing effect (a 47% decrement in performance when verbal descriptions were given), other-race faces showed no such decline in performance.

"Face space" models. Although the configural–featural hypothesis has received much attention, other researchers have examined the particular manner in which faces might be represented in memory. Likely the most ambitious work involves that of Valentine and his colleagues (Valentine, 1991; Valentine & Bruce, 1986; Valentine & Endo, 1992) in the development of an exemplar-based model of facial memory. Although Valentine and colleagues conceded the notion of a configural–featural distinction in the type of facial features that individuals may encode, they disputed Diamond and Carey's (1986) proposal that a fundamental change in the underlying processing strategy occurs under inversion (Valentine, 1988; Valentine & Bruce, 1988). Rather, Valentine (1988) proposed that, in conjunction with the notion of *schema theory* pioneered by Goldstein and Chance (1980), an exemplar-model reflecting "the acquisition of knowledge of how faces vary" may account for the effects of inversion, race, and distinctiveness (Valentine, 1988, p. 485).

Generally speaking, Valentine's (1991) multi-dimensional space (MDS) framework holds that the representational system may be thought of as a hypothetical space in which faces are stored based on various dimensions representing features or sets of features. The model posits that these dimensions are based on an individual's prior experience with the stimulus set and thus are best suited for representation of own-race faces, due to a reliance on appropriate featural and/or configural information. As a result of this encoding, own-race faces are spread more evenly throughout the MDS and are better individuated from one another at retrieval. Conversely, other-race faces are poorly represented (and, thus, more tightly clustered in the MDS) due to the encoding of less appropriate featural and/or configural information. Valentine's (1991) model also posits, however, that with increasing experience, other-race faces may be better represented once the relevant (invariant) aspects of other-race faces are learned.

In a test of the MDS framework, Chiroro and Valentine (1995) examined the effects of race, typicality, and level of perceptual experience within the cross-race paradigm. Although the influence of rated distinctiveness on recognition of own-race faces had been widely known (Brigham, 1990; Hosie & Milne, 1995), the manner in which it might interact with race and perceptual experience had not been investigated. Based on the assumptions of the MDS model, Chiroro and Valentine predicted that only individuals who had considerable previous experience with other-race faces (high-contact) would demonstrate distinctiveness ef-

fects for both own-race and other-race faces. This was due largely to the notion that such individuals should be able to distinguish between typical and distinctive other-race faces based on features they had extracted through prior experience. In contrast, low-contact individuals were predicted to demonstrate no differences in performance on the distinctiveness dimensions of other-race faces. Overall, their results indicated the predicted four-way interaction such that distinctiveness effects for low-contact individuals were confined to own-race faces. On the other hand, high-contact individuals demonstrated significant effects of distinctiveness regardless of the race of the face.

Race-feature hypothesis. An alternative to Valentine's (1991) MDS model was proposed by Levin (1996) in explaining the paradoxical effect that individuals are slower at classifying the race of an own-race face compared with that of an other-race face. This other-race classification advantage (ORCA) was observed by Valentine and Endo (1992) and was explained as resulting from strong activation due to the high-density cluster of other-race faces in the representational system (MDS). Levin (1996) proposed an alternative to this explanation in which the ORCA was said to arise from a "facilitated classification process" (p. 1366). In particular, Levin suggested that other-race faces were more quickly classified due to an automated process in which race-specific coding is performed without regard for other individuating information, which is largely ignored.

In testing this race-feature hypothesis, Levin (1996) observed that participants demonstrating a large ORB in recognition memory also demonstrated a large ORCA when compared with other individuals (see also Levin & Lacruz, 1999). Given that the ORB observed was driven largely by false alarm responses to other-race faces, Levin argued that participants' coding of race alone was insufficient to discriminate between other-race faces, leading to a tendency to respond "seen before" during test. Levin further proposed that individuals having greater experience with other-race persons would be less likely to generate the race-feature response, but instead would initially seek out individuating information for later use. Although he did not test this possibility, Levin's observation is analogous to that of skill differences in the "basic level" categorization effect (K. E. Johnson & Mervis, 1997, 1998; Tanaka & Taylor, 1991). Namely, whereas novices respond to stimuli most quickly based on a basic level categorization (e.g., bird), experts respond just as quickly at the basic, subordinate (e.g., wren), and even sub-subordinate levels (e.g., Carolina wren). Thus, experts' conceptual knowledge of domain-relevant features appears to allow them faster access to multiple levels of identification. Similarly, individuals with more experience with other-race faces may have faster access to identity information by way of their conceptual knowledge of individuating features.

Taken together, a perceptual learning approach to understanding the ORB has considerable potential for explaining its cognitive origins. The focus on encoding-based processes within the configural–featural and race-feature hypotheses may stimulate future empirical and theoretical progress. In addition, the representational model put forth by Valentine and colleagues (Valentine, 1991; Valentine & Endo, 1992) has provided a testable framework within which both general and effect-specific approaches to memory for faces may interact. The current meta-analysis was designed to aid researchers in further exploring perceptual learning

aspects of the ORB by providing aggregate estimates of the effect across several performance measures.

Meta-Analysis

The present review of the ORB paradigm has yielded many testable hypotheses concerning both the general reliability of the effect and the various mechanisms posited for its occurrence. Our meta-analysis took the approach advocated by Hedges and Olkin (1985) in which a mean weighted effect size for the sample of studies was initially calculated, followed by prediction of effect size based on moderating variables (see B. T. Johnson, Mullen, & Salas, 1995, for a discussion of various approaches). In particular, we were interested in examining ORB effect size estimates for basic measures of hits (correctly identifying a face as "old") and false alarms (incorrectly identifying a face as "old"), as well as aggregate signal detection estimates of discrimination accuracy (the standardized distance between the means of the "new" and "old" distributions) and response criterion (the level of familiarity necessary for an individual to categorize a given stimulus as "old" vs. "new"; for a review of signal detection theory, see Green & Swets, 1966). Second, in testing the validity of several theoretical mechanisms posited in the literature, we also provide estimates of the influence of racial attitudes and self-rated interracial contact on other-race memory performance, as well as an estimate of the correlation between attitudes and contact as measured across studies. Finally, in addition to overall effect size analyses, eight moderating variables (described below) are examined across the four performance measures.

Method

Studies

A total of 91 independent effect sizes described in 39 research articles were located, representing the responses of 4,996 participants. Of the 39 research articles, 6 (15%) were unpublished manuscripts or theses/dissertations. Studies were obtained using several methods, including (a) searches of *PsycINFO, Sociofile,* and *Dissertation Abstracts* databases and using the key words "face memory," "face recognition," and "face identification" along with the key words "race" and "ethnicity"; (b) cross-referencing with the three previous meta-analyses (Anthony et al., 1992; Bothwell et al., 1989; Shapiro & Penrod, 1986) and various reviews on the effect (Brigham & Malpass, 1985; Chance & Goldstein, 1996; R. C. Lindsay & Wells, 1983); and (c) contact with colleagues in the field who may have had knowledge of fugitive literature that had neither been published nor presented at a conference.

Inclusion–Exclusion Criteria

To be included in the analysis, studies must have involved a within-subjects test of participants' memory for own-race and other-race faces. The statistical difference in performance on these two sets of stimuli for each participant is defined as the ORB. Note that, in contrast to several previous meta-analyses (Anthony et al., 1992; Bothwell et al., 1989), studies that involved only a single race of participants were included in addition to studies that involved races other than Whites and Blacks. Reasons for excluding studies involved (a) the lack of sufficient data from which to compute an effect size (Bruce, Beard, & Tedford, 1997; Caroo, 1988; Horowitz & Horowitz, 1938; Luce, 1974; Malpass,

1988), (b) the use of a between-subjects design and analysis (Caroo, 1986; E. S. Elliot, Wills, & Goldstein, 1973), or (c) the implementation of various methodological procedures that might obscure interpretation of the effect size estimate, such as unequal presentation rates for own-race and other-race faces (Byatt & Rhodes, 1998; Doty, 1998; Goldstein & Chance, 1985; Lavrakas et al., 1976; Padgett, 1997; Valentine & Bruce, 1986).

Coded Variables

Based on the suggestions of Lipsey (1994), moderator variables were selected by way of three general categories of study descriptors.[1] First, we examined variables that were of substantive experimental and applied interest in characterizing the reliability and generality of the ORB effect, including the race of the participant and the type of memory task used. Fifty-six percent of the samples were reported as White, and 32% were reported as Black. The remaining 12% of samples included individuals of Arab/Turkish, Asian, and Hispanic origin. The majority (91%) of studies used a recognition paradigm, whereas 9% of studies used a (simultaneous and target-present) lineup identification task. Briefly, recognition paradigms involve presenting participants with a set of faces that they must later recognize from a group of "old" and "new" faces. Identification paradigms are generally more applicable to the eyewitness situation and involve presenting participants with a single face (either from a photograph or a short video) that they must later identify from a group (or photo lineup) of 6–8 similar faces.

Second, we assessed methodological or procedural aspects of studies such that we might identify possible sources of distortion involving boundary conditions under which the ORB might be observed. Such variables included (a) whether test stimuli were identical (72%) or different (28%) from those used at study, (b) whether races of face were presented and tested in a blocked (19%) or mixed (81%) fashion, (c) the amount of time participants were permitted to study individual faces (minimum = 0.12 s; maximum = 4 min; median = 3 s), and (d) the length of the retention interval between study and test phases (minimum = immediate; maximum = 3 weeks; median = 2 min).

Finally, we also considered other extrinsic study characteristics, including the date of publication or presentation and whether the effect size estimate was taken from a published or unpublished manuscript. Of the studies included for analysis, 27% were published in the 1970s, 33% in the 1980s, and 40% in the 1990s. Fifteen percent of these studies were unpublished and took the form of a conference presentation or a thesis/dissertation.

Measure of Effect Size

Our measure of effect size for the performance variables (i.e., hits, false alarms, and discrimination accuracy) was a single sample estimate equivalent to Hedge's g^U. This effect size was computed simply as the mean difference between own-race and other-race performance divided by the sample standard deviation, or

$$g = (\mu_{own} - \mu_{other})/S_D \quad (1)$$

To control for skewness in estimating the true population parameter, g was transformed to g^U by way of Equation 2:

[1]To assess the reliability of coding study moderator variables, two raters generated independent codings for each variable across studies. Rate of agreement across all variables ranged between 93% and 100%.

$$g^{U} = c(m)*g, \quad (2)$$

where

$$c(m) = 1 - (3/[(4*df) - 1]). \quad (3)$$

To assess the influence of both attitudes and contact on the ORB, as well as the correlation between the two measures across studies, r coefficients were recorded for each independent sample, after which r was transformed to Fisher's Z_r by way of Equation 4:

$$Z_r = .5*\log_e[(1 + r)/(1 - r)]. \quad (4)$$

All formulae were obtained from Rosenthal (1994). Effect sizes demonstrating the ORB will be positive for measures of hits, discrimination accuracy, and response criterion, and negative for the measure of false alarms. Likewise, positive estimates for the racial attitude and interracial contact measures indicate that positive attitude toward and increased contact with other-race individuals leads to better performance on other-race faces.

Results

Weighted Effect Size Analyses

To examine the pattern of effect sizes for each measure, estimates were weighted as a function of their independent sample sizes, after which the results were analyzed across studies. For each measure, the mean weighted effect size (g^{U}) is presented, in addition to a test of the significance of the estimate (Z), and the associated 95% confidence intervals.

Hits and false alarms. The mean weighted effect size for the proportion of hit responses across studies ($k = 74$) demonstrated a significant ORB, $g^{U} = .24$, $Z = 15.43$, $p < .001$, with 95% confidence intervals of .21 and .27. In practical terms, an odds-ratio analysis indicated that participants were 1.4 times more likely to correctly identify a previously viewed own-race face when compared with performance on other-race faces. For false alarm responses, the mean weighted effect size across studies ($k = 53$) also indicated a significant ORB, $g^{U} = -.39$, $Z = 22.24$, $p < .001$, with 95% confidence intervals of $-.42$ and $-.35$. Participants were 1.56 times more likely to falsely identify a novel other-race face when compared with performance on own-race faces.

Taken together, these results illustrate a "mirror effect" pattern in which other-race faces receive a lower proportion of hits and a higher proportion of false alarms when compared with own-race faces (Figure 1). The mirror effect has been termed a "regularity" of recognition memory and has been demonstrated for such variables as frequency, distinctiveness, and study time (see Glanzer & Adams, 1985, 1990). Although the theoretical mechanisms of this effect are often debated between models (Glanzer & Adams, 1990; Hintzman, 1988; Hirshman, 1995; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997), many studies have shown that the aggregate measure of discrimination accuracy is generally influenced when mirror effects are observed. Other researchers have noted changes in response criterion estimates as well; however, substantial differences in discrimination accuracy between stimuli must be present for the criterion effect to be observed (Hirshman, 1995; McClelland & Chappell, 1998). Hence, we were

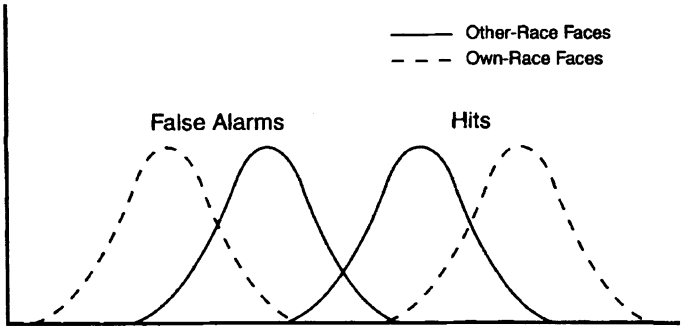

Figure 1. "Mirror-effect" pattern demonstrated in hit and false alarm responses to own-race and other-race faces.

interested to see whether ORB differences would occur only on estimates of discrimination accuracy, or on estimates of response criterion as well.

Discrimination accuracy. The mean weighted effect size for the measures of discrimination accuracy across studies ($k = 56$) was $g^U = .82$, a significant ORB, $Z = 42.32$, $p < .001$, with 95% confidence intervals of .78 and .85. Overall, the ORB in discrimination accuracy accounted for 15% of the variability across studies, and participants were 2.23 times more likely to accurately discriminate an own-race face as new versus old when compared with performance on other-race faces.

Response criterion. Unfortunately, only six studies ($k = 14$) actually calculated a response criterion measure across participants. Of the 14 independent samples, 11 demonstrated a significant ORB effect ($\alpha = .05$) such that other-race faces yielded a more liberal criterion when compared with performance on own-race faces. The remaining 3 samples demonstrated nonsignificant patterns. To further assess this effect, a studywise response criterion analysis was conducted in which the mean hit and false alarm rates for each study were used to calculate a response criterion estimate (see Macmillan & Creelman, 1990). The mean weighted effect size for the estimates of response criterion across studies ($k = 49$) was $g^U = .30$, a significant ORB, $Z = 17.91$, $p < .001$, with 95% confidence intervals of .26 and .33. Overall, this small effect of response criterion in the ORB accounted for only 1% of the variability across studies and indicated that own-race faces generally yielded a more conservative criterion when compared with performance on other-race faces.

In summary, the pattern of results for discrimination accuracy measures was consistent with the mirror effect pattern that was observed in the hit and false alarm responses. Given the significant size of the discrimination accuracy effect, the presence of a response criterion effect in the ORB was expected (Hirshman, 1995). A recent model of recognition memory proposed by McClelland and Chappell (1998) provided an account of this pattern of results by simulating the process of differentiation (Gibson, 1969). As discussed previously, differentiation has been implicated in the various perceptual learning approaches to explaining the ORB. In the Discussion section, we consider the merits of McClelland and Chappell's model and its theoretical implications for the ORB.

Racial attitudes. Researchers have long posited that attitudes toward other-race persons may be responsible for the ORB in face memory. However, as noted, empirical results have not generally supported this notion. To assess the validity of this hypothesis, we examined the pattern of correlations between racial attitudes and performance on other-race faces across studies ($k = 14$). The mean weighted effect size across studies indicated no significant relationship, $Z_r = -.01$, $Z = .25$, with 95% confidence intervals of $-.08$ and .06. Hence, there appears to be no evidence of a direct influence of racial attitudes on the ORB.

Interracial contact. Researchers have also posited that interracial contact should influence the degree of ORB demonstrated by any given individual. To assess this relationship across studies, we examined the pattern of correlations between self-rated interracial contact and discrimination of other-race faces ($k = 29$). The mean weighted effect size across studies demonstrated a significant relationship, $Z_r = .13$, $Z = 5.34$, $p < .001$, with 95% confidence intervals of .08 and .18. Overall, contact appears to play a small, yet reliable, mediating role in the ORB, accounting for approximately 2% of the variability across participants. This seemingly weak relationship between self-rated contact and the ORB may be due to limitations in the range of variability present in such measures. Future studies may wish to further explore alternative methods of assessing interracial contact.

Attitude–contact relationship. As noted previously, we have found evidence of a relationship between attitudes toward other-race persons and self-rated contact in our lab. It is conceivable that although individuals' attitudes have no direct influence on their memory for other-race faces, racial attitudes may yet play a mediating role by way of their relation to individuals' social experience with other-race persons. The mean weighted effect size between interracial attitudes and contact across studies ($k = 10$) demonstrated a significant relationship, $Z_r = .36$, $Z = 11.42$, $p < .001$, with 95% confidence intervals of .30 and .42. In general, individuals with more positive attitudes toward other-race persons tend to rate themselves as experiencing more interracial contact when compared with individuals with more negative attitudes.

Moderator Effects

A test of the homogeneity of variances across the sample of weighted effect sizes (hit, false alarm, discrimination accuracy, and response criterion measures) indicated a significant degree of variability, exceeding that expected on the basis of sampling error alone, Qs $> 1{,}000$, ps $< .001$. Thus, the design moderators discussed earlier were used to predict the variability across the sample of effect sizes. A weighted least-squares regression analysis (Hedges, 1994) was conducted for each measure across the three sets of moderator variables (i.e., reliability and generalizability, methodological characteristics, and extrinsic study factors). Effect sizes in the analysis were weighted as a function of their sample size. Due to the sensitivity of this fixed-effects analysis, we took a more conservative approach

and discuss only those moderator effects with $Z_j \geq 3.30$ or $\alpha = .001$.[2] Significant effects resulting from this criterion yielded semipartial correlations (r_s) ranging in magnitude from .11 to .33. Table 1 provides a summary of moderator effects (Z_j) across the four performance measures.

Reliability and generalizability. The first set of moderators assessed whether the ORB was reliable across racial/ethnic groups and whether the effect was generalizable to the type of memory task. Similar to that of Anthony et al. (1992), results indicated that White participants demonstrated a significantly larger ORB when compared with Black participants with regard to the measure of discrimination accuracy, $Z_j = 6.91, p < .001$. This effect appeared to stem largely from differences in the magnitude of false alarm responses, $Z_j = 9.50, p < .001$. However, Whites and Blacks did not differ in the magnitude of the ORB on either proportion of hits or estimates of response criterion, Z_js $\leq .79$. White participants also demonstrated a significantly larger ORB when compared with participants grouped in the "other" racial/ethnic category. This effect was observed reliably in hit, false alarm, and response criterion estimates, Z_js ≥ 8.14, ps $< .001$. However, the analysis of discrimination accuracy was not significant, $Z_j = 1.13$. Mean weighted effect sizes for each racial/ethnic group across the four performance measures are displayed in Table 2.

Analysis of the effect sizes found in recognition versus lineup identification paradigms yielded no significant difference with regard to the measure of false alarm responses, $Z_j = 1.55$. However, there was a tendency for studies using an identification paradigm ($g^U = .45$) to yield a larger ORB for proportion of hits when compared with studies using a recognition paradigm ($g^U = .22$), $Z_j = 2.76$, $p < .01$. Nevertheless, it is evident that the ORB effect is generalizable to both recognition and lineup identification tasks. As only a small proportion (9%) of the samples involved the use of an identification task, future studies utilizing the lineup paradigm would be valuable.

Methodological characteristics. The second set of moderators examined various methodological aspects that might influence the magnitude of effects observed across studies. First, studies were coded for whether they utilized the identical or different facial photographs at study and test and for whether the presentation of stimuli was mixed or blocked by race/ethnicity. Results indicated that the type of stimulus (i.e., identical vs. different) significantly influenced estimates of the ORB on the proportion of hits and estimates of response criterion, Z_js ≥ 3.42, ps $< .001$. This effect of stimulus type was also apparent in the proportion of false alarms, $Z_j = 3.27, p < .01$, though not at the $\alpha = .001$ level.

---

[2]This conservative criterion ($\alpha = .001$) for study moderators was chosen due to the sensitivity of the "fixed effects" analysis. Given the exploratory nature of our investigation, we felt that such a criterion might allow us to examine a range of variables that would likely be replicable under direct empirical investigation. A more conservative, "random effects" model was also run on the sample of studies (see Raudenbush, 1994). Results indicated that White participants yielded a significantly larger own-race bias (ORB) on false alarm responses when compared with both Black, $Z = 2.50$, $p < .05$, and other racial/ethnic participants, $Z = 2.45$, $p < .05$. White and other participants also exhibited a significant difference in the response criterion estimates, $Z = 2.13$, $p < .05$. Additionally, limiting the amount of study time significantly increased estimates of the ORB on aggregate measures of discrimination accuracy, $Z = -2.19$, $p < .05$. No other moderator effects were found to be significant, Zs ≤ 1.36.

Table 1
Influence of Moderator Variables (Z_j) Across Measures of Hits, False Alarms, Discrimination Accuracy, and Response Criterion

Moderator	Hits (k = 74)	False alarms (k = 53)	Discrimination accuracy (k = 56)	Response criterion (k = 49)
Race/ethnicity of participant				
White vs. Black	0.79	9.50***	6.91***	0.12
White vs. Other	8.14***	10.36***	1.13	11.55***
Task				
Identification vs. recognition	2.76	1.55	—	—
Stimulus				
Photo vs. face	3.42***	3.27	0.16	7.34***
Order of study				
Blocked vs. mixed	4.47***	1.13	5.49***	0.81
Study time (in seconds)	0.60	2.65	11.70***	1.19
Retention interval (in minutes)	0.45	0.86	1.55	7.17***
Year of study	2.50	5.46***	10.50***	0.50
Status of study				
Published vs. unpublished	0.69	0.63	1.74	1.59

***$p < .001$.

In addition, presentation of stimuli (i.e., blocked vs. mixed) significantly influenced ORB estimates of discrimination accuracy, $Z_j = 4.47$, $p < .001$, largely as a function of the proportion of hits, $Z_j = 5.49$, $p < .001$. As displayed in Table 3, the pattern of weighted means demonstrated that significantly larger ORB effects were observed if facial photographs were altered from study to test and if the presentation of faces was blocked by race/ethnicity.

Studies were also coded for the length of time participants studied each target face (in seconds), and the length of the retention interval between study and test phases of the experiment (in minutes). Results indicated that the amount of study time influenced estimates of the ORB on measures of discrimination accuracy, $Z_j = 11.70$, $p < .001$, $r_s = -.29$. The direction of the effect indicated that reducing the amount of study time for each face significantly increased the magnitude of the ORB, largely as a result of an increase in the proportion of false alarm responses to other-race faces, $Z_j = 2.65$, $r_s = -.09$. This effect of exposure time is similar to the findings of Anthony et al. (1992) across their meta-analytic

Table 2
Weighted Effect Size Estimates (g^U) on Performance Measures as a Function of Race/Ethnicity of Participant

Race/ethnicity of participant	Hits	False alarms	Discrimination accuracy	Response criterion
Whites	0.35	−0.62	1.06	0.38
Blacks	0.32	−0.15	0.66	0.32
Others	0.04	−0.22	0.74	−0.21

Table 3
Weighted Effect Size Estimates (g^U) on Performance Measures as a Function of Changes in Stimuli and Order of Study by Race of Face

Moderator	Hits	False alarms	Discrimination accuracy	Response criterion
Stimuli at study and test				
Identical	0.20	−0.36	0.76	0.14
Different	0.37	−0.48	0.82	0.53
Order of study by race of face				
Mixed	0.21	−0.38	0.79	0.30
Blocked	0.45	−0.45	1.18	0.37

sample of White participants. Length of the retention interval had a significant influence on the size of the ORB across estimates of response criterion, $Z_j = 7.17$, $p < .001$, $r_s = .18$. The direction of the effect indicated that lengthening the retention interval induced more liberal responding to other-race faces.

Extrinsic study factors. As a final set of moderator variables, effect sizes were coded for whether they had been taken from a published or unpublished manuscript and for the date of the manuscript's publication or presentation. Results indicated no significant differences in the magnitude of effect sizes taken from published and unpublished manuscripts, Z_js ≤ 1.74. However, a rather interesting effect was found for date of study, most significantly across measures of false alarm and discrimination, Z_js ≥ 5.46, ps $< .001$. Weighted means for each decade are presented in Table 4. It appears that whereas the size of the ORB has significantly decreased over time for measures of discrimination accuracy, $r_s = -.22$, and proportion of hits, $r_s = -.06$, it has significantly increased over time for the proportion of false alarms, $r_s = -.17$. Curiously, this effect does not hold for estimates of response criterion.

Influence of date of study on estimates of attitude and contact. With regard to estimates of racial attitude and interracial contact, we also assessed the effect of date of study on the magnitude of effects observed. Whereas the estimated influence of racial attitudes on recognition of other-race faces has significantly decreased over the past 3 decades, $Z_j = 16.67$, $p < .001$, $r_s = -.46$, the influence of interracial contact on recognition has significantly increased, $Z_j = 9.28$, $p < .001$, $r_s = .40$ (see Table 5). As noted previously, the increase in magnitude of

Table 4
Weighted Effect Size Estimates (g^U) on Performance Measures as a Function of Date of Study

Date of study	Hits	False alarms	Discrimination accuracy	Response criterion
1970s	0.32	−0.28	1.35	0.39
1980s	0.23	−0.38	0.72	0.11
1990s	0.21	−0.41	0.64	0.32

Table 5
Weighted Effect Size Estimates (Z_r) of Influence of Racial Attitudes and Interracial Contact as a Function of Date of Study

Date of study	Attitudes	Contact
1970s	0.06	−0.01
1980s	0.02	0.19
1990s	−0.08	0.27

effect of contact over the past 3 decades may be due to a cohort effect resulting from increases in the opportunities for interracial contact between groups (Chance & Goldstein, 1996). Alternatively, the increase may be due to improved precision and validity in the measures used to assess interracial contact. Nevertheless, it is increasingly evident that the contact hypothesis plays a vital role in our conception of the ORB.

Discussion

The present meta-analysis has empirically reviewed over 30 years of research on the ORB in memory for faces. Thirty-nine research articles were located, involving the combined responses of nearly 5,000 participants. Analyses examined differences in performance on own-race and other-race faces across measures of hit and false alarm responses and across aggregate measures of discrimination accuracy and response criterion. Results of hit and false alarm rates illustrated an ORB mirror-effect pattern in which own-race faces produced a higher proportion of hits and a lower proportion of false alarms compared with other-race faces (see Figure 1). Consistent with this effect, measures of discrimination accuracy demonstrated a significant, moderately sized ORB, accounting for 15% of the variability across samples. Measures of response criterion also showed a significant ORB; however, this effect was considerably smaller, accounting for only 1% of the variability across samples.

In addition, estimates of the influence of both racial attitudes and interracial contact on the ORB were examined across studies. Although no influence of racial attitudes was present in the sample, a small, yet significant, effect of interracial contact was found, accounting for approximately 2% of the variability across the sample. Although racial attitudes appeared to have no direct influence on the ORB, a possible mediating role was indicated by a moderately strong relationship between racial attitudes and interracial contact, accounting for 13% of the variability.

Several study moderators were also examined across the various measures. Results indicated that White participants were more likely to demonstrate the ORB, especially with regard to false alarm responses. Additionally, ORB effects were more likely in measures of discrimination accuracy when presentation and testing were blocked by race of face and when study time was reduced. Measures of response criterion demonstrated ORB effects when stimuli differed between study and test and when the retention interval between study and test was increased. Finally, date of study had a significant influence on both false alarm

and discrimination measures. Results indicated that, over the past 3 decades, the ORB effect appears to have become most prominent in false alarm responses. Measures of the influence of racial attitudes and interracial contact were also affected by date of study, such that the effect of racial attitudes on other-race face recognition has decreased, whereas the effect of interracial contact has increased in more recent years.

Theoretical Implications

The pattern of hit and false alarm responses across studies exhibited a mirror-effect pattern (see Figure 1). This pattern of responses has been demonstrated across a number of manipulations in the literature and has been deemed a "regularity" of recognition memory (Glanzer & Adams, 1985, 1990). It is interesting that this mirror-effect pattern is often captured in aggregate signal detection measures of discrimination accuracy and response criterion, consistent with our meta-analytic results. Much of the debate regarding this phenomenon has involved whether the mirror effect pattern results from a change in the response criterion for each stimulus set, or whether the effects observed on the response criterion measures represent an actual change in the psychological sense of familiarity resulting from the manipulation.

In support of the latter hypothesis, McClelland and Chappell (1998) have proposed a model of recognition memory involving a mechanism of differentiation. As discussed previously, differentiation is a process in which the perceiver focuses attention toward invariant cues that provide the best basis for discriminations within a given stimulus set (Gibson, 1969). McClelland and Chappell model this process by proposing that individuals store features of a given stimulus in memory and that these features (and their associated probabilities) are updated in the representation each time the individual encounters the particular stimulus, thereby resulting in an increase in the psychological sense of familiarity. Furthermore, this increase in the strength of the representation is accompanied by a decrease in the likelihood of responding to a novel, unrelated stimulus. Thus, as McClelland and Chappell conclude, "familiarity breeds differentiation" (p. 726).

McClelland and Chappell's (1998) model effectively simulates the mirror-effect pattern across measures of both discrimination accuracy and response criterion. In doing so, the authors note that the response criterion effects are reproduced despite the fact that the model actually holds the response criterion constant across the stimulus manipulation. Thus, the fluctuation in response criterion is produced as a function of changes in the distributions of new and old items across the familiarity continuum, and not as a result of shifts in the location of the criterion itself. A recent model by Shiffrin and Steyvers (1997) also reproduced these results and was similarly based on the process of differentiation; however, some conceptual differences do exist between the two approaches.

With regard to the ORB, McClelland and Chappell's (1998) model suggests that individuals store own-race faces more accurately and efficiently with respect to the appropriate featural and configural information represented in memory. This accuracy and efficiency may be the result of prior experience (or familiarity) with own-race faces that has led to the ability in attending to the proper invariant aspects of the face. Other-race faces, on the other hand, appear to be encoded in

a less efficient manner, in which fewer or inappropriate cues are selected for storage. When later presented with a recognition task, such differences in encoding result in both differential discrimination accuracy and criterion of responding to own-race versus other-race faces. However, this apparent difference in response criterion occurs as a byproduct of differentiation processes in which the underlying distributions of own- and other-race faces shift along the familiarity continuum. In practical terms, our general familiarity with other-race faces, in the absence of an appropriate representation of features in memory, leads to differential responding in acknowledging the familiarity of the face. As such, this apparent difference in response criterion indicates the role of increased variability in the encoding of featural and/or configural information of other-race faces when compared with the more consistent (less variable) representation of own-race faces.

In summary, the mirror-effect pattern across hit and false alarm responses, together with the associated discrimination accuracy and response criterion effects, suggest a process of differentiation consistent with several recent models of recognition memory (McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997). The implications of this type of model are consistent with the perceptual learning framework outlined previously, including research on the configural–featural (Diamond & Carey, 1986; Rhodes et al., 1989) and race-feature hypotheses (Levin, 1996), as well as the representational model proposed by Valentine and his colleagues (Chiroro & Valentine, 1995; Valentine, 1991). Furthermore, the importance of prior research on the influence of interracial contact, particularly with regard to the effects of discrimination training (e.g., Malpass et al., 1973) and prior experience with other-race faces (Chiroro & Valentine, 1995; Li et al., 1998), are substantiated within this theoretical framework. Although we have previously discussed the potential importance of response criterion measures in the ORB based on findings in our lab (Slone et al., 2000), few studies currently in the literature have documented this effect. Future research that more thoroughly investigates the importance of response criterion can further distinguish its role in the differentiation process.

Applied Considerations

From an applied perspective, several issues merit further discussion. First, the magnitude of the ORB that has been found across many studies, accounting for 15% of the variance in discrimination accuracy, indicates that this is an issue of considerable practical importance. Although our analyses demonstrated that the overall magnitude of the effect on discrimination accuracy has decreased over the past 2 decades, it was also observed that the influence of false alarm responses on the ORB has actually increased during that same period. We believe this to be of great practical significance, as it is precisely the existence of false alarms, namely the erroneous identification of an individual who is not the perpetrator, with which attorneys, judges, and researchers have been most concerned. For example, a recent U.S. Department of Justice report focused on 28 cases in which felony convictions were overturned due to subsequent DNA analyses. In over 85% of those cases, erroneous eyewitness identifications (i.e., false alarms) were the

primary evidence that led to the original conviction (Connors, Lundregan, Miller, & McEwan, 1996).

Second, our moderator analyses indicated that both recognition and lineup identification tasks yield similar ORB estimates across studies. Although a trend was present for lineup tasks to demonstrate a larger ORB effect on correct identifications, more studies involving the use of lineup tasks are needed to better assess the reliability of this effect. Furthermore, as R. C. Lindsay and Wells (1983) noted some time ago, it is important that researchers also manipulate the presence or absence of the target such that they might examine the influence of *diagnosticity* (i.e., the ratio of correct identifications to false identifications) in the other-race lineup situation.

Our moderator analyses also demonstrated that the amount of study time significantly influenced discrimination accuracy in the ORB, particularly through an increase in false alarm responses to other-race faces when study time is limited. Although the application of the laboratory-based term "study time" to the crime situation may seem forced, it should be noted that many crimes involving eyewitnesses occur in a matter of seconds (e.g., assaults, murders, some robberies). This short period of time would involve very limited "study time" for the eyewitness, hence increasing the chances of subsequent false alarms (i.e., mistaken identifications) in cross-race situations.

Moderator analyses also indicated that the length of the retention interval between study and test influenced the ORB through a change in the response criterion. More specifically, this effect indicated that as the length of time increased between study and test, participants increasingly adopted a more liberal response criterion when responding to other-race faces. This liberal response criterion indicated that participants required less evidence from memory (e.g., familiarity or memorability of the face) to respond that they had previously seen an other-race face. In actual cases, the time between viewing the suspect at the crime and later attempting an identification can range between days, weeks, months, and even years. Given this influence of response criterion, the legal community should be cautious of cross-race identifications attempted after such extensive delays.

In the eyewitness literature, researchers have made a distinction between *system variables*—those that are, at least in principle, controllable by the criminal justice system (e.g., interviewing techniques) and *estimator variables*—those that can be manipulated experimentally, but that are not controllable in actual cases; their influence can only be "estimated" post hoc (e.g., the ORB) (Wells, 1978; Wells, Wright, & Bradfield, 1999). Some have suggested that greater research attention should be directed toward system variables, because research results may be more directly applicable to police procedures and legal policy. However, there is one related aspect of the ORB that does involve system variables, namely the procedures used in the construction of identification lineups (Brigham, Meissner, & Wasserman, 1999; Brigham & Ready, 1985). Brigham and Ready (1985) found that race influenced the manner in which individuals constructed lineups, such that both Blacks and Whites used a looser criterion (i.e., more faces were seen as similar to each other, and therefore as useful in a lineup) when constructing lineups of other-race faces as compared with constructing own-race lineups. Hence, there was a tendency to construct fairer lineups (in which the faces were

actually similar to one another) when working with own-race faces. Further research on this application of the ORB seems warranted.

Legal "Safeguards" to the ORB in Eyewitness Identification

As we have discussed elsewhere (Brigham, Wasserman, & Meissner, 1999), several purported "safeguards" are available to defendants accused primarily on the basis of eyewitness evidence, including cross-examination by defense counsel, cautionary instructions to jurors, and expert testimony regarding eyewitness evidence. Although cross-examination has not been shown effective in allowing jurors to distinguish accurate from inaccurate eyewitnesses (R. C. Lindsay, Wells, & O'Connor, 1989; R. C. Lindsay, Wells, & Rumpel, 1981), cautionary jury instructions may have some potential (Cutler, Dexter, & Penrod, 1990; Greene, 1988; Katzev & Wishart, 1985), assuming that they contain accurate information. Unfortunately, such instructions are typically written by legal scholars who have little knowledge of the research findings.

What might more appropriate model jury instructions include? Based on the survey responses of researchers classed as "eyewitness experts" (Kassin et al., 1989) and the results of research meta-analyses, useful model jury instructions could summarize the negative impacts of several factors on the accuracy of eyewitness memory, each of which was listed by over 70% of the experts in the Kassin et al. survey (see also Leippe, 1995). These include short exposure time, high stress, misleading postevent information, and biased lineup instructions. Model instructions could also describe potential problems due to unconscious transference, cross-race identifications, unfair lineups, and the use of showups. They could also point out that expressed confidence or certainty about an identification is not a strong indicator of accuracy (Bothwell, Deffenbacher, & Brigham, 1987; Penrod & Cutler, 1995; Sporer, Penrod, Read, & Cutler, 1995).

The most commonly cited jury instructions are likely those in *United States v. Telfaire* (1972), in which the U.S. Court of Appeals for the District of Columbia endorsed the use of a cautionary instruction on eyewitness evidence. The *Telfaire* instructions state that the juror should evaluate whether the witness "had the capacity and an adequate opportunity to observe the defendant," and whether the witness's identification "was the product of his [sic] own recollection." Jurors are told that they may also take into account "the strength of the identification [certainty]," whether the identification "may have been influenced by the circumstances under which the defendant was presented to him [sic] for identification," and the "length of time that lapsed between the occurrence of the crime and the next opportunity of the witness to see the defendant" (Cutler & Penrod, 1995, pp. 255–256).

Although *Telfaire* was seen by some as a positive step, researchers have noted many shortcomings. For example, the instructions fail to specify in which direction each factor should influence an evaluation of the eyewitness. Furthermore, the *Telfaire* decision was based largely on the five factors listed by the Supreme Court in *Neil v. Biggers* (1972). These factors included (a) the witness's opportunity to view the suspect during the crime, (b) the length of time between the crime and the subsequent identification, (c) the level of certainty demonstrated by the witness during the identification, (d) the (apparent) accuracy of the witness's

prior description of the suspect, and (e) the witness's degree of attention during the crime. However, only two of these five factors have been clearly supported by research findings (see Brigham, Wasserman, & Meissner, 1999), namely opportunity to view the suspect and the retention interval between viewing and identification of the suspect. Finally, many factors shown by research to be relevant to eyewitness accuracy, such as the ORB, stress, weapon focus, lineup bias, and so forth, are not mentioned in the *Telfaire* instructions.[3]

On a more positive note, the New Jersey Supreme Court recently held that in cases involving a cross-race identification, the defendant is entitled to jury instructions specifically warning jurors about the potential for misidentification of other-race persons (*State v. Cromedy,* 1999). In this case, a Black intruder sexually assaulted a White college student in her apartment. Eight months later the victim saw a man on the street whom she believed to be her assailant. The man was immediately picked up, and the woman identified him 15 min later in a one-person "showup." It is interesting to note that the victim had failed to identify the same man from a photograph lineup only 2 days after the initial assault! The New Jersey Supreme Court, citing in its decision some of the studies included in our meta-analysis, ruled that a cross-race identification, as a subset of eyewitness identification, requires a special jury instruction in the appropriate case. Namely, the instruction should be given when the cross-racial identification is a critical issue in the case, especially when other evidence does not corroborate it. Unfortunately, the instruction advocated by the New Jersey Supreme Court was not directional. The instruction indicated that jurors "may consider, if you think it is appropriate to do so, whether the cross-racial nature of the identification has affected the accuracy of the witness's original perception and/or accuracy of a subsequent identification," without indicating what that effect might be.

With regard to expert testimony, the present meta-analysis results provide additional material that could be presented by an eyewitness expert in cases involving disputed eyewitness evidence. The present findings provide strong evidence of the reliability of the ORB effect, based on the responses of almost 5,000 respondents. The analyses yield meaningful indexes of the strength of the effect, namely that it accounts for 15% of the variance in discrimination accuracy or, alternatively, that participants were over 2.2 times as likely to accurately identify own-race faces as new versus old, when compared with performance on other-race faces. The findings indicate that the majority of errors for other-race faces are false alarms, that is, incorrectly identifying an other-race face as having been seen before. This is the type of error that is generally seen as most harmful in a crime situation. The results show that the ORB is not related to the level of racial prejudice. Finally, factors such as study time and retention interval play an important role in determining when the ORB is most likely to occur.

Given both the reliability of the ORB shown in the present analysis (especially with regard to false alarm responses) and the general agreement among researchers regarding the importance of the phenomenon (Kassin et al., 1989;

[3]In his concurring opinion in *Telfaire,* Chief Judge Bazelon urged that juries be warned of the pitfalls of cross-racial identification. Unfortunately, this caution was not included in the final version of the instructions.

Yarmey & Jones, 1983), we advocate the use of expert testimony in cases involving disputed cross-racial eyewitness evidence. Although prior research has demonstrated that the factors influencing eyewitness testimony often reach beyond jurors' common knowledge (Brigham & Bothwell, 1983; Devenport, Penrod, & Cutler, 1997), the courts have often prohibited expert testimony on eyewitness identification, including the ORB (e.g., *People v. Dixon,* 1980; *United States v. Hudson,* 1989; *United States v. Watson,* 1978), ruling that such testimony would not be helpful to jurors. However, in a recent case, *United States v. Norwood* (1996), the U.S. District Court for the District of New Jersey ruled in support of expert testimony on cross-racial identification, along with several other factors. In its decision, the court reasoned that such expert testimony would not confuse or overwhelm the jury. Rather, the "defendant's expert's proposed testimony regarding cross-racial identification was sufficiently tied to facts of [the] case and would be helpful to [the] jury" (p. 1133). The decision relied heavily on the 1985 *United States v. Downing* decision, which held that such expert testimony should (a) properly "fit" the particular features of the case, (b) be based on reliable scientific principles, and (c) not confuse or overwhelm the jury. Expert testimony on cross-racial identifications was also found to be helpful to the jury in *United States v. Stevens* (1984) and *United States v. Smith* (1984).

In closing, the present meta-analysis has yielded many intriguing findings that appear both to bolster our current understanding of the mechanisms responsible for the ORB effect and to illuminate new directions for future research. We believe that previous research has sufficiently underscored the robustness of the phenomenon and illustrated the potential for moderator variables in defining its limits. The current analysis sought only to bring together these findings and to discuss the potential for various theoretical frameworks that might account for the pattern of results across studies. Overall, the ORB was found to be a reliable and generalizable phenomenon, deserving of further theoretical consideration. Moreover, the strong influence of false identifications in the ORB indicates that this issue is of great practical importance as well.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Ahissar, M. & Hochstein, S. (1998). Perceptual learning. In V. Walsh & J. Kulikowski (Eds.), *Perceptual constancy: Why things look as they do* (pp. 455–498). New York: Cambridge University Press.

Allport, G. W. & Kramer, B. M. (1946). Some roots of prejudice. *Journal of Psychology, 22,* 9–39.

Anthony, T., Copper, C. & Mullen, B. (1992). Cross-racial facial identification: A social cognitive integration. *Personality & Social Psychology Bulletin, 18,* 296–301.

*Ayuk, R. E. (1990). Cross-racial identification of transformed, untransformed, and mixed-race faces. *International Journal of Psychology, 25,* 509–527.[A,B,C]

*Barkowitz, P. & Brigham, J. C. (1982). Recognition of faces: Own-race bias, incentive, and time delay. *Journal of Applied Social Psychology, 12,* 255–268.[C,D]

Benton, A. L., Hamsher, K. deS., Varney, N. R. & Spreen, O. (1983). *Contributions to neuropsychological assessment.* New York: Oxford University Press.

*Berger, D. G. (1969). *They all look alike.* Unpublished doctoral dissertation, Vanderbilt University.[A,E,F]

Biederman, I. & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert

systems analysis of a difficult perceptual learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 640–645.

Bothwell, R. K., Brigham, J. C. & Malpass, R. S. (1989). Cross-racial identification. *Personality & Social Psychology Bulletin, 15,* 19–25.

Bothwell, R. K., Deffenbacher, K. A. & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology, 72,* 691–695.

Brigham, J. C. (1981). The accuracy of eyewitness evidence: How do attorneys see it? *Florida Bar Journal, 55,* 714–721.

Brigham, J. C. (1990). Target person distinctiveness and attractiveness as moderator variables in the confidence–accuracy relationship in eyewitness identifications. *Basic & Applied Social Psychology, 11,* 101–115.

*Brigham, J. C. (1993). College students' racial attitudes. *Journal of Applied Social Psychology, 23,* 1933–1967.[G]

*Brigham, J. C. & Barkowitz, P. (1978). Do "They all look alike"?: The effect of race, sex, experience, and attitudes on the ability to recognize faces. *Journal of Applied Social Psychology, 8,* 306–318.[C,E,F,G]

Brigham, J. C. & Bothwell, R. K. (1983). The ability of prospective jurors to estimate the accuracy of eyewitness identifications. *Law & Human Behavior, 7,* 19–30.

*Brigham, J. C., Maass, A., Snyder, L. D. & Spaulding, K. (1982). Accuracy of eyewitness identifications in a field setting. *Journal of Personality and Social Psychology, 42,* 673–681.[A,B,E]

Brigham, J. C. & Malpass, R. S. (1985). The role of experience and contact in the recognition of faces of own- and other-race persons. *Journal of Social Issues, 41(3),* 139–155.

*Brigham, J. C. & Meissner, C. A. (2000, March). Representation and memory for same- and other-race faces. In J. C. Brigham (Chair), *What do we know about the "own-race bias" in face recognition?* Symposium conducted at the biennial meeting of the American Psychology-Law Society, New Orleans, LA.[G]

Brigham, J. C., Meissner, C. A. & Wasserman, A. W. (1999). Applied issues in the construction and expert assessment of photo lineups. *Applied Cognitive Psychology, 13,* S73–S92.

*Brigham, J. C. & Ready, D. J. (1985). Own-race bias in lineup construction. *Law & Human Behavior, 9,* 415–424.[G]

Brigham, J. C., Wasserman, A. W. & Meissner, C. A. (1999). Disputed eyewitness identification evidence: Important legal and scientific issues. *Court Review, 36,* 12–25.

*Brigham, J. C. & Williamson, N. L. (1979). Cross-racial recognition and age: When you're over 60, do they still "all look alike?" *Personality & Social Psychology Bulletin, 5,* 218–222.[C]

Brigham, J. C. & WolfsKeil, M. P. (1983). Opinions of attorneys and law enforcement personnel on the accuracy of eyewitness identifications. *Law & Human Behavior, 7,* 337–349.

Bruce, A. J., Beard, K. W. & Tedford, S. (1997). African Americans' and Caucasian American faces. *Journal of General Psychology, 124,* 143–156.

*Buckhout, R. & Regan, S. (1988). Explorations in research on the other-race effect in face recognition. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues: Volume 1. Memory in everyday life* (pp. 40–46). New York: Wiley.[A,B]

*Burgess, M. C. R. (1997). *The cross-race effect in facial recognition: A function of expertise?* Unpublished doctoral dissertation, Florida State University.[C]

*Byatt, G. & Rhodes, G. (1998). Recognition of own-race and other-race caricatures: Implications for models of face recognition. *Vision Research, 38,* 2455–2468.[E]

*Carroo, A. W. (1986). Other race recognition: A comparison of Black American and African subjects. *Perceptual & Motor Skills, 62,* 135–138.[E]

*Carroo, A. W. (1987). Recognition of faces as a function of race, attitudes, and reported cross-racial friendships. *Perceptual & Motor Skills, 64,* 319–325.[A,B,C,E,F]

Carroo, A. W. (1988). Arousal in differential recognition. *Perceptual & Motor Skills, 67,* 696.

Carter, L. F. (1948). The identification of "racial membership." *Journal of Abnormal & Social Psychology, 51,* 339–341.

Chance, J. E. & Goldstein, A. G. (1979). Reliability of face recognition performance. *Bulletin of the Psychonomic Society, 14,* 115–117.

Chance, J. E. & Goldstein, A. G. (1987). Retention interval and face recognition: Response latency measures. *Bulletin of the Psychonomic Society, 25,* 415–418.

Chance, J. E. & Goldstein, A. G. (1996). The other-race effect and eyewitness identification. In S. L. Sporer, R. S. Malpass, & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 153–176). Mahwah, NJ: Erlbaum.

*Chance, J. E., Goldstein, A. G. & McBride, L. (1975). Differential experience and recognition memory for faces. *Journal of Social Psychology, 97,* 243–253.[A]

*Chance, J. E., Turner, A. L. & Goldstein, A. G. (1982). Development of differential recognition for own- and other-race faces. *Journal of Psychology, 112,* 29–37.[A,B,C]

*Chiroro, P. & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *Quarterly Journal of Experimental Psychology, 48A,* 879–894.[A,B,C,E]

Christensen, E. E., Murry, R. C., Holland, K., Reynolds, J., Landay, M. J. & Moore, J. G. (1981). The effect of search time on perception. *Radiology, 138,* 361–365.

Connors, E., Lundregan, T., Miller, N. & McEwan, T. (1996). *Convicted by juries, exonerated by science: Case studies in the use of DNA evidence to establish innocence after trial.* Washington, DC: National Institute of Justice.

*Cross, J. F., Cross, J. & Daly, J. (1971). Sex, race, age, and beauty as factors in recognition of faces. *Perception and Psychophysics, 10,* 393–396.[A,B,E]

Cutler, B. L., Dexter, H. R. & Penrod, S. D. (1990). Nonadversarial methods for sensitizing jurors to eyewitness evidence. *Journal of Applied Social Psychology, 20,* 1197–1207.

Cutler, B. L. & Penrod, S. D. (1995). *Mistaken identification: The eyewitness, psychology, and the law.* Cambridge, England: Cambridge University Press.

Devenport, J. L., Penrod, S. D. & Cutler, B. L. (1997). Eyewitness identification evidence: Evaluating commonsense evaluations. *Psychology, Public Policy, & Law, 3,* 338–361.

*Devine, P. G. & Malpass, R. S. (1985). Orienting strategies in differential face recognition. *Personality and Social Psychology Bulletin, 11,* 33–40.[C]

Diamond, R. & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115,* 107–117.

Doty, N. D. (1998). The influence of nationality on the accuracy of face and voice recognition. *American Journal of Psychology, 111,* 191–214.

Elliott, D. N. & Wittenberg, B. H. (1955). Accuracy of identification of Jewish and non-Jewish photographs. *Journal of Abnormal & Social Psychology, 51,* 339–341.

Elliott, E. S., Wills, E. J. & Goldstein, A. G. (1973). The effects of discrimination training on the recognition of White and Oriental faces. *Bulletin of the Psychonomic Society, 2,* 71–73.

Ellis, H. D., Davies, G. M. & McMurran, M. M. (1979). Recall of White and Black faces by White and Black witnesses using the Photofit system. *Human Factors, 21,* 55–59.

*Ellis, H. D. & Deregowski, J. B. (1981). Within-race and between-race recognition of transformed and untransformed faces. *American Journal of Psychology, 94,* 27–35.[A,B,C]

Ellis, H. D., Deregowski, J. B. & Shepherd, J. W. (1975). Descriptions of White and Black faces by White and Black subjects. *International Journal of Psychology, 10,* 119–123.

*Fallshore, M. & Schooler, J. W. (1995). Verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1608–1623.[A]

Farah, M. J., Wilson, K. D., Drain, M. & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105,* 482–498.

Feinman, S. & Entwisle, D. R. (1976). Children's ability to recognize other children's faces. *Child Development, 47,* 506–510.

*Galper, R. E. (1973). "Functional race membership" and recognition of faces. *Perceptual & Motor Skills, 37,* 455–462.[A]

Gibson, E. J. (1969). *Principles of perceptual learning and development.* New York: Appleton-Century-Crofts.

Glanzer, M. & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13,* 8–20.

Glanzer, M. & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 5–16.

Goldstein, A. G. (1979). Race-related variation of facial features: Anthropometric data I. *Bulletin of the Psychonomic Society, 13,* 187–190.

Goldstein, A. G. & Chance, J. E. (1971). Visual recognition memory for complex configurations. *Perception & Psychophysics, 9,* 237–241.

Goldstein, A. G. & Chance, J. E. (1976). Measuring psychological similarity of faces. *Bulletin of the Psychonomic Society, 7,* 407–408.

Goldstein, A. G. & Chance, J. E. (1978). Judging face similarity in own and other races. *Journal of Psychology, 98,* 185–193.

Goldstein, A. G. & Chance, J. E. (1979). Do "foreign" faces really look alike? *Bulletin of the Psychonomic Society, 13,* 111–113.

*Goldstein, A. G. & Chance, J. E. (1980). Memory for faces and schema theory. *Journal of Psychology, 105,* 47–59.[A,B,C]

Goldstein, A. G. & Chance, J. E. (1985). Effects of training on Japanese faces recognition: Reduction of the other-race effect. *Bulletin of the Psychonomic Society, 23,* 211–214.

Green, D. & Swets, J. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Greene, E. (1988). Judge's instruction on eyewitness testimony: Evaluation and revision. *Journal of Applied Social Psychology, 18,* 252–276.

Haider, H. & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology, 30,* 304–337.

Haider, H. & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 172–190.

Hedges, L. V. (1994). Fixed effects models. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 285–299). New York: Sage.

Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Helsen, W. & Pauwels, J. M. (1993). The relationship between expertise and visual information processing in sport. In J. L. Starkes, & F. Allard (Eds.), *Cognitive issues in motor expertise* (pp. 109–134). Amsterdam: North-Holland.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 302–313.

Horowitz, E. L. & Horowitz, R. E. (1938). Development of social attitudes in children. *Sociometry, 1,* 301–338.

Hosie, J. A. & Milne, A. B. (1995). Distinctiveness and memory for unfamiliar faces. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition* (pp. 95–112). London: Routledge.

*Jalbert, N. L. & Getting, J. (1992). Racial and gender issues in facial recognition. In F. Losel, D. Bender, & T. Bliesener (Eds.), *Psychology and law: International perspectives* (pp. 309–316). Berlin, Germany: Walter de Gruyter.[A,B]

Johnson, B. T., Mullen, B. & Salas, E. (1995). Comparison of three major meta-analytic approaches. *Journal of Applied Psychology, 80,* 94–106.

Johnson, K. E. & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 126,* 248–277.

Johnson, K. E. & Mervis, C. B. (1998). Impact of intuitive theories on feature recruitment throughout the continuum of expertise. *Memory & Cognition, 26,* 382–401.

Kamphaus, R. W., Beres, K. A., Kaufman, A. S. & Kaufman, N. L. (1996). The Kaufman Assessment Battery for Children (K–ABC). In C. S. Newmark (Ed.), *Major psychological assessment instruments* (pp. 348–399). Boston, MA: Allyn & Bacon.

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.

Katzev, R. D. & Wishart, S. S. (1985). The impact of judicial commentary concerning eyewitness identifications on jury decision making. *Journal of Criminal Law & Criminology, 76,* 733–745.

*Lavrakas, P. J., Buri, J. R. & Mayzner, M. S. (1976). A perspective on the recognition of other race faces. *Perception & Psychophysics, 20,* 475–481.[F]

Leippe, M. R. (1995). The case for expert testimony about eyewitness memory. *Psychology, Public Policy, and Law, 1,* 909–959.

Lesgold, A., Glaser, R., Rubinson, H., Klopfer, D., Feltovich, P. & Wang, Y. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In M. Chi, R. Glaser, & M. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale, NJ: Erlbaum.

*Levin, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1364–1382.[A,B,C]

Levin, D. T. & Lacruz, I. (1999, November). *An alternative to the encoding expertise explanation for the cross-race recognition deficit.* Poster presented at the annual meeting of the Psychonomic Society, Los Angeles, CA.

*Li, J. C., Dunning, D. & Malpass, R. S. (1998, March). *Cross-racial identification among European-Americans: Basketball fandom and the contact hypothesis.* Paper presented at the biennial meeting of the American Psychology–Law Society, Redondo Beach, CA.[A,B,C,E,F]

Lickson, J. (1974). *David Charles: The story of the Quincy Five.* Tallahassee, FL: Mockingbird Press, Inc.

*Lindsay, D. S., Jack, P. C. & Christian, M. A. (1991). Other-race face perception. *Journal of Applied Psychology, 76,* 587–589.[A,E]

Lindsay, R. C. & Wells, G. L. (1983). What do we really know about cross-race eyewitness identification? In S. M. A. Lloyd-Boystock, & B. R. Clifford (Eds.),

Evaluating witness evidence: Recent psychological research and new perspectives (pp. 219–234). New York: Wiley.

Lindsay, R. C., Wells, G. L. & O'Connor, F. J. (1989). Mock-juror belief of accurate and inaccurate eyewitnesses: A replication and extension. *Law & Human Behavior, 13,* 333–339.

Lindsay, R. C., Wells, G. L. & Rumpel, C. M. (1981). Can people detect eyewitness-identification accuracy within and across situations? *Journal of Applied Psychology, 66,* 79–89.

Lindzey, G. & Rogolsky, S. (1950). Prejudice and identification of minority group membership. *Journal of Abnormal & Social Psychology, 45,* 37–53.

Lipsey, M. W. (1994). Identifying potentially interesting variables and analysis opportunities. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 111–124). New York: Russell Sage Foundation.

Luce, T. S. (1974). The role of experience in inter-racial recognition. *Personality & Social Psychology Bulletin, 1,* 39–41.

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

Malina, A. C., Bowers, D. A., Millis, S. R. & Uekert, S. (1998). Internal consistency of the Warrington Recognition Memory Test. *Perceptual & Motor Skills, 86,* 1320–1322.

*Malpass, R. S. (1974, September). *Racial bias in eyewitness identification.* Paper presented at the 82nd Annual Convention of the American Psychological Association, New Orleans, LA.[A]

Malpass, R. S. (1981). Training in face recognition. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 271–285). London: Academic Press.

Malpass, R. S. (1988). Psychological differentiation, gaze and face recognition. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues: Volume 1. Memory in everyday life* (pp. 145–150). New York: Wiley.

Malpass, R. S. (1990). An excursion into utilitarian analyses, with side trips. *Behavioral Science Research, 24,* 1–15.

Malpass, R. S., Corey, D., Parada, M., Chavez, J., Bowles, S. & McQuiston, D. (1998, March). *Reliability of face recognition: Take two.* Paper presented at the biennial meeting of the American Psychology–Law Society, Redondo Beach, CA.

Malpass, R. S., Erskine, D. M. & Vaughn, L. L. (1988, April). *Matching own- and other-race faces.* Paper presented at the annual meeting of the Eastern Psychological Association, Buffalo, NY.

*Malpass, R. S. & Kravitz, J. (1969). Recognition for faces of own- and other-race faces. *Journal of Personality and Social Psychology, 13,* 330–334.[A,B,C]

*Malpass, R. S., Lavigueur, H. & Weldon, D. E. (1973). Verbal and visual training in face recognition. *Perception & Psychophysics, 14,* 283–292.[A,B,C]

McClelland, J. L. & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105,* 724–760.

Meissner, C. A. & Brigham, J. C. (in press). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology.*

Myles-Worsley, M., Johnston, W. A. & Simons, M. A. (1988). The influence of expertise on x-ray image processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 553–557.

Neil v. Biggers, 409 U.S. 188 (1972).

*Ng, W. & Lindsay, R. C. L. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology, 25,* 217–232.[A,B,C,D,E]

*O'Toole, A. J., Deffenbacher, K. A., Valentin, D. & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition, 22,* 208–224.[C]

Padgett, K. (1997). *A categorical investigation of the other-race effect in cross-race facial recognition.* Unpublished honor's thesis, Florida State University.

Penrod, S. D. & Cutler, B. L. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law, 1,* 817–845.

People v. Dixon, 410 N.E. 2d 252 (Ill. App. 1980).

*Platz, S. J. & Hosch, H. M. (1988). Cross-racial/ethnic eyewitness identification: A field study. *Journal of Applied Social Psychology, 18,* 972–984.[A,B,E,F]

Proctor, R. W. & Dutta, A. (1995). Perceptual skill. *Skill acquisition and human performance* (pp. 33–65). London: Sage Publications.

Prospero, M., Corey, D., Malpass, R. S., Parada, M. & Schreiber, N. (1996, February). *Is face recognition an ability? On the reliability of face recognition.* Paper presented at the biennial meeting of the American Psychology–Law Society, Hilton Head, SC.

*Rattner, A., Weimann, G. & Fishman, G. (1990). Cross-ethnic identifications and misidentifications by Israelis. *Sociology & Social Research, 74,* 73–79.[A,B]

Raudenbush, S. W. (1994). Random effects models. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301–322). New York: Sage.

Reingold, E. M., Charness, N., Pomplun, M. & Stampe, D. M. (in press). Visual span in expert chess players: Evidence from eye movements. *Psychological Science.*

*Rhodes, G., Brake, S., Taylor, K. & Tan, S. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology, 80,* 313–331.[A]

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 111–124). New York: Sage.

Secord, P. F., Bevan, W. & Katz, B. (1956). The Negro stereotype and perceptual accentuation. *Journal of Abnormal & Social Psychology, 53,* 78–83.

Shapiro, P. N. & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100,* 139–156.

Shea, J. B. & Paull, G. (1996). Capturing expertise in sports. In K. A. Ericsson (Ed.), *The road to excellence* (pp. 321–336). Mahwah, NJ: Erlbaum.

Shepherd, J. W. (1981). Social factors in face recognition. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 55–79). London: Academic Press.

Shepherd, J. W. & Deregowski, J. B. (1981). Races and faces: A comparison of the responses of Africans and Europeans to faces of the same and different races. *British Journal of Social Psychology, 20,* 125–133.

*Shepherd, J. W., Deregowski, J. B. & Ellis, H. D. (1974). A cross-cultural study of recognition memory for faces. *International Journal of Psychology, 9,* 205–211.[A,B,C,D]

Shiffrin, R. M. & Steyvers, M. (1997). A model of recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166.

*Slone, A. E., Brigham, J. C. & Meissner, C. A. (2000). Social and cognitive factors affecting the own-race bias in Whites. *Basic & Applied Social Psychology, 22,* 71–84.[A,B,C,D,E,F,G]

Soukop, V. M., Bimbela, A. & Schiess, M. C. (1999). Recognition memory for faces: Reliability and validity of the Warrington Recognition Memory Test (RMT) in a neurological sample. *Journal of Clinical Psychology in Medical Settings, 6,* 287–293.

*Sporer, S. L. (1992). *Das Wiedererkennen von Gesichtern* [Recognizing faces]. Weinheim, Germany: Beltz/Psychologie Verlags Union. [Experiments 15, 16, and 17].[A,B,C,D]

*Sporer, S. L. (1999, July). *The own-race bias in Germany: Testing the contact hypothesis with Turks and Germans.* Paper presented at the 4th European Conference on Psychology and Law of the American Psychology–Law Association and the European Psychology–Law Association, Dublin, Ireland.[A,B,C,D]

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

State v. Cromedy, 158 N.J. 112 (1999).

*Swope, T. M. (1994). *Social experience, illusory correlation and facial recognition ability.* Unpublished master's thesis, Florida State University.[A,B,C,E,F,G]

Tanaka, J. W. & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23,* 457–482.

*Teitelbaum, S. & Geiselman, R. E. (1997). Observer mood and cross-racial recognition of faces. *Journal of Cross-Cultural Psychology, 28,* 93–106.[A,B,C]

United States v. Downing, 753 F.2d 1224 (3rd Cir. 1985).

United States v. Hudson, 884 F.2d 1016 (7th Cir. 1989).

United States v. Norwood, 939 F.Supp. 1132 (D. N.J. 1996).

United States v. Smith, 736 F.2d 1103 (6th Cir. 1984).

United States v. Stevens, 736 F.2d 1103 (6th Cir. 1984).

United States v. Telfaire, 469 F.2d 552 (D.C. Cir. 1972).

United States v. Watson, 587 F.2d 365 (7th Cir. 1978).

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79,* 471–491.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race on face recognition. *Quarterly Journal of Experimental Psychology, 43A,* 161–204.

Valentine, T. & Bruce, V. (1986). The effect of race, inversion and encoding activity upon face recognition. *Acta Psychologica, 61,* 259–273.

Valentine, T. & Bruce, V. (1988). Mental rotation of faces. *Memory & Cognition, 16,* 556–566.

*Valentine, T. & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *Quarterly Journal of Experimental Psychology, 44A,* 671–703.[A,B,C]

Walk, R. D. (1978). Perceptual learning. In E. C. Carterette, & M. P. Friedman (Eds.), *Handbook of perception: Vol. IX. Perceptual processing* (pp. 257–297). New York: Academic Press.

Warrington, E. K. (1984). *Recognition Memory Test.* Windsor, Ontario, Canada: NFER-Nelson.

Wechsler Memory Scale: Technical manual (3rd ed.). (1997). San Antonio, TX: Psychological Corporation.

*Weimann, G., Fishman, G. & Rattner, A. (1986). Social distance and misidentification. *International Journal of Comparative Sociology, 27,* 217–225.[A,B]

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality & Social Psychology, 36,* 1546–1557.

Wells, G. L., Wright, E. F. & Bradfield, A. L. (1999). Witnesses to crime: Social and cognitive factors governing the validity of people's reports. In R. Roesch, S. D. Hart, & J. R. P. Ogloff (Eds.), *Psychology and law: The state of the discipline* (pp. 53–87). New York: Kluwer Academic/Plenum.

*Wright, D. B., Boyd, C. E. & Tredoux, C. G. (1999, June). *Own-race bias in South Africa and England.* Paper presented at the biennial meeting of the Society of Applied Research in Memory & Cognition, Boulder, CO.[A,B,C,E]

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

Yarmey, A. D. & Jones, H. P. T. (1983). Is the psychology of eyewitness identification a matter of common sense? In S. Lloyd-Bostock, & B. R. Clifford (Eds.), *Evaluating witness evidence: Recent psychological research and new perspectives* (pp. 13–40). Chichester, England: Wiley.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81,* 141–145.

Note. Studies contributed one or more of the following estimates: [A]hits; [B]false alarms; [C]discrimination; [D]response criterion; [E]racial attitudes—accuracy; [F]interracial contact—accuracy; [G]attitudes—contact.

Received April 12, 2000
Revision received October 20, 2000
Accepted October 24, 2000 ■

ORDER FORM

Start my 2001 subscription to *Psychology, Public Policy, and Law!* ISSN: 1076-8971

Send me a Free Sample Issue ❑

___ $31.00, APA Member/Affiliate _____
___ $61.00, Individual Non-Member _____
___ $136.00, Institution _____
In DC add 5.75% sales tax / In MD add 5% sales tax _____
TOTAL AMOUNT ENCLOSED $_____

Subscription orders must be prepaid. (Subscriptions are on a calendar basis only.) Allow 4-6 weeks for delivery of the first issue. Call for international subscription rates.

SEND THIS ORDER FORM TO:
American Psychological Association
Subscriptions
750 First Street, NE
Washington, DC 20002-4242

Or call (800) 374-2721, fax (202) 336-5568.
TDD/TTY (202)336-6123. Email: subscriptions@apa.org

AMERICAN PSYCHOLOGICAL ASSOCIATION

❑ Check Enclosed (make payable to APA)
Charge my: ❑ VISA ❑ MasterCard ❑ American Express
Cardholder Name _____
Card No. _____ Exp. date _____
_____
Signature (Required for Charge)
Credit Card _____
Billing Address _____
City _____ State _____ Zip _____
Daytime Phone _____
SHIP TO:
Name _____
Address _____
City _____ State _____ Zip _____
APA Customer # _____

GAD01

PLEASE DO NOT REMOVE – A PHOTOCOPY MAY BE USED

Judge's Instruction on Eyewitness Testimony: Evaluation and Revision[1]

EDITH GREENE[2]

University of Colorado

Studies of the reliability of eyewitness identification show that such testimony may frequently be inaccurate; because of this inherent unreliability, the law has established certain safeguards to the use of eyewitness evidence. One safeguard has been the development of an instruction that a judge may use to focus jurors' attention on the eyewitness issue. The effectiveness of this instruction has never been assessed, although other studies confirm that jurors frequently misunderstand or incorrectly use instructions they get from the judge. The purpose of these studies was to evaluate comprehension of this instruction in the context of a videotaped trial and to develop a simplified instruction that would be easier for jurors to understand. Compared to jurors who heard the existing instruction, those with the revised version were more knowledgeable of the factors to consider when listening to eyewitness testimony and were less likely to convict the defendant. A sample of superior court judges in the U.S. thought the simplified instruction was more effective than the existing version at conveying the intended legal concepts to the jury, but also rated it as more strongly biased toward the defense.

One common method of gathering evidence before trial is the use of eyewitness testimony. Eyewitnesses often play a crucial role both in criminal cases, particularly robberies and rapes, and in civil cases. The issue of eyewitness reliability arises not only when a victim or witness must identify another person, but whenever a witness must recall details about a past incident. Studies of the reliability of eyewitness testimony indicate that eyewitnesses' recollections are frequently inaccurate. Recent reviews of this work can be found in Yarmey (1979), Loftus (1979), Brigham (1983), and Wells and Loftus (1984).

[1]This research was supported by a Graduate Research Fellowship from the National Institute of Justice, Department of Justice, and by a grant from the National Science Foundation (Law and Social Sciences Program), and was submitted as a dissertation to the Department of Psychology at the University of Washington. I thank Elizabeth Loftus, chair of doctoral committee, for her helpful comments, and Carol Van Buren for assistance in data collection.

[2]Requests for reprints should be sent to Edith Greene, Department of Psychology, University of Colorado, Colorado Springs, CO 80933.

Journal of Applied Social Psychology, 1988, **18**, 3, pp. 252–276
.Copyright © 1988 by V. H. Winston & Sons, Inc. All rights reserved.

B-51

Because of the inherent unreliability of eyewitness testimony and the powerful impact it can have on a jury (e.g., Cavoukian, 1980; Loftus, 1979; Saunders, Vidmar, & Hewitt, 1983; Wells, Lindsay, & Tousignant, 1980), the law has established certain safeguards to the use of eyewitness identifications. One safeguard has been the development of an instruction for trial judges to use to focus jurors' attention on the potential unreliability of eyewitness testimony. One such instruction, drafted in the case of *U.S. v. TELFAIRE* (1972), is referred to as the TELFAIRE instruction. The full text is given in Table 1.

The TELFAIRE instruction merits attention because it is the most widely used standardized instruction relating to eyewitness identifications (Woocher, 1977) and because we know nothing about its effectiveness. In most cases, the judge determines the language of the jury instructions after conferring with lawyers representing both sides. In recent years, both federal and state jurisdictions have standardized their instructions in order to ease the judge's task and to ensure that the instructions are legally adequate. The TELFAIRE instruction has been adopted by both federal and state jurisdictions and has become the most frequently used cautionary message on eyewitness reliability. To date, all but one federal circuit court have approved use of a cautionary instruction on eyewitness identification, and most state courts have followed suit (Holtshouser, 1983).

Because of the importance and widespread use of the TELFAIRE instruction it is worth examining the extent to which jurors accurately comprehend and use the information contained in that instruction. With the exception of one study that examined whether jurors' verdicts change if they were given a cautionary message (Katzev & Wishart, 1985) this is a wholly unexplored area. To date, no one knows if a cautionary instruction can be an effective safeguard against potentially erroneous testimony by an eyewitness. Obviously, many judges assume that it is.

What effect should a comprehensible instruction have on jurors' decisionmaking? For one, it should clearly and concisely inform jurors about the appropriate legal standards to be used in a given case (e.g., that the identity of the suspect as the person who committed a crime must be proven *beyond a reasonable doubt*). Second, it should educate jurors about the circumstances under which certain identifications are likely to be reliable and when they may not be. The first step in evaluation of the effectiveness of the TELFAIRE instruction will be to determine whether these two conditions are being met—whether the TELFAIRE message can properly inform jurors about the law and educate them about circumstances that can lead to unreliable eyewitness identifications. This was the purpose of Experiment 1.

Table 1

Telfaire Instruction

One of the most important issues in this case is the identification of the defendant as the perpetrator of the crime. The Government has the burden of proving identity, beyond a reasonable doubt. It is not essential that the witness himself be free from doubt as to the correctness of the witness' statement. However, you, the jury, must be satisfied beyond a reasonable doubt of the accuracy of the identification of the defendant before you may convict. If you are not convinced beyond a reasonable doubt that the defendant was the person who committed the crime, you must find the defendant not guilty.

Identification testimony is an expression of belief or impression by the witness. Its value depends on the opportunity the witness had to observe the offender at the time of the offense and to make a reliable identification later.

In appraising the identification testimony of a witness, you should consider the following:

1. Are you convinced that the witness had the capacity and an adequate opportunity to observe the defender?

Whether the witness had an adequate opportunity to observe the offender at the time of the offense will be affected by such matters as how long or short a time was available, how far or close the witness was, how good were lighting conditions, whether the witness had had occasion to see or know the person in the past.

2. Are you satisfied that the identification made by the witness subsequent to the offense was the product of his own recollection? You may take into account both the strength of the identification and the circumstances under which the identification was made.

If the identification by the witness may have been influenced by the circumstances under which the defendant was presented to him for identification, you should scrutinize the identification with great care. You may also consider the length of time that lapsed between the occurrence of the crime and the next opportunity of the witness to see the defendant, as a factor bearing on the reliability of the identification.

B-52

3. Finally, you must consider the credibility of each identification witness in the same way as any other witness, consider whether the witness is truthful, and consider whether the witness had the capacity and opportunity to make a reliable observation of the matter covered in his testimony.

I again emphasize that the burden of proof on the prosecutor extends to every element of the crime charged, and this specifically includes the burden of proving beyond a reasonable doubt the identity of the defendant as the perpetrator of the crime with which the defendant stands charged. If after examining the testimony, you have a reasonable doubt as to the accuracy of the identification, you must find the defendant not guilty.

What if the TELFAIRE instruction proves to be less effective than jurists think it is? There was reason to believe beforehand that this would be the case because other studies have shown that jurors often misunderstand or misuse the judge's instructions (Alfini, Sales, & Elwork, 1982; Severance, Greene, & Loftus, 1984). A second step in this process was an attempt to simplify and clarify the language of the instruction so that the deficiencies uncovered in Experiment 1 could be remedied. Revisions of the cautionary message were based both upon specific errors jurors made in response to questions about the TELFAIRE instruction in Experiment 1 and upon principles of psycholinguistics that other researchers have used to simplify jury instructions (Alfini et al., 1982; Charrow & Charrow, 1979; Severance et al., 1984). The revised instruction was intended to give jurors more information about factors that influence an eyewitness' memory and to present this information in a more comprehensive way than the TELFAIRE instruction did. The purpose of Experiment 2 was to pit the TELFAIRE instruction against its revised counterpart to see if revision improves jurors' understanding of the eyewitness issue.

Revisions of the cautionary instruction will be merely academic exercises, however, unless trial judges accept and utilize them. Because the ultimate goal of this research is to develop a comprehensible instruction that can be used in court, it is extremely important to assess the legal community's reactions to this simplified message. Accordingly, Experiment 3 is a survey of 300 superior court judges in this country; its purpose was to obtain feedback on the likely acceptance of the TELFAIRE and revised instructions in courts of law today.

Experiment 1

Method

Subjects

Subjects were 127 University of Washington students who were also registered voters and thus eligible for jury duty. They participated in groups of five or more, and each deliberating jury had five to nine people. There was a total of 21 juries.

Stimulus Materials

A videotaped trial was made based upon a second-degree assault case in which the defendant allegedly threw a bottle that struck and partially blinded another person. The incident occurred in a tavern. The trial was an enactment, filmed in an actual courtroom with an ex-judge, practicing attorneys, a court reporter and bailiff, and credible actors as witnesses. The facts were intentionally balanced so the defendant was not clearly innocent or guilty. There were two versions of the trial. In both versions, no one actually saw the defendant throw the bottle, but the barmaid identified him as the person who might have done so. Three factors known to influence eyewitness identification (lighting, distance from the scene, and clarity of view) were differentially represented in the two versions. In the "strong identification" version, the barmaid testified that the defendant was seated directly below an overhead light, his table was situated near her position behind the bar, and she had an unobstructed view of him. In the "weak identification" version, she testified that the defendant was sitting in a dimly lit area of the tavern, that he was not close to the bar, and that her view of his table was partially obstructed. The trial included opening arguments, trial testimony, instructions from the judge and closing arguments. All subjects heard instructions about their general duties as jurors, the elements of the crime, and the meanings of the presumption of innocence, burden of proof, and reasonable doubt. These instructions are currently used in criminal trials in the state of Washington (*Washington Pattern Instructions,* 1977). The trial lasted approximately 1-1/2 hours.

Design and Procedure

A 2 × 2 factorial design was used. Approximately half of the subjects viewed the strong identification version of the trial; the other half viewed the weak version. For half of the subjects, the TELFAIRE instruction was

B-53

included as part of the judge's charge to the jury. The other half did not hear an instruction on eyewitness reliability. After watching the trial, subjects deliberated for up to 30 minutes. Each jury was given a copy of the judge's instructions, asked to select a foreperson, and told that their verdict must be unanimous. The deliberations were tape-recorded. Finally, each juror completed a questionnaire that pertained to information given during the trial and to the jury's deliberations.

Dependent Measures

The questionnaire first asked for the jury's verdict (guilty/not guilty/hung) and then asked subjects whether they agreed with this verdict. Subjects were asked how difficult it was for them individually to decide on a verdict, and how difficult it was for their jury to reach consensus. Additional questions asked subjects to rate the believability and importance of the testimony of all the witnesses, and whether certain pieces of evidence pointed toward the defendant's innocence or toward guilt. There were several questions that tapped the deliberation process: the clarity of certain instructions; the extent to which the group discussed the judge's instructions; and the interactions among fellow jurors. Eight multiple-choice questions about factors known to influence eyewitness identification were included. These questions covered the issues of a witness' stress level, opportunity to observe the offense and confidence level when making an identification, the presence of a weapon, the standard of proof to be met in proving identity, and the most fair way to conduct a lineup.

Results

Verdicts

Table 2 presents information on jurors' verdicts. Overall, 17 jurors, comprising two juries and 13% of the sample, found the defendant guilty; 48 jurors, comprising eight juries and 38% of the sample, found the defendant not guilty; and 62 jurors, comprising 11 juries and 49% of the sample, reported that their jury was hung. When the instruction was absent, two juries voted to convict (both groups had seen the strong version of the identification). When the identification was weak and jurors received an eyewitness instruction, they were less likely to be hung than when the evidence was weak and no instruction was given.

The presence of an instruction also made it easier for juries to reach a consensus when the evidence against the defendant was weak. On a scale

Table 2

Jurors' Verdicts With and Without Telfaire *Instruction (in percentages): Experiment 1*

	Guilty	Not guilty	Hung
No *Telfaire* instruction			
Strong identification	41.9	19.4	38.7
Weak identification	3.0	33.3	63.6
Telfaire instruction			
Strong identification	6.5	19.4	74.2
Weak identification	3.1	78.1	18.8
	13.63	37.55	48.83
	N = 17	N = 48	N = 62

from 1-7 where 1 = extremely easy for my jury to agree on a verdict and 7 = extremely difficult to agree on a verdict, the mean for jurors with weak evidence and TELFAIRE = 3.38; with weak evidence and no instruction = 4.78; with strong evidence and TELFAIRE = 4.81; with strong evidence and no instruction = 3.74. Overall Version × Instruction interaction: $F(1,122) = 10.90, p = .001$. Post-hoc Newman-Keuls tests showed that jurors with weak evidence and TELFAIRE differed significantly from both those with weak evidence and no instruction and those with strong evidence and TELFAIRE.

An important statistical issue is raised by these data. Because subjects deliberated in groups, it can be argued that their post-deliberation responses are not independent and that analyses should use groups (nested within conditions) as the appropriate error term for significance tests. In fact, some recent jury simulation studies have used groups-within-conditions analyses (e.g., Tanford & Penrod, 1982), although many employ analyses based on individual subjects (e.g., Cowan, Thompson, & Ellsworth, 1984; Hastie, Penrod, & Pennington, 1984). The obvious difficulty involved in groups-within-conditions analyses is one of statistical power. Several hundred jurors must often be run before the appropriate significance tests can be performed (Tanford and Penrod, for example, had 732 subjects in

their experiment). Although jurors could have been influenced in different ways during the course of deliberating, the decision was made here to use subjects-within-conditions as the error term in data analyses. There is considerable precedent for this technique, and it lends obvious simplification to data collection, although its limitation in assessing the effects of juror interaction should be recognized.

Comprehension Questions

The questionnaire included eight multiple-choice questions designed to tap jurors' general understanding of eyewitness testimony. In theory, if the TELFAIRE instruction is effective in educating jurors about factors to consider when hearing eyewitness evidence, then subjects who were given the instruction should correctly answer more questions than jurors without the instruction. In practice, however, the TELFAIRE instruction had no effect on performance, as there was no significant difference in the number of questions answered correctly by different groups of jurors (of 8 questions, the mean number correct for the pattern instruction group was 4.78; for the no instruction group, 4.61, $F < 1$). An item analysis confirmed this result: none of the eight questions was more often answered correctly by jurors who heard the TELFAIRE message.

Weight of the Evidence

One question asked jurors to rate the importance of the eyewitness testimony. An effective instruction could conceivably affect the weight that jurors place on the witness' identification. If an instruction clearly informed jurors about factors that make certain identifications unreliable, then jurors might place more weight on identifications made under good conditions and less weight on identifications made under adverse conditions. Contrary to this hypothesis, however, there were no significant differences in rated importance of the eyewitness as a function of the version of the trial that subjects saw or whether they heard the eyewitness instruction. Possibly, a ceiling effect occurred here, as the mean rating for the importance of the eyewitness in influencing verdict was 6.09 on a scale from 1 to 7 where 1 = Not at all and 7 = Extremely important. To many jurors, the eyewitness was clearly the most important person to take the stand.

Another question designed to measure the weight subjects gave to the eyewitness' testimony asked them how believable her account seemed. They answered on a scale from 1-7 where 1 = Not at all believable and 7 = Very believable. There was no effect on rated believability of the version of the trial that subjects saw. They did rate the eyewitness as slightly less believable

if they were given the TELFAIRE instruction, although this was a very marginal effect. The mean rating for jurors who heard a cautionary instruction = 5.35; without an instruction = 5.70; $F(1,123) = 2.52, p < .10$.

Finally, jurors were asked to select the witness who gave the single most important piece of testimony. With TELFAIRE and weak evidence, 52% of subjects chose the eyewitness. With TELFAIRE and strong evidence, 68% said the eyewitness gave the most important testimony. The instruction seems moderately effective at informing jurors of the factors that can make an identification difficult: When those factors were present and the instruction given, fewer jurors chose the eyewitness' testimony as most important. On the other hand, when the instruction was absent, 85% of jurors with weak evidence and 39% of jurors with strong evidence chose the eyewitness. This counter-intuitive finding resulted because without an instruction, jurors with strong evidence were more likely than jurors with weak evidence to select an "earwitness" or the defendant as the person with the most important testimony. Relative to these subjects, jurors who heard the instruction responded in a more predictable way.

Effectiveness of Judge's Instructions

Jurors were asked whether the judge's instructions were effective in helping them understand the law (on a scale from 1-5 where 1 = Not at all effective and 5 = Extremely effective). The mean score overall was 3.51. There were no significant main effects. There was a significant Version × Instruction interaction, $F(1,122) = 3.93$, $p < .05$. Post-hoc tests indicated that jurors who saw the weak identification *with* the TELFAIRE instruction (mean = 3.68) and jurors who saw the strong identification *without* the instruction (mean = 3.65) gave the highest ratings. The mean for jurors who saw the weak identification *without* the instruction = 3.33; and for the strong identification *with* the instruction, 3.39.

Discussion

The results of Experiment 1 are mixed on the effectiveness of the eyewitness instruction that is currently being used in many jurisdictions. Jurors who heard this instruction were no more sensitive to factors known to be problematic to eyewitness identification than were jurors who had no instruction. This is not a surprising result because the TELFAIRE instruction made no mention of how stress or the presence of a weapon can affect eyewitness memory, nor did the instruction refer to standards of proof or lineup fairness. Yet it is just this increase in semantic knowledge that

B-55

would most clearly signal an improvement in jurors' understanding of the eyewitness issue.

However, an eyewitness instruction can be effective in a second, unrelated way: It can focus jurors' attention on the identification so they can scrutinize the conditions under which it was made. The weight attributed to the eyewitness' testimony might be affected as a result. Data from Experiment 1 are equivocal on this point. Subjects did not rate the eyewitness' testimony as more (or less) important when the TELFAIRE message was given, but when they heard that instruction in combination with a weak identification, fewer selected the eyewitness as the person who gave the most important testimony. When the instruction was presented in combination with a strong identification, more jurors chose the eyewitness as having the most important testimony. In at least one respect then, the TELFAIRE instruction is having the effect of cautioning jurors about an unreliable identification, and simultaneously bolstering their belief in an identification made under more optimal conditions.

Experiment 2

There were two basic concerns that directed revision of the TELFAIRE instruction. One was to simplify the language of the message in order to make it clear and logical to jurors. The second was to integrate findings from research on eyewitness testimony into the instruction. But there is a fine line to tread here. Because many trial court decisions have been overturned because of incorrect jury instructions (Nieland, 1979), it would be foolish to alter the existing instruction too drastically. Therefore, throughout the process of revision, care was taken to ensure that the revised instruction bear some similarity to the pattern instruction, but that it impart more information and do so logically and coherently.

Several principles of psycholinguistics tested in other studies (e.g., Charrow & Charrow, 1979; Elwork, Sales, & Alfini, 1977; Severance et al., 1984) were used to simplify the wording of the TELFAIRE instruction. Changes were made both in terms of specific grammatical constructions and in overall organization:

1. Legal terminology and uncommon words were eliminated.
2. Passive constructions, embedded phrases and compound sentences were eliminated.
3. Changes in organization were made to insure a logical structure within each paragraph and between paragraphs. "Hierarchical structuring" as described by Myers and Jones

(1979) was used. In a hierarchical structure, general high-level concepts are broken down into their component parts and perhaps reintegrated.

Some changes in the instruction were prompted by a desire to inform jurors about research on eyewitness testimony, particularly findings that may go against common-sense notions. The specific additions to be made were determined in part by errors subjects made on the comprehension questions in Experiment 1. They did not know, for example, that the relationship between a witness' confidence and the accuracy of his or her identification can be weak (for a review of this issue, see Wells & Murray, 1984). Unless jurors are informed that even an extremely confident witness might still be wrong, the cautionary instruction has missed its mark, no matter how well it is comprehended. Other added information concerned a witness' stress at the time of the crime, the fairness of a lineup, and the fact that identification testimony is an expression of the eyewitness' belief. Recent research shows that high levels of stress interfere with a person's ability to accurately recall an event (see Deffenbacher, 1983, for a review of the relationship between stress and memory), and that certain lineup and identification techniques can bias witnesses and increase the likelihood of a mistaken identification (see Malpass & Devine, 1984). The full text of the revised instruction is given in Table 3.

The major purpose of Experiment 2 was to evaluate jurors' comprehension of the revised instruction relative to the pattern TELFAIRE message and to do so in an ecologically valid way. Subjects in this experiment were people currently serving on jury duty who watched the videotaped trial in an actual courtroom and retired to the jury room to deliberate. These more realistic procedures provided a test of the usefulness of the revisions for actual jurors.

A revised instruction that effectively communicates information about eyewitnessing should affect jurors in several ways. First, their understanding or semantic knowledge of the factors to consider when weighing eyewitness testimony should increase. Second, a comprehensible instruction should help jurors more easily agree on a verdict. With a clear instruction, they might realize when an eyewitness is likely to be accurate and, barring confounding evidence, return a guilty verdict. Jurors would also be made aware of the conditions under which an accurate identification is unlikely and, in these circumstances, decide to acquit. Third, a revised instruction might change the weight that jurors attribute to certain witnesses or pieces of testimony. For example, they may put less weight on the eyewitness' testimony if an instruction effectively conveys the types of circumstances

B-56

Table 3

Revised Instruction

One of the major issues in this case is the identification of the defendant as the person accused of committing a crime. The prosecution has the burden of proving beyond a reasonable doubt, not only that a crime was committed, but that the defendant was the person who committed it.

Identification testimony is an expression of belief by an eyewitness about a person who may have committed a crime. You should keep in mind that identifying a person who committed a crime may be very different from recognizing a friend you see repeatedly. In evaluating the testimony of any eyewitness, you should consider two sets of factors: first, factors present when the incident occurred; second, factors affecting the later identification.

Factors present when the incident occurred are:
a) how much time was available for observation;
b) how well the scene was lit;
c) how far the eyewitness was from the incident.

You should also consider how well the eyewitness could see and hear at the time. For example, if a witness is afraid or distracted, his or her capacity to perceive and remember is reduced.

A second set of factors affect later identification. You should consider how much time passed between the incident and the identification. For example, identification errors increase as time passes. You should also consider the circumstances surrounding the identification. For example, an identification made from a fair lineup of similar individuals is more reliable than other forms of identification such as viewing a suspect alone. You should also consider how certain the eyewitness was in making an identification. Certainty may or may not mean that the identification is accurate.

If, after considering all of these factors, you have a reasonable doubt about the accuracy of the eyewitness' identification of the defendant as the person who may have committed a crime, then you must find the defendant not guilty.

that could distort memory, and these circumstances are present in the case being tried.

Method

Subjects

The Court Administrator and Superior Court Judges for King County, Washington, agreed to let us enlist persons who were currently serving as jurors to be subjects in this experiment. During the two weeks they serve, jurors typically spend time waiting to be assigned to a courtroom. We solicited subjects from among those who were waiting, with the understanding that their participation was voluntary and that they would be excused from the experiment if they were called for jury selection. In all, 139 people volunteered to take part in this experiment. There were five to nine people on each jury, and a total of 24 juries.

Stimulus Materials

Subjects watched one of the two versions of the videotaped trial that had been used in Experiment 1. All jurors heard instructions defining their general duties, the elements of second degree assault, and the presumption of innocence/burden of proof/reasonable doubt. The eyewitness instruction was either the pattern TELFAIRE instruction or the revised instruction.

Design and Procedure

A 2×3 factorial design was used. Approximately half of the subjects saw the strong identification version of the trial and half saw the weak identification. One third of the jurors heard the TELFAIRE instruction as part of the judge's charge to the jury; one third heard the revised instruction; and one third had no instruction on eyewitness reliability.

Before deliberations began, subjects were given a pre-deliberation questionnaire that asked for their individual verdicts. They then deliberated for 30 minutes and completed a post-deliberation questionnaire.

Dependent Measures

To measure the effects of the eyewitness instruction on jurors before they discussed the case, we asked subjects to complete a questionnaire before they deliberated. They were asked if they personally thought the defendant was guilty or not guilty, and how certain they were of their verdict. The

B-57

post-deliberation questionnaire was similar, but not identical, to the questionnaire used in Experiment 1. Jurors gave their group verdict, then told the extent to which they agreed with that decision, and how difficult it was personally, and as a group, to reach a verdict. Several questions related to the believability of witnesses and importance of their testimony. Questions relating to the deliberation process were slightly modified. Jurors were asked whether each of the judge's instructions was clear to them as they first heard it, and whether there was discussion of any of these instructions during deliberation. Finally, they were asked six multiple-choice questions pertaining to eyewitness identification (these questions had been used in Experiment 1), and several demographic questions.

Results

Pre-deliberation Questionnaire

When questioned before deliberating, 41.2% of jurors thought the defendant was guilty and 58.8% thought he was not guilty. Subjects who had the revised instruction were less likely to convict than jurors in other conditions (22.5% guilty verdicts with the revised instruction vs. 53.1% with the TELFAIRE instruction and 44.7% with no eyewitness instruction; $\chi^2 = 8.86$, $2df$, $p < .01$). There was also a significant effect of the version of the identification: Jurors who viewed the strong identification were more likely to convict than jurors who saw the weak identification version (50% guilty verdicts for strong identification vs. 33.8% for weak identification; $\chi^2 = 3.66$, $1df$, $p < .05$).

Post-deliberation Questionnaire

Verdicts. Overall, 2 juries (11% of the subjects) returned a guilty verdict after deliberations, 12 juries (47% of all subjects) said not guilty, and 10 juries (42% of the subjects) were hung. The version of the identification had an effect: Not one jury convicted the defendant in the weak identification condition whereas two juries (24% of the subjects in this condition) convicted the defendant when the identification was strong.

It was predicted that a comprehensible instruction should result in more convictions in the strong identification version and more acquittals in the weak identification condition, in comparison to verdicts from jurors with a pattern instruction or no cautionary message. The instruction manipulation did have an effect, but not quite as predicted: Fewer juries convicted the defendant when they heard the revised instruction. These data are shown in Table 4.

Table 4

Verdicts After Deliberation (in percentages): Experiment 2

	Guilty	Not guilty	Hung
Telfaire instruction ($N = 49$)			
Weak identification	0	41.4	58.6
Strong identification	35.0	35.0	30.0
Revised instruction ($N = 41$)			
Weak identification	0	73.2	26.8
Strong identification	0	75.0	25.0
No cautionary instruction ($N = 49$)			
Weak identification	0	42.3	57.7
Strong identification	34.8	21.7	43.5
	$\overline{X} = 10.8$	$\overline{X} = 46.8$	$\overline{X} = 42.4$
	$N = 15$	$N = 65$	$N = 59$

Subjects were asked how difficult it was for their jury to reach a consensus. It was easier for people who saw the strong identification to reach a verdict. On a scale from 1-7 where 1= extremely difficult and 7 = extremely easy, the mean for jurors who saw the weak identification = 2.65 and for jurors with the strong identification, 3.54; $F(1,139) = 5.08, p = .025$. It was slightly easier for jurors with the revised instruction to agree on a verdict than jurors with no cautionary instruction (mean for revised group = 3.63; for pattern group = 3.16; for no instruction group = 2.79; $F(2,130) = 2.39, p = .08$).

Comprehension Questions

This questionnaire included six of the multiple-choice questions used in Experiment 1. Jurors who heard the revised instruction answered more questions correctly than did jurors in the other two conditions, although the

effect was marginal. The mean number correct with revised instruction = 3.33, with pattern = 2.80, and with no instruction= 2.66; $F(2,130) = 2.51, p = .07$. Post hoc t-tests showed that the revised instruction group differed significantly from the group with no instructions. An item analysis shows that for four of the six questions, the highest percentage of correct responses came from jurors with the revised instruction. In three of the four instances, Z-tests confirmed that the proportion of correct responses was higher for jurors with revised instructions than jurors who heard the pattern instruction ($p < .05$).

Weight of the Evidence

Subjects were asked to select the one witness who gave the most important testimony. It was predicted that compared to the pattern instruction, a revised instruction might cause jurors to put less weight on the eyewitness' account when the identification seemed unreliable and more weight on her story when the identification was strong. When the identification was weak, fewer jurors with the revised instruction than with the pattern instruction selected the eyewitness as the person with the most important testimony ($p < .05$ by Z-test). When the identification was strong, equal proportions of subjects in the revised-instruction juries and in the pattern-instruction juries selected the barmaid's testimony as the most important. A complementary result came from data on the importance of the defendant's testimony: When the identification was weak, more jurors with the revised instruction than with the pattern instruction selected the defendant as the witness who gave the most important testimony ($p < .05$ by Z-test). These percentages are given in Table 5.

Jurors also were asked how believable certain witnesses were. When the revised instruction was given, the believability of the defendant rose. On a scale from 1-7 where 1 = Extremely unbelievable and 7 = Extremely believable, the mean for jurors with the pattern instruction = 3.32, with the revised = 4.34, and with no cautionary instruction = 3.53; $F(2,129) = 4.32$, $p = .015$. Post-hoc tests showed that the revised group differed significantly from the other two ($p < .05$). The defendant's testimony was rated as slightly more believable when the revised instruction was combined with a weak identification (mean = 4.57) rather than with the strong identification (mean = 4.10), although this effect was only marginally significant, $p < .10$.

Finally, jurors were asked whether certain evidence pointed more toward the defendant's innocence or toward his guilt. One factor that jurors evaluated was the testimony of the barmaid that she believed, but was not certain, that the defendant threw the bottle. This is an important question because the pattern and revised instruction gave quite different information

Table 5

Percentage of Jurors Who Selected the Eyewitness or Defendant as the Most Important Witness: Experiment 2

	Eyewitness	Defendant
Telfaire instruction ($N = 49$)		
Weak identification	76	3
Strong identification	60	20
Revised instruction ($N = 41$)		
Weak identification	52	19
Strong identification	60	20
No cautionary instruction ($N = 49$)		
Weak identification	25	13
Strong identification	54	5

about the relationship between a witness' confidence and the accuracy of her identification. (Recall that the revised instruction suggested that confidence does not imply accuracy.) Because jurors who heard the pattern instruction had no reason to discount the barmaid, they may rate her testimony as indicative of guilt, whereas jurors with the revised instruction may see her uncertainty as indicative of innocence. Using a scale from 1-7 where 1 = Strong indication of defendant's innocence and 7 = Strong indication of guilt, subjects' mean score with pattern instruction = 2.91, with revised = 2.34, and with no instruction = 3.19; $F(2,129) = 2.99, p < .05$. Jurors who heard the revised instruction were less likely to indicate that the barmaid's testimony implied guilt.

Effectiveness of Judge's Instructions

Jurors were asked to rate how much they discussed the judge's instructions on a scale from 1-5 where 1 = No discussion and 5 = Thorough discussion. Jurors who heard the revised eyewitness instruction spent less

time discussing the judge's instructions than did jurors in the other conditions (mean for revised instruction = 2.37, for pattern = 2.85, and for no instruction group = 2.91; $F(2,129) = 4.65, p < .01$). Post-hoc tests confirmed that the mean score from jurors with the revised instruction was significantly less than the other two means. A related finding was that jurors who heard the revised instruction were less likely to use a printed packet of the judge's instructions left for them during deliberations. Only 31.6% of jurors in the revised instruction condition reported that someone on their jury referred to the copy of the instructions, whereas 88.9% of jurors with the pattern instruction and 70.8% with no eyewitness instruction indicated that someone on their jury used this packet (Z-test: pattern vs. revised = $4.09, p < .05$).

Discussion

Data from Experiment 2 show that a simplified cautionary instruction can affect jury decisionmaking in several ways. Jurors who heard the revised instruction were more apt to be correct in their answers to several multiple-choice questions than jurors who heard the pattern instruction or no cautionary message. These questions tapped jurors' comprehension of significant factors to consider when evaluating eyewitness accounts. Although this result is not extremely powerful, it suggests that the simplified instruction is somewhat more effective than the pattern instruction in imparting information and in clarifying misconceptions about eyewitness testimony.

In terms of jurors' final verdicts, the revised instruction led to a higher percentage of not-guilty verdicts than did the TELFAIRE instruction. This was true for polls taken both before and after deliberation. It was predicted that a simplified instruction should have this effect only in combination with a weak identification, that is, an effective instruction should alert jurors to the fact that the witnessing conditions were very poor and should result in more acquittals. In this experiment, the revised instruction led to more acquittals, regardless of the version of the identification that jurors saw. There are two possible explanations for this result. One is simply that the revised instruction is strongly biased in favor of the defendant and calls into question the reliability of all eyewitness testimony. A second explanation has to do with the notion of a witness' certainty and the possibility that that could mediate verdicts. Recall that the revised instruction informed jurors about the minimal correlation between certainty and accuracy, something that was absent from the pattern instruction. Even after hearing this, subjects took the witness' certainty into account. Jurors with the revised instruction were somewhat more convinced than other jurors that the

barmaid's uncertainty about her identification might imply innocence. Perhaps then, this additional information about certainty caused jurors not to disregard the accuracy/confidence relationship entirely, but, instead, made jurors with the revised instruction more apt to acquit.

The revised instruction influenced the weight that jurors attributed to the testimony of some witnesses. When asked to select the witness with the most important testimony, fewer subjects who had the revised instruction selected the eyewitness barmaid and more chose the defendant, compared to jurors with the pattern instruction. This result fits with the fact that jurors in this condition were more likely to find the defendant not guilty. A collateral finding is that jurors who heard the revised instruction were more apt to believe the defendant (especially with the weak identification) than were jurors in other conditions.

Jurors who heard the revised instruction said they spent less time discussing the judge's instructions and were less apt to refer to a copy of the instructions during their deliberations. Perhaps these jurors understood their eyewitness instruction better than did jurors with the pattern instruction, and because eyewitness identification was the central issue in the case, they had less need to discuss the instructions in total. Conceivably, a comprehensible instruction could shorten deliberation time.

Experiment 3

No matter how plainly stated an instruction might be, unless it is approved by trial judges, it might never be used in court. Because ultimately the trial judge decides which instructions the jury will hear, a sample of these judges was asked what they thought of the pattern TELFAIRE instruction and its revised counterpart. In addition, judges were asked how they usually deal with the issue of eyewitness testimony: do they routinely deliver a cautionary instruction? Do they give the TELFAIRE instruction, their own instruction, or one provided by an attorney? Do they allow an expert to testify about eyewitness reliability? Information from judges across the country can provide a representative picture of the ways this issue is handled in trial courts.

Method

Subjects

From a list of the names and addresses of all superior court judges in the U.S., five judges were randomly selected from each state, the District of Columbia and Puerto Rico, and each received a questionnaire. In addition,

B-60

because of their higher likelihood of responding, all 40 local superior court judges were sent a questionnaire. In all, 300 subjects received questionnaires. Half were asked to evaluate the TELFAIRE instruction, half the revised instruction. Responses were received from 102 judges, or 34% of the sample. Of these, 89 were completed questionnaires. The remaining judges gave their comments only, or indicated that they tried civil cases exclusively. the TELFAIRE instruction was evaluated by 52 judges (58.4%); the revised instruction by 37 judges (41.5%). These percentages are not significantly different.

Questionnaire

All judges were asked five questions about their instruction:

1. How effective do you think the instruction will be in conveying the intended legal concepts to the jury? (On a scale from 0-4 where 0 = Not at all effective and 4 = Extremely effective).
2. What, if any, aspects of the instruction do you find particularly helpful? What aspects are troublesome?
3. Overall, would you consider this a proper instruction to give to the jury? (Yes, No, Uncertain).
4. Do you perceive bias favoring either party in the language of the instruction? (On a scale from –3 to +3 where –3 = Strong bias in favor of defense, 0 = No bias, +3 = Strong bias in favor of prosecution.)
5. Any additional comments about the instruction.

Finally, judges were asked several questions about the procedures they typically follow in eyewitness cases.

Results and Discussion

When judges were asked "How effective do you think this instruction will be in conveying the intended legal concepts to the jury?", the revised instruction was rated as more effective than the pattern TELFAIRE version. Of the judges who had the revised instruction, 62.5% thought it was very effective or extremely effective; only 23.4% of judges who read the pattern instruction thought it was that effective. The overall responses are given in Table 6 ($\chi^2 = 16.34$, 4*df*, $p < .005$). This may indicate that the simplified language and more coherent organization can better communicate information to jurors. At any rate, judges felt that legal

Table 6

Responses to Question "How Effective Is the Instruction at Conveying the Intended Legal Concepts to the Jury?" Experiment 3

	Telfaire instruction	Revised instruction
Not at all effective	27.7	9.4
Slightly effective	19.1	9.4
Moderately effective	29.8	18.8
Very effective	19.1	28.1
Extremely effective	4.3	34.4

Numbers given are percentages

(N = 79)

concepts were being more effectively expressed by the revised language than by the pattern instruction.

Responses to the question about whether the instruction was a proper one for the jury illustrate the controversy over this instruction. Most judges thought that neither instruction was proper to give to the jury (78% of the judges with the pattern instruction and 53% with the revised instruction gave this response), although more judges said "Yes" to the revised instruction than to the TELFAIRE instruction (36.6% vs. 12.0%, overall $\chi^2 = 7.59, 2df, p < .05$). This general disapproval may reflect the fact that many states (including Washington) disallow cautionary instructions, so even a comprehensible message would be met with disfavor.

The final scaled question asked about bias favoring either party in the language of the instruction. Not surprisingly, judges thought both instructions were biased in favor of the defense. A more interesting question then becomes "How biased are these instructions?" The revised instruction was thought to be more biased than the TELFAIRE version. Of the judges who read the revised instruction, 61.3% rated it as carrying a "strong defense bias," whereas only 34% of judges who had the TELFAIRE instruction gave it this rating (overall $\chi^2 = 8.78, 3df, p < .05$). This result might have been due to the fact that the revised instruction included specific examples of the ways that identifications can go awry, for example, that if a

witness is afraid or distracted, his or her capacity to perceive and remember is reduced.

Some of the most interesting data came from judge's written comments about the instructions. There was very little difference in the general character of their remarks depending upon which instruction they reviewed. Judges liked the fact that the instruction explained the burden of proof and reasonable doubt criterion with regard to the identification issue. They generally approved of the statements that told jurors about the factors to consider in assessing eyewitness reliability. What they questioned, however, was the fact that the instructions seemed to comment on the evidence, noting that jurors alone, using their judgement and common sense, are to decide what weight an eyewitness' testimony will carry.

What about jurors' common-sense notions? Is it true, as some judges suggested, that the instruction simply tells jurors what they already know. Recent studies of common beliefs about eyewitness accounts (Brigham & Bothwell, 1983; Loftus, 1979) have concluded that knowledge about the likely accuracy of eyewitnesses is not within the common knowledge of most people.

Could it be that judges overestimate people's knowledge in their suggestion that the information in cautionary instructions is commonsensical? The Loftus and Brigham and Bothwell studies suggest this. It may be that judges themselves are unaware of the fallibility of eyewitness evidence. Consequently, they might not recognize the need to educate jurors and certainly would not be motivated to give instructions which stress caution in evaluating eyewitness evidence. This unfamiliarity with the relevant research may have been at the root of some of the superior court judges comments about the cautionary instructions.

What are the standard procedures for judges in trials that involve eyewitness testimony? Judges were first asked to estimate the percentage of trials over which they preside where eyewitness identification is a major issue. Their mean response was 26% and median response, 17.2%. In response to the question about whether they use instructions they wrote themselves, those offered by a defense attorney, or the pattern TELFAIRE instructions, judges' answers indicated that the three types of instructions have been used by approximately equal numbers of judges (17.4%, 23.5%, and 23.5% of the sample, respectively); although when an instruction is given, it is most likely to be a version written by the judge (45% of the sample gave this response, 23% said they would be most likely to use an attorney's version, and 32% said they would be most likely to use the TELFAIRE instruction). By contrast, expert testimony is used by many fewer judges and in only a small percentage of eyewitness cases.

General Discussion

The purpose of these studies was to evaluate jurors' comprehension of the standardized pattern instruction about eyewitness reliability and to make revisions where the instruction seemed lacking. The simplified language was intended to be easier for jurors to understand and to increase their knowledge of the factors to consider when evaluating eyewitness testimony.

The results of these experiments suggest that the pattern TELFAIRE message is not especially effective at educating jurors about the vagaries of eyewitness memory. The revised instruction did better on this count: Jurors who heard it were more knowledgeable about the factors to consider when listening to an eyewitness testify. Although several studies have shown that people have certain misconceptions about eyewitness reliability, this finding indicates that the revised instruction is indeed telling jurors something they do not already know. Finally, judges thought that the revised instruction was more effective at conveying the intended legal concepts to the jury than was the pattern instruction. At least in terms of its comprehensibility and the ease with which it communicates information to jurors, the revised instruction seems the preferred one.

The revised instruction was also intended to make it easier for jurors to discriminate between reliable and unreliable identifications—in a sense, to know whom to believe. The revised instruction seemed less able to teach discrimination than to teach doubt. Jurors who heard it were less likely to convict the defendant, regardless of whether the identification was weak or strong. Conceivably, a change in the wording of the instruction may have shifted subjects' decision criteria away from the existing standard. As a consequence, jurors who heard the revised instruction were less apt to believe the testimony of the eyewitness and more apt to believe the defendant's story. A corollary result is that judges who evaluated these instructions thought the revised version was more biased in favor of the defense than was the pattern instruction.

To what extent can these results be generalized to other persons and situations? Because the experimental situation used to test jurors' understanding did not mirror exactly the conditions of a real trial, one must be cautious about overgeneralization. Still, there is room for cautious optimism. The setting used in Experiment 2 closely resembled that of a real trial, and actual jurors served as subjects. Also, judges' evaluations provided an additional test of the usefulness of this research.

This work has illustrated the difficulties jurors have in knowing how to evaluate eyewitness testimony when they are given no instruction or the customary TELFAIRE admonition on eyewitness reliability. It has also moved us one step closer to development of a clear and complete eyewitness instruction. Perhaps an instruction that is somewhat less biased but equally clear and complete will be the next refinement. Jurors would then know how to evaluate an eyewitness account without doubting much of what the witness says.

B-62

References

Alfini, J. J., Sales, B. D., & Elwork, A. (1982). *Making jury instructions understandable.* New York: Michie/Bobbs-Merrill.

Brigham, J. C. (1983). Psychological factors in eyewitness identification. *Journal of Criminal Justice,* **11,** 47–56.

Brigham, J. C., & Bothwell, R. (1983). Prospective jurors' ability to predict eyewitness identification accuracy. *Law and Human Behavior,* **7,** 19–30.

Cavoukian, A. (1980). *Eyewitness testimony: The ineffectiveness of discrediting information.* Paper presented at the American Psychological Association Meeting, Montreal.

Charrow, R. P., & Charrow, V. R. (1979). Making legal language understandable: A psycholinguistic study of jury instructions. *Columbia Law Review,* **79,** 1306–1374.

Cowan, C., Thompson, W., & Ellsworth, P. C. (1984). The effects of death qualification on jurors' predisposition to convict and on the quality of deliberation. *Law and Human Behavior,* **8,** 53–80.

Deffenbacher, K. A. (1983). The influence of arousal on reliability of testimony. In B. R. Clifford & S. Lloyd-Bostock (Eds.), *Evaluating witness evidence* (pp. 235–251). Chicester, England: Wiley.

Elwork, A., Sales, B. D., & Alfini, J. J. (1977). Juridic decisions: In ignorance of the law or in light of it? *Law and Human Behavior,* **1,** 163–190.

Hastie, R., Penrod, S. D., & Pennington, N. (1984). *Inside the jury.* Cambridge, MA: Harvard Press.

Holtshouser, S. E. (1983). Eyewitness identification testimony and the need for cautionary jury instructions in criminal cases. *Washington University Law Quarterly,* **60,** 1387–1433.

Katzev, R. D., & Wishart, S. S. (1985). *The impact of judicial commentary concerning eyewitness identifications on jury decisionmaking. Journal of Criminal Law & Criminology,* **76,** 733–745.

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

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

Myers, E. R., & Jones, C. (1979). Language and jury instructions. *In the Jury,* Law Reform Commission of Canada.

Nieland, R. G. (1979). Assessing the impact of pattern jury instructions. *Judicature,* **62,** 185.

Saunders, D. M., Vidmar, N., & Hewitt, E. C. (1983). Eyewitness testimony and the discrediting effect. In B. R. Clifford & S. M. A. Lloyd-Bostock (Eds.), *Evaluating witness evidence* (pp. 57–78). Chicester, England: Wiley.

Severance, L. J., Greene, E., & Loftus, E.F. (1984). Towards criminal jury instructions that jurors can understand. *Journal of Criminal Law and Criminology,* **75,** 198–233.

Tanford, S., & Penrod, S. (1982). Biases in trials involving defendants charged with multiple offenses. *Journal of Applied Social Psychology,* **12,** 453–480.

United States v. TELFAIRE, 469 F.2d 552 (D. C. Cir. 1972)

Washington Pattern Instructions—Criminal. (1977). St. Paul: West Publishing Co.

Wells, G. L., Lindsay, R. C. L., & Tousignant, J. P. (1980). Effects of expert psychological advice on juror judgments in eyewitness testimony. *Law and Human Behavior,* **4,** 275–286.

Wells, G. L., & Loftus, E. F. (Eds.). (1984). *Eyewitness testimony: Psychological perspectives.* Cambridge: Cambridge University Press.

Wells, G. L., & Murray, D. M. (1984). Eyewitness confidence. In G. L. Wells & E. F. Loftus (Eds.), *eyewitness testimony: Psychological perspectives* (pp. 155–170). Cambridge: Cambridge University Press.

Woocher, F. D. (1977). Did your eyes deceive you? Expert psychological testimony on the unreliability of eyewitness identification. *Stanford Law Review,* **29,** 969–1030.

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

B-63

Beyond Fear Appeals: Negative and Positive Persuasive Appeals to Health and Self-Esteem

MARGARET R. ROBBERSON AND RONALD W. ROGERS[1]

University of Alabama

The present research was designed to extend our knowledge of fear appeals and attitude change in two directions. The first purpose was to compare a traditional negative, or fear, appeal to its counterpart, an appeal emphasizing the positive consequences of accepting a communicator's recommendation. The second purpose was to test the assumption of the theory of protection motivation that people are motivated to protect themselves from psychological and social, as well as physical, threats. The results disclosed that, compared to a no-message control condition, the traditional negative appeal to health was more persuasive than the positive appeal. A positive appeal to self-esteem, however, was superior to a negative appeal to self-esteem. Several implications for health psychology were discussed. First, the traditional negative appeals to health may be more persuasive than positive appeals. Second, people may be persuaded to adopt healthy lifestyles for reasons other than health per se (i.e., to enhance self-esteem). The following tentative conclusions were based on unexpected findings and should be interpreted with caution. The decision to use a positive or negative appeal should consider the number of available adaptive responses. Finally, the data suggested that mass media health campaigns should use both negative and positive appeals.

A revolution in health care occurred at the turn of the century. Public sanitation and mass immunization drastically reduced the morbidity and mortality caused by infectious diseases, which had been the leading causes of death in the late 1800s and early 1900s. The infectious diseases, such as influenza and tuberculosis, are no longer the leading killers. Today, one-half of all deaths in the United States are due to unhealthy behavior or lifestyle (Center for Disease Control, 1979). If the last revolution in health care was the battle against infectious diseases, the second revolution will be the effort to prevent disease and promote health (Michael, 1982). The health care community has realized that the next major step in improving the general health of people in developed countries will be more behavioral than

[1]Requests for reprints should be sent to Ronald W. Rogers, Department of Psychology, The University of Alabama, Box 2968, Tuscaloosa, AL 35486.

Journal of Applied Social Psychology, 1988, **18,** 3, pp. 277–287.
Copyright © 1988 by V. H. Winston & Sons, Inc. All rights reserved.

JUDGES' CAUTIONARY INSTRUCTIONS ON EYEWITNESS TESTIMONY

Gabriella Ramirez, Dennis Zemba and R. Edward Geiselman, Ph.D.

There is little indication that judges' pattern instructions concerning eyewitness testimony sensitize jurors to relevant witnessing conditions. Two experiments were conducted to evaluate the pattern Telfaire *instructions versus a revised set of instructions drafted from typical eyewitness expert testimony. In Experiment 1,* Telfaire *reduced subject-jurors' sensitivity to eyewitness evidence, producing skepticism or overbelief depending on the timing of their presentation during the trial. In Experiment 2, the revised instructions preserved juror sensitivity to the evidence and significantly affected jurors' knowledge of eyewitness factors. The jurors were able to recall more of the revised instructions and rated them as being more effective and understandable, as well as more likely to affect their verdicts. Implications for assuring juror awareness of eyewitness issues are discussed.*

It has been estimated that each year approximately 77,000 individuals are arrested in the United States as suspects in cases where the principal evidence against the accused is testimony from eyewitnesses (1). The Rand Corporation (2) reported that the principal determinant of whether or not a criminal case is solved is the completeness and accuracy of eyewitness accounts. Agents from both law enforcement (3) and the defense (4) make similar claims about the importance of eyewitness testimony. Given the potential for eyewitness evidence to determine the outcome of a criminal proceeding, it is critical that the testimony be obtained and evaluated carefully. In a study of 500 cases where a felony conviction was overturned because of clear and convincing new evidence, 60 percent of the cases hinged on eyewitness identifications (5). In a 1986 survey of judges,

Copyright 1996 *American Journal of Forensic Psychology*, Volume 14, Issue 1. The *Journal* is a publication of the American College of Forensic Psychology, P.O. Box 5870, Balboa Island, California 92662.

prosecutors, and police, 8 out of 10 respondents concluded that, in their experience, witness error was the most common source of false conviction (5). Wells (6) has documented 1,000 additional such cases since 1986.

A number of psycholegal approaches have been proposed in response to the problem of fallible eyewitness memory. Some social scientists have emphasized the value of early intervention during the investigations phase of the process. This intervention would take the form of better interview techniques for use by police to help eyewitnesses provide more complete and accurate reports (7). Another "solution" often introduced by the prosecution and some law-enforcement personnel is to deny the problem and do nothing. It follows from this conception that a unique and meaningful event such as an encounter with a criminal must have been etched permanently into the witness's memory, resulting in near-perfect recollection. It has been our experience, however, that many law-enforcement professionals readily acknowledge the problems and limitations of eyewitness reports, although they may not fully understand the factors involved (8). Likewise, an apparent motivation for the prosecution's position to deny the problem is to avoid the appearance of any doubt about the veracity of favorable eyewitnesses.

Some social scientists, while acknowledging the existence of the problem of fallible eyewitness recollection, concur that nothing is to be gained at present by a mixture of social scientists, counsels, and juries in the courtroom (9-11). In contrast, legal scholar and social scientist Hugo Munsterberg (12) acknowledged the fallibility of eyewitness testimony at the turn of the century. He advanced the argument that "experimental psychology has reached a stage at which it seems natural and sound to give attention also to its possible service for the practical needs of life" (p. 8). He further stated that "while the court makes fullest use of all the modern scientific methods when...a drop of dried blood is to be examined...the same court is completely satisfied with...common prejudice and ignorance when a mental product, especially the memory report of a witness, is to be examined" (p. 45).

More recently, it has been asserted that "research on evidence produced by eyewitnesses to an event is among the most systematic and theoretically developed in the psychology of law" (13). Thus, the foundation was laid for the use of eyewitness expert testimony to assist the trier of fact in a court of law. By 1988, some 450 trials in 25 states had admitted expert testimony on eyewitness evidence (14). Lindsay (15) states "Expert testimony on eyewitness issues easily meets the legal criteria of admissibility; most people have little idea of how to evaluate the accuracy of an eyewitness identification, underestimate the importance of many relevant variables, and apply significantly wrong expectations regarding other variables" (p. 381).

Given the adversarial nature of a criminal trial, psychologists hired to testify as expert witnesses typically are hired by the representatives for one side in the case. McCloskey, Egeth, and McKenna (11) have argued that the introduction of eyewitness expert testimony raises ethical issues and possibly will lead to a "battle of the experts." Further, the results of the available research on the effects of eyewitness expert testimony in mock trials is mixed. There is some evidence that jurors become more sensitized to eyewitnessing and identification conditions with the aid of the expert testimony (16, 17), while other studies suggest that jurors simply become more skeptical of eyewitnesses in general (18, 19). The use of a court-appointed expert on eyewitness testimony has been suggested as an alternative to adversarial expert testimony (20). Unfortunately, Cutler, Dexter, and Penrod (18) found that subjects who heard nonadversarial expert testimony were no better at differentiating between the good versus poor witnessing conditions than subjects who did not hear the expert testimony. The nonadversarial testimony did cause some juror skepticism, however.

At present, the preferred alternative to eyewitness expert testimony in the United Kingdom is judge's cautionary instructions (21). In the past 20 years, both federal and state courts in the United States also have encouraged judges to read instructions to jurors that specify factors they should consider when evaluating eyewitness evidence. In *Neil v. Biggers* (22), the

U.S. Supreme Court listed five factors that should be contained in judge's instructions to the jury in cases where eyewitness evidence is central to the case. These factors are: the opportunity for witnesses to view the criminal at the time of the crime, the length of time between the crime and the later identification, the level of certainty shown by the witnesses at the identification, the witnesses' degree of attention during the crime, and the accuracy of the witnesses' prior description of the criminal.

More recently, in *State v. Warren* (23), Kansas judges are encouraged to instruct jurors in the following seven factors: the opportunity the witness had to observe, the emotional state of the witness including that which might be caused by the use of a weapon, whether the witness had observed the defendant(s) on earlier occasions, whether a significant amount of time elapsed between the crime and later identification, whether the witness ever failed to identify the defendant(s) or made any inconsistent identification, the degree of certainty demonstrated by the witness, and whether there are any other circumstances that may have affected the accuracy of the eyewitness identification.

A major difference between eyewitness expert testimony and these pattern instructions from a judge is that the factors are listed without any guidance to the jurors on how to interpret those factors. For example, both sets of instructions include a reference to the witnesses' level of certainty, but most jurors would probably infer that this means a confident witness should be believed and a witness who is uncertain should be less believed. The majority of laboratory and field evidence suggests that this is not the case. Further, in *Biggers* a reference is made to the accuracy of the witnesses' prior description, but the available evidence indicates that we cannot predict the accuracy of an identification from a witnesses' prior verbal description (24). Thus, the lack of specificity of the pattern instructions suggests that they would not be an effective alternative to eyewitness expert testimony.

There has been some research carried out on the effectiveness of another set of pattern instructions that arose from *U.S. v. Telfaire* (25). In that case, Melvin Telfaire was found guilty of robbery based on the sole

ntained in judge's
:e is central to the
view the criminal
rime and the later
ses at the identifi-
ne, and the accu-

s are encouraged
portunity the wit-
ss including that
r the witness had
her a significant
fication, whether
ade any inconsis-
l by the witness,
have affected the

imony and these
sted without any
or example, both
evel of certainty,
onfident witness
be less believed.
iat this is not the
uracy of the wit-
ates that we can-
sses' prior verbal
ttern instructions
to eyewitness ex-

ectiveness of an-
Telfaire (25). In
based on the sole

identification of a Mr. Peregory. No other persons witnessed the robbery and no physical evidence was ever recovered. The court of appeals overturned the conviction on the premise that the trial court had failed "to give a special instruction on identification in which the case turned on the testimony of a single witness," even in the absence of a request by the defense counsel. The appeals court concluded that "Trial courts should as a matter of routine include an identification instruction which emphasizes to the jury the need for finding that the circumstances of identification are convincing...and in cases where identification is a major issue, the court should not rely on defense counsel to request such a charge." The *Telfaire* instructions make some reference to some of the factors typically presented by eyewitness experts, such as the confidence and carry-over effect issues, but these instructions show the same lack of specificity as the *Biggers* and *Warren* instructions. It is not surprising, therefore, that the *Telfaire* instructions generally have not been found to sensitize jurors to the eyewitnessing and identification conditions in a case (26, 27). Nevertheless, the *Telfaire* instruction has become the most widely used standardize instruction in the U.S. (28, 29).

While the use of cautionary instructions for eyewitness testimony has been greeted by some as a positive step, many researchers contend that instructions to the jury are often misunderstood (27, 30). Elwork, Sales and Alfini (31), for example, found that jurors could answer only 15 percent of 89 questions on judge's instructions from criminal trials accurately. The need for technically accurate statements of the law often results in instructions replete with legal terms that are incomprehensible to the jury, leading to juror confusion (32). Also, instructions on eyewitness testimony typically are embedded in a lengthy set of judicial instructions such that jurors may not give them adequate attention (33). Research on this issue has been inconclusive. Katzev and Wishart (34) found that subject-jurors who were presented cautionary, eyewitness-identification instructions embedded in general judge's instructions returned fewer guilty verdicts than subjects who only received the general instructions. In contrast, Cutler, Penrod, and Dexter (26) found no effect of the *Telfaire* instructions on

mock jurors' verdicts while Greene (27) found some juror skepticism with the instructions. Greene (27) found significant juror skepticism of eyewitness testimony when the *Telfaire* instructions were revised to make them more comprehensible to the jurors.

EXPERIMENT 1

Experiment 1 was designed to evaluate the effectiveness of the *Telfaire* instructions when given to jurors prior to the presentation of the evidence in the trial. In each of the studies of cautionary eyewitness-identification instructions, the instructions were presented only at the close of the trial. While there is no legal mandate regarding when cautionary instructions must be delivered to the jury, it has been the practice of the courts to give them to the jury just before the members begin their deliberations. This practice follows from judges' beliefs in jurors' "recall readiness" (35). Judge E. B. Prettyman questioned the wisdom of this practice when he wrote: "What manner of mind can go back over a stream of conflicting statements of alleged facts...or even the existence of the witnesses, and retrospectively fit all these recollections into a pattern of evaluations and judgments given to him for the first time after the events; the human mind cannot do so" (36, p. 1066). Concurring with Judge Prettyman's assertion, Kalven and Zeisel (37) found that jurors typically form firm opinions about a defendant's guilt before the close of the trial.

Research on the utility of presenting other instructions at the beginning of a trial, such as requirements of proof and determination of facts, has shown promise for helping jurors to focus on relevant evidence and to follow legal guidelines in their decision making (30, 38, 39). While Cruse and Brown (40) found no effect of presenting legal definitions at the beginning of a trial versus at the end, they did find an effect of the frequency of presentation, with multiple exposures at both the beginning and end increasing jurors' ability to apply the definitions in evaluating the testimony. There also is some evidence that presenting the legal definition of the alleged crime both before and after trial improves jurors' comprehension of those instructions (39).

There are two often-cited reasons for not giving cautionary instructions at the beginning of a trial. First, this practice may be impractical because judges will not know which instructions to give until after the evidence is heard. Second, this practice would put undue strain on the resources of the court. The results from a survey of judges by Heuer and Penrod (38) does not support either claim. The researchers asked 29 judges to either include or not to include preliminary instructions in actual trials. In all trials, the instructions were presented at the conclusion of the evidence. Based on the judges' experience in this study, they disagreed with the assertion that preliminary instructions were impractical because it was not possible to know which instructions to give until after the evidence was presented. They also disagreed that the use of preliminary instructions would be disruptive to the trial process.

The purpose of the present experiment was to evaluate the effectiveness of using preliminary *Telfaire* instructions for sensitizing jurors to factors likely to affect eyewitness identification performance. Mock jurors viewed a 30-minute videotape of a simulated trial in which the pivotal evidence was that of a single eyewitness (42). Based on the results from the available studies of other types of instructions, it was hypothesized that presenting the *Telfaire* instructions before the testimony would be more effective in sensitizing jurors to factors affecting eyewitness identification than presenting them only after the testimony. Based on the results of the study by Cruse and Brown (40), where frequency of presentation was found to affect jurors' performance, a condition also was included where the *Telfaire* instructions were presented both at the beginning and end of the trial. Each of these three conditions were compared to a control condition where no cautionary instructions concerning eyewitness identification were presented.

METHOD

Subjects

The subjects were 349 undergraduate students (139 males, 210 females) enrolled in introductory psychology classes at the University of

California, Los Angeles. Cutler et al. (26) reported that the differences in sensitivity to eyewitness evidence between undergraduates and eligible jurors were negligible. Subjects were randomly assigned to one of eight different conditions. The average number of subjects in each cell was 44. All subjects received course credit for their participation. The degrees of freedom vary on some statistical tests because of missing values on some variables from some subjects.

Design

The data matrix formed a 4 (instruction condition: before and after, before-only, after-only, none) by 2 (eyewitnessing condition: good, poor) by 2 (verdict: guilty, not guilty) array. The design was completely between subjects. In addition to a 2-way analysis of verdict, the dependent variables were: confidence rating, *Telfaire* elements listed as rationale for the verdict, *Telfaire* elements listed from memory, number of correct answers on a multiple-choice test concerning the facts of the case, and a believability rating for the eyewitness.

The Trial

The videotaped trial was a modified version of materials produced by Cutler et al. (1988). The trial concerned a case in which a liquor store was robbed by a lone gunman. Without the *Telfaire* instructions, the trial lasted approximately 30 minutes. The trial began with the prosecution and defense giving their opening statements, followed by the testimony of four witnesses. The first witness for the prosecution was the lone eyewitness who also was the clerk and victim of the crime. This woman was then cross-examined by the defense. The second witness was the investigating police officer who was questioned first by the prosecution and then by the defense. The defense then called two witnesses, a character witness and the defendant, who were questioned in turn by the defense and the prosecution. Following all four witnesses, the jurors heard the closing arguments of the prosecution and the defense, followed by the judge's instructions. The counsels in the trial were practicing attorneys in Madison, Wisconsin. Finally, the judge delivered his general instructions. These in-

the differences in
ıates and eligible
:d to one of eight
each cell was 44.
n. The degrees of
g values on some

:fore and after, be-
ın: good, poor) by
npletely between
e dependent vari-
s rationale for the
ıf correct answers
ase, and a believ-

rials produced by
a liquor store was
ns, the trial lasted
osecution and de-
testimony of four
: lone eyewitness
woman was then
the investigating
ın and then by the
acter witness and
ıse and the prose-
the closing argu-
ıe judge's instruc-
in Madison, Wis-
ctions. These in-

structions included an explanation of the requirements of burden of proof, definitions of reasonable doubt, the definition of the crime for which the defendant was charged, and general points that the jury should consider when evaluating the credibility of the testimony of any witness.

Two versions of the evidence in the trial were constructed by altering the testimony of the prosecution witnesses. In the poor eyewitnessing and identification condition, the eyewitness testified that: a) the robber was wearing a hat that covered his hairline; b) the robber was waving a handgun throughout the robbery; c) she had identified the robber in a lineup two weeks after the robbery had occurred; and d) when viewing the lineup of suspects, the police led her to believe that the robber was in the lineup. In contrast, in the good witnessing and identification condition, the eyewitness testified that: a) the robber's face and hair were not concealed; b) the robber had a gun but kept it hidden most of the time; c) she identified the robber just two days after the robbery; and d) when viewing the lineup of suspects, she was told that the robber may or may not be in the lineup. In each condition, the police officer corroborated the eyewitness's testimony concerning the lineup procedure.

The insertion of the *Telfaire* instructions in the trial was manipulated across four conditions. They either were or were not inserted immediately before the prosecutor's opening statement and they either were or were not inserted at the end of the judge's general instructions to the jury. The complete set of *Telfaire* instructions are presented in Appendix A. The two witnessing and identification conditions were completely crossed with the four instruction conditions to produce eight different conditions.

Materials and Procedure

Subjects were tested in groups ranging in size from one to six. At the beginning of each session, the subjects were asked to complete the juror bias scale (JBS) devised by Kassin and Wrightsman (43) to determine pretrial bias on the part of jurors toward the prosecution or defense. The JBS is intended to assess jurors' pretrial bias that a defendant is guilty of

committing the crimes for which he or she is charged and their threshold of certainty deemed necessary for convicting a defendant.

After completing the JBS and viewing one of the eight versions of the trial, each subject was given a sheet of paper to record the verdict (guilty, not guilty). Then, each subject was administered a questionnaire that contained several secondary dependent measures in the following order: a) a confidence rating for the verdict, b) three reasons for rendering the verdict, c) free recall of the judge's instructions concerning eyewitness testimony, d) the believability of the eyewitness, and e) six multiple-choice questions to assess the juror's memory for the evidence presented in the trial.

Finally, the subjects were asked to comment on the judge's instructions in a free format as follows: "In just a few words, what did you think of the judge's instructions that he read at the (beginning, end, beginning and the end) of the trial?"

RESULTS

Verdict

A log-linear analysis of variance was conducted to assess the effects of the instruction condition and eyewitnessing conditions on the subjects' verdicts. The percentage of subjects who voted guilty in each condition is presented in Table 1.

Table 1. Percentage of subjects in Experiment 1 voting guilty as a function of eyewitnessing conditions and instruction condition.

	Eyewitnessing Conditions	
Instruction Condition	Good	Poor
Before-and-After	50.0	47.1
After-Only	28.1	22.7
Before-Only	54.1	27.7
Control (no *Telfaire*)	50.0	27.7

Overall, approximately two-thirds of the subjects voted not guilty, which is consistent with the reasonable-doubt standard in a trial that revolves around the testimony of a lone eyewitness. The analysis showed that the verdicts were affected by an interaction between Instruction Condition and Eyewitnessing Condition [$\chi^2(3) = 8.04$, $p < .05$]. Separate chi-square analyses revealed that the subjects exhibited differential sensitivity to the eyewitnessing conditions only in the control (no instruction) and before-only instruction conditions ($ps < .05$). In addition, when the eyewitnessing conditions were good, the subjects were less likely to vote guilty in the after-only instruction condition than in the other three conditions [$\chi^2 (3) = 8.68$, $p < .05$]; and when the eyewitnessing conditions were poor, the subjects were more likely to vote guilty in the before-and-after instruction condition than in the other three conditions [$\chi^2 (3) = 7.59$, $p < .10$]. In summary, sensitivity to the eyewitness evidence was minimized when the *Telfaire* instructions were presented at the end of the trial, and this effect was to promote either juror overbelief or skepticism depending on whether the instructions were also presented at the beginning of the trial.

Confidence in Verdict

A three-way analysis of variance was conducted on the jurors' confidence ratings with the factors being eyewitnessing condition, instruction condition, and verdict (guilty, not guilty). Only the main effect of verdict was significant, with the subjects being significantly more confident in guilty verdicts than not guilty verdicts [$F (1,333) = 26.22$, $p < .01$]. This result is consistent with the appropriate interpretation of the reasonable doubt instruction. None of the other main effects or interactions were significant, with all $Fs < 1.57$. The overall average confidence rating was 3.52 on the 5-point scale, which is between "somewhat" and "very confident."

Rationale for Verdict

An average of 40 percent of the three reasons given by the subjects for their verdicts pertained to eyewitness issues contained in the *Telfaire* in-

structions. A three-way analysis of variance showed that the number of elements listed which are contained in *Telfaire* was not significantly related to eyewitnessing condition, instruction condition, or verdict (all Fs < 2.45). Thus, elements of the eyewitness evidence were just as likely to be cited as rationale for guilty verdicts as not guilty verdicts, and this was the case irrespective of the nature of the eyewitnessing conditions or the cautionary instruction conditions. The null effect with respect to eyewitnessing conditions is consistent with the results of Lindsay (15, p. 381).

Memory for the Telfaire Instructions

Subjects were asked to list everything they could remember from the judge's instructions concerning eyewitness testimony. The responses were scored for 10 elements from the *Telfaire* instructions which pertain to factors affecting eyewitness performance. The scoring was carried out by a panel of five research assistants working as a group. The average number of elements listed was 1.98. A three-way analysis of variance showed only a significant effect of instruction condition [$F(3,333) = 12.69$, $p < .01$], with memory for the instructions being a direct function of the number of times the instructions were presented: Before-After = 3.13, Before = 2.38, After = 2.30, and Control = 0.18. It is worthy of note that even when the instructions were presented twice, the subjects recalled only 31 percent of the elements. Memory for the instructions did not depend significantly on the nature of the eyewitnessing conditions or the subject's choice of verdict.

Memory for the Testimony

Consistent with the findings of Cutler et al. (1989), the subjects had little difficulty remembering the facts disclosed in the testimony of the trial. On average, subjects scored over 92 percent on the test of six multiple-choice questions about evidence in the case. This high level of performance was not affected by the eyewitnessing conditions, the instruction condition, or the verdict rendered (all Fs < 1.0).

at the number of
t significantly re-
r verdict (all Fs <
ıst as likely to be
, and this was the
litions or the cau-
ɛspect to eyewit-
y (15, p. 381).

nember from the
ıe responses were
which pertain to
'as carried out by
The average num-
variance showed
33) = 12.69, p <
ction of the num-
· = 3.13, Before =
te that even when
d only 31 percent
ıend significantly
ıbject's choice of

the subjects had
testimony of the
test of six multi-
ıigh level of per-
ıs, the instruction

Believability of Eyewitness

The average rating of believability of the clerk (eyewitness) was 2.46 on the 5-point scale, which falls midway between "somewhat" and "not very believable." This dependent variable formed a pattern similar to that of the confidence ratings. Subjects who voted guilty found the clerk to be significantly more believable (3.56) than subjects who voted not guilty (1.92), with $F(1,333) = 18.24$, $p < .01$. The believability scores were not significantly affected by the eyewitnessing conditions or the instruction condition. This was surprising given that the nature of the testimony by the clerk (eyewitness) was the basis for the eyewitnessing factor, and that the verdicts were significantly affected by the instruction condition. However, a debriefing of some of the subjects revealed that the question about believability was taken to address the sincerity of the eyewitness, not the accuracy of the identification. Inspection of the videotapes indicated that the clerk did not appear differentially sincere in the different versions of the trial. Leippe (44) also notes that a juror's conception of eyewitness believability is a function of factors other than just the witnessing conditions, including prior beliefs, message content, and message delivery style.

Juror Bias Scale (JBS)

This sample of subjects showed a wide range of scores on the JBS (14-70), with the average in the mid-range (42.91) on the 90-point scale. The sample was divided into two groups, those scoring above the mean and those scoring below the mean, to form two levels of a JBS factor. This factor was combined in a 3-way, log-linear analysis of variance with eyewitnessing condition and instruction condition as the other factors and verdict as the dependent variable. The main effect of JBS was significant [χ^2 (1) = 4.79, $p < .05$], with subjects scoring above the mean on the JBS (pro-prosecution) being more likely to vote guilty (45 percent) than subjects scoring below the mean on the JBS (21 percent). This bias did not interact with either the eyewitnessing condition or the instruction condition. Thus, the pro-prosecution bias of some subjects was not modulated by the eyewitness evidence or by the cautionary instructions.

Independent t-tests showed that the JBS scores were not related to the number of *Telfaire* elements cited by the subjects as rationale for their verdicts ($t < 1.0$), nor were the JBS scores related to the number of elements from the *Telfaire* instructions listed by the subjects after observing the trial ($t < 1.0$). Thus, pre-trial bias did not appear to affect either the subjects' consideration of eyewitness issues in rendering their verdicts or their attention to the cautionary instructions.

Discriminators of Verdict

A discriminant-function analysis was performed to predict verdict as a binary variable. The predictors were: eyewitnessing condition, instruction condition, number of *Telfaire* elements given as rationale for the verdict, number of *Telfaire* elements listed after the trial, juror pretrial bias (JBS), "believability" rating for the eyewitness, and confidence rating for the verdict. A combination of four predictors was found to accurately classify 81.3 percent of the subjects with respect to their verdicts [$F(4,346) = 27.64$, $p < .001$]. The four predictors were: the eyewitnessing conditions (the evidence), JBS (pre-trial bias), believability of the eyewitness (perceived sincerity), and confidence in the verdict (interpretation of reasonable doubt). Conspicuously absent from the predictor equation were independent contributions from instruction condition, number of *Telfaire* elements given as rationale for the verdict, and number of elements listed from the *Telfaire* instructions. The effects of *Telfaire* were not as salient in the subject-jurors' decisions as other factors.

Comments on the Instructions

The subjects were asked to comment on the judge's instructions at the end of the experiment. Seventy percent of the comments were negative and only four percent of the comments pertained explicitly to the eyewitness instructions. The subjects found the instructions to be "too long, boring, repetitious, confusing, and hard to remember." The *Telfaire* portion of the judge's instructions obviously was not salient to the subject-jurors, with only a few subjects commenting that the instructions emphasized the

not related to the
ationale for their
ıe number of ele-
ts after observing
› affect either the
ş their verdicts or

redict verdict as a
dition, instruction
le for the verdict,
retrial bias (JBS),
ıce rating for the
ccurately classify
dicts [F(4,346) =
essing conditions
f the eyewitness
erpretation of rea-
or equation were
umber of *Telfaire*
of elements listed
e not as salient in

instructions at the
ıts were negative
itly to the eyewit-
be "too long, bor-
Telfaire portion of
he subject-jurors,
ıs emphasized the

eyewitness or were biased against the eyewitness. This outcome is consistent with the findings of Greene (27).

DISCUSSION

The results of this experiment indicate that the pattern *Telfaire* instructions do not promote juror sensitivity to the quality of eyewitnessing and identification conditions presented in evidence. Presenting the instructions after the evidence reduced the subject-jurors' sensitivity to the eyewitness evidence and led them to vote not guilty regardless of the nature of the eyewitness evidence. Greene (27) also found some skepticism in jury verdicts with *Telfaire* as compared to a no-instruction condition, with an 11 percent drop in discrimination between good and poor conditions in one experiment and a 31 percent drop in a second experiment. In contrast, presenting the instructions both before and after the evidence in the present experiment led the subjects to vote guilty regardless of the nature of the eyewitness evidence. When the instructions were presented only before the evidence, the subjects' sensitivity to the eyewitness evidence was preserved, but to no greater extent than was obtained without any cautionary instructions about eyewitness testimony. This pattern is important given that the timing and frequency of the instructions appear to influence jurors' verdicts in different directions, causing either skepticism or overbelief.

It also is worthy of note that none of the variables pertaining to the presentation, recollection, or use of the instructions were significant discriminators of verdict beyond what could be predicted from pre-trial bias, the eyewitness evidence, and the subject-jurors' confidence. Further, the subjects could list only 31 percent of the elements contained in *Telfaire* even when the instructions were presented twice. It does not appear to be the case that the present subjects were generally inattentive or that the materials were poorly constructed because the subjects were 92 percent accurate on a test of the facts of the case. Instead, it appears that *Telfaire* is not a reliable alternative to eyewitness expert testimony.

From the variables measured in this experiment, it is not clear why presenting the instructions after the evidence should cause skepticism of the eyewitness whereas presenting the instructions both before and after the evidence should cause overbelief in the eyewitness. However, we speculate that with the instructions presented after the evidence, the subjects concluded that they may have missed something about the eyewitness and they became skeptical. This is troublesome given that judges tend to present the cautionary instructions at the end of the trial (38). With the instructions presented before the evidence, the subjects were more likely to rely on their recollection of the quality of the evidence and to discriminate between good and poor eyewitnessing conditions. However, with a re-presentation of the instructions after the evidence, the subjects tended to overemphasize the role of the eyewitness whom they tend to believe apriori, even under poor eyewitnessing conditions (45). Thus, juror discrimination of the eyewitness evidence was preserved with *Telfaire* presented only before the evidence, but even then discrimination was not improved beyond that observed without any cautionary instructions about eyewitness testimony.

Given the negative pattern of results with the *Telfaire* instructions, a revised set of instructions was constructed based on 13 factors that are typically discussed by eyewitness experts (46). These instructions were tested in Experiment 2 in comparison to the *Telfaire* instructions in a second trial that revolved around the testimony of a lone eyewitness.

EXPERIMENT 2

Several researchers have concluded that comprehension of legal language does not lie entirely on the complexity of the legal concepts, but rather on problematic linguistic constructions (27, 32, 47-49). By altering these constructions, they found comprehension to be improved irrespective of the legal complexity of the material. The language used in the present revised instructions was selected based on the psycholinguistic methodology proposed by Charrow and Charrow and on the modifications of *Telfaire* tested by Greene (27). In addition, however, the content of the instructions was modified and elaborated to include statements about 13

is not clear why
ıse skepticism of
before and after
ss. However, we
vidence, the sub-
about the eyewit-
n that judges tend
ial (38). With the
were more likely
e and to discrimi-
However, with a
ubjects tended to
ıd to believe apri-
ıs, juror discrimi-
elfaire presented
vas not improved
ons about eyewit-

re instructions, a
3 factors that are
instructions were
ructions in a sec-
witness.

sion of legal lan-
gal concepts, but
-49). By altering
nproved irrespec-
used in the pres-
ıolinguistic meth-
modifications of
he content of the
tements about 13

factors thought by most eyewitness experts to reliably affect the performance of a significant percentage of eyewitnesses (50, 51). These 13 factors were: duration of view, distance for view, lighting, high levels of stress, weapon focus, prior exposure to the suspect, delay of identification, lineup fairness, pressures to choose from a lineup, effects of post-event information, prior identifications, cross-race identification, and the relationship between eyewitness confidence and identification accuracy. The complete set of revised instructions are presented in Appendix B.

It was hypothesized that the revised cautionary instructions would result in greater comprehension of eyewitness issues, as measured by a multiple-choice test, than either the *Telfaire* instructions or no instructions at all. The effectiveness of the revised instructions on sensitization to the eyewitness evidence was evaluated by varying the eyewitnessing and identification conditions as in Experiment 1. A greater percentage of guilty verdicts were expected when the eyewitnessing conditions were portrayed as having been good than poor. Subjects viewed a 30-minute trial where the only incriminating evidence was the testimony from a lone eyewitness. All subjects were presented the judge's general instructions on reasonable doubt, the burden of proof, and the evidence standards at the end of the trial. One group of subjects also received the *Telfaire* instructions embedded in the judge's instructions whereas a third group received the revised instructions.

METHOD

Subjects

The subjects were 338 undergraduate volunteers (135 males and 203 females) from introductory psychology courses at the University of California, Los Angeles. The average number of subjects per cell was 56. Subjects received course credit for their participation. The degrees of freedom vary on some statistical tests because of missing values on some variables from some subjects.

Design

The data matrix formed a 3 (instruction condition: revised, *Telfaire*, none) by 2 (eyewitnessing condition: good, poor) by 2 (verdict: guilty, not guilty) array. The design was completely between subjects. In addition to a 2-way analysis of verdict, the dependent variables were: confidence rating, instruction elements listed as rationale for the verdict, instruction elements listed from memory, number of correct answers on a multiple-choice test concerning factors affecting eyewitness testimony, number of correct answers on a multiple-choice test concerning the facts of the case, and a believability rating for the eyewitness.

The Trial and Judge's Instructions

Two general versions of a simulated trial were constructed and were recorded on videotape. The versions differed only with respect to the quality of the eyewitnessing and identification conditions portrayed in the evidence. The case involved a Caucasian male accused of burglarizing the home of a Caucasian female. The only direct evidence against the defendant was the testimony of the victim, who unexpectedly entered her house while the perpetrator was still present. The witnesses for the prosecution were the victim and a detective who arrested the defendant on the night of the burglary. The defense witnesses were the defendant and his male friend who provided him with an alibi. The friend claimed that they were both at his house doing plumbing work in the kitchen at the time of the crime. After closing arguments, the judge read general instructions on reasonable doubt, the burden of proof, and the elements needed in order to convict a person with the charge of burglary. For one group of subjects, the *Telfaire* instructions were embedded within the judge's general instructions whereas a third group received the revised instructions (see Appendix B).

Two versions of the evidence in the trial were constructed by altering the testimony of the prosecution witnesses. In the poor eyewitnessing and identification condition, the eyewitness testified that: a) she viewed the burglar for five seconds; b) she viewed the burglar at a distance of 15 feet; c) the burglar carried a gun; d) she was "very, very upset and scared" by

revised, *Telfaire*,
verdict: guilty, not
:cts. In addition to
re: confidence rat-
ict, instruction ele-
rs on a multiple-
imony, number of
: facts of the case,

structed and were
ith respect to the
ıs portrayed in the
of burglarizing the
against the defen-
entered her house
or the prosecution
ant on the night of
ant and his male
ied that they were
at the time of the
nstructions on rea-
ieeded in order to
group of subjects,
udge's general in-
:tructions (see Ap-

:ructed by altering
:yewitnessing and
a) she viewed the
listance of 15 feet;
:et and scared" by

the incident; e) the lighting was from a small porch light; f) two weeks passed before she was asked to make an identification from a lineup; g) she failed to identify the burglar on an earlier occasion from mugshots; h) the police at the lineup told her that "they thought they had the guy," i) only "some" of the people in the lineup were of the same height, weight, and hair color as the suspect, and j) she said she had not seen him before.

In contrast, in the good witnessing and identification condition, the eyewitness testified that: a) she viewed the burglar for 20 seconds; b) she viewed the burglar at a distance of five feet; c) the burglar did not carry a weapon; d) she was "surprised" by the incident; e) the lighting was from the living room light where she encountered the burglar; f) two hours passed before she was asked to make an identification from a lineup; g) she identified the burglar prior to the lineup from mugshots; h) the police at the lineup told her that "the suspect may not be there," i) all of the people in the lineup were of the same height, weight, and hair color as the suspect, and i) she said she had not seen him before. In both versions of the trial, the eyewitness stated that she was "sure" of her identification. This level of confidence was selected such that it could be taken either way by the subject-jurors.

Six 30-minute videotapes were produced to cross the two eyewitnessing and identification conditions (good, poor) with the three instruction conditions (*Telfaire*, revised instructions, or none). In the revised instructions, the language employed in *Telfaire* was simplified based on the suggestions of Charrow and Charrow (47) and Greene (27) toward making it more comprehensible for the lay person. Legal terminology, uncommon words, and embedded phrases were eliminated. Changes also were made in the organization of the instructions to insure a logical structure within and between paragraphs. Finally, thirteen eyewitness factors were incorporated into the instructions toward educating the subjects about eyewitness phenomena and to sensitize jurors to good versus poor eyewitnessing conditions. The revised instructions are presented in Appendix B.

Procedure and Test Materials

Groups of subjects ranging in size from one and eight were assigned randomly to the six conditions. After viewing one of the six videotapes, each subject was given a sheet of paper to record the verdict (guilty, not guilty). Then, each subject was administered a questionnaire that contained several secondary dependent measures in the following order: a) a confidence rating for the verdict, b) three reasons for rendering the verdict, c) free recall of the judge's instructions concerning eyewitness testimony, d) the believability of the eyewitness, e) the comprehensibility of the instructions, and f) ten multiple-choice questions to assess the juror's knowledge of eyewitness factors (the effects of lighting, distance for view, and duration for view were assumed to be obvious and therefore were not tested).

RESULTS

Verdict

A log-linear analysis of variance was conducted to assess the effects of the instruction condition and eyewitnessing conditions on the subjects' verdicts. The percentage of subjects who voted guilty in each condition is presented in Table 2.

Table 2. Percentage of subjects in Experiment 2 voting guilty as a function of eyewitnessing conditions and instruction condition.

	Eyewitnessing Conditions	
Instruction Condition	Good	Poor
Revised Instructions	24.6	5.5
Telfaire Instructions	10.5	7.4
Control (no instruction)	29.0	4.3

Overall, approximately 88 percent of the subjects voted not guilty. While consistent with the reasonable-doubt standard in a trial that revolves around the testimony of a lone eyewitness, this percentage is considerably higher than the 67 percent of the subjects who voted not guilty in Experiment 1. The analysis showed that the verdicts were affected by an interaction between Instruction Condition and Eyewitnessing Conditions [$\chi^2(2) = 5.47$, $p = .065$], which was marginally not significant at the $p < .05$ level. Although the effect showed a strong trend, the floor effect for guilty verdicts in the poor viewing condition probably contributed to the nonsignificant result. Inspection of Table 2 reveals that when the eyewitnessing conditions were good, the subjects were less likely to vote guilty in the *Telfaire* instruction condition than in the other two conditions, whereas when the eyewitnessing conditions were poor, the instruction condition did not affect performance. In summary, sensitivity to the eyewitness evidence was minimized when the *Telfaire* instructions were presented at the end of the trial, and this effect was to promote juror skepticism.

Confidence in Verdict

A three-way analysis of variance was conducted on the jurors' confidence ratings with the factors being eyewitnessing condition, instruction condition, and verdict (guilty, not guilty). The results showed that confidence was not associated significantly with any of these three variables or by with an interaction between the variables (all $Fs < 1.21$). The overall average confidence rating was similar to that obtained in Experiment 1 (3.41 on the 5-point scale), which is between "somewhat" and "very confident." It is not readily apparent why the subjects in this experiment were not more confident in their guilty verdicts (as in Experiment 1) than their not guilty verdicts.

Rationale for Verdict

An average of 32 percent of the three reasons given by the subjects for their verdicts pertained to eyewitness issues contained in the revised instructions. This list of elements from the revised instructions was used for

this scoring because it is more comprehensive that *Telfaire*. A three-way analysis of variance showed that the number of elements listed as rationale for verdict was not significantly related to eyewitnessing condition [$F < 1.00$], but it was related to the instruction condition [$F(2, 299) = 16.33$, $MSe = 0.76$, $p < .001$]. The subjects were more likely to have given eyewitness elements as rationale for their verdicts after hearing the revised instructions (46 percent) than either *Telfaire* (16 percent) or no instructions (17 percent). The interaction between these variables was not significant ($F < 1.00$), and elements of the eyewitness evidence were no more likely to be cited as rationale for not guilty verdicts than guilty verdicts.

Memory for the Instructions

Subjects were asked to list everything they could remember from the judge's instructions concerning eyewitness testimony. The responses were scored for 13 elements from the revised instructions which pertain to factors affecting eyewitness performance. The scoring was carried out by a panel of five research assistants working as a group. The average number of elements listed was 2.29. A three-way analysis of variance showed only a significant effect of instruction condition [$F(2,299) = 10.89$, $p < .001$], with the number of eyewitness factors listed being 3.80, 2.18, and 0.89 for the revised, *Telfaire*, and no-instruction conditions respectively. A Tukey's posttest showed that both the revised and *Telfaire* instruction conditions differed from the control condition and they differed from one another (both $ps < .05$). It is worthy of note that even with the revised instructions, the subjects were able to list only 29 percent of the 13 elements. Memory for the instructions did not depend significantly on the nature of the eyewitnessing conditions, the subject's choice of verdict, or an interaction between those variables (all $Fs < 1.00$).

Believability of the Eyewitness

The average rating of believability of the burglary victim (eyewitness) was 2.78 on the 5-point scale, which is slightly under "somewhat" believable. The pattern of results was consistent with those from Experiment 1. Subjects who voted guilty found the victim/eyewitness to be significantly

aire. A three-way
listed as rationale
ng condition [F <
(2, 299) = 16.33,
to have given eye-
earing the revised
nt) or no instruc-
les was not signifi-
nce were no more
guilty verdicts.

member from the
he responses were
ich pertain to fac-
s carried out by a
e average number
iance showed only
10.89, p < .001],
2.18, and 0.89 for
ctively. A Tukey's
ruction conditions
from one another
vised instructions,
elements. Memory
the nature of the
t, or an interaction

ictim (eyewitness)
somewhat" believ-
rom Experiment 1.
to be significantly

more believable (3.41) than subjects who voted not guilty (2.00) [F(1,299) = 7.57, p < .01], but the believability scores were not significantly affected by the eyewitnessing conditions, the instruction condition, or an interaction with either of these variables (all Fs < 1.64). Hence, it is hypothesized that the question about believability was taken by the subjects to address the sincerity of the eyewitness, as was found in Experiment 1, rather than the accuracy of the identification. The victim/witness did not appear differentially sincere in the different versions of the trial.

Discriminators of Verdict

A discriminant-function analysis was performed to predict verdict as a binary variable. The predictors were: eyewitnessing conditions, instruction condition, number of eyewitness elements given as rationale for the verdict, number of eyewitness elements listed after the trial, "believability" rating for the eyewitness, confidence rating for the verdict, and knowledge of eyewitness factors (as measured by the 10-item multiple-choice test administered at the end of the experiment). A combination of four predictors was found to accurately classify 79.2 percent of the subjects with respect to their verdicts [F(4,307) = 24.28, p < .001]. This percentage is very similar to that accounted for in Experiment 1 by the variables collected there. The four significant predictors in the present experiment were: the eyewitnessing conditions (the evidence), believability of the victim/eyewitness (perceived sincerity), and the believability of the defendant and his friend (the defense witnesses).

The first two predictors are consistent with the results of Experiment 1. In contrast, the subjects' confidence in their verdicts was not related to verdict (see above). The subjects were more confident of guilty verdicts in Experiment 1 as would be expected from applying the reasonable doubt standard. Instead, the defense witnesses further discriminated the verdicts in the present experiment. This difference probably reflects differences in the nature of the evidence and the demeanor of the witnesses portrayed in the two trials.

As in Experiment 1, there were no significant contributions to the discrimination of verdicts from either instruction condition or the number of eyewitness elements listed after the trial. Although the subjects who gave more eyewitness elements as rationale for their verdicts were more likely to vote not guilty in this experiment, the effects of the judge's instructions were once again not as salient in the subject-jurors' verdicts as were other factors.

Effectiveness and Understanding of the Instructions

Subjects were asked to rate on a 5-point scale the effectiveness of the judge's instructions about eyewitness testimony in helping them decide on a verdict. The reported effectiveness of the instructions differed significantly between conditions [$F(1,299 = 4.05, p < .05$]. The average ratings of effectiveness were: Revised = (3.21), *Telfaire* = (2.31) and no cautionary instructions = (2.39). In addition, the subjects rated the revised instructions to be more understandable regardless of the eyewitnessing conditions, with [$F(2,299) = 3.89, p < .05$]. The average ratings were: Revised = 3.48, *Telfaire* = 2.66, and no cautionary instructions = 2.57. Tukey's posttests showed that the revised instructions were rated to be both more effective and understandable (all p s $< .05$).

Expert Knowledge of Eyewitness Factors

A 10-item multiple choice test was used to measure subjects' expert knowledge of eyewitness phenomena at the end of the experiment. A significant difference was found for instruction condition [$F(2,299) = 4.05, p < .05$]. Subjects who heard the revised instructions tended to score higher on the test than subjects who either heard *Telfaire* or no cautionary instructions (7.02 vs. 5.28 and 5.10). In addition, a significant positive correlation was found between scores on the test and recall of eyewitness elements after the trial [$r(311) = +.47, p < .001$]. Neither identification condition, verdict, nor an interaction between these variables was related to knowledge of the eyewitness factors (all Fs < 1.00). Thus, while the mere accessibility of eyewitness elements was related to expert knowledge

utions to the dis-
or the number of
ıbjects who gave
were more likely
lge's instructions
cts as were other

ectiveness of the
g them decide on
differed signifi-
ıverage ratings of
nd no cautionary
vised instructions
ssing conditions,
: Revised = 3.48,
Tukey's posttests
th more effective

subjects' expert
xperiment. A sig-
(2,299) = 4.05, p
ed to score higher
no cautionary in-
cant positive cor-
all of eyewitness
her identification
ables was related
. Thus, while the
expert knowledge

of eyewitness factors, and both the revised and *Telfaire* instructions increased this accessibility (see above), only the revised instructions showed a significant effect toward improving the subjects' expert knowledge of eyewitness factors.

Table 3. Percentage of subjects in Experiment 2 giving the Expert answer as a function of instruction condition.

	Instruction Condition			
Eyewitness Factor	Revised	*Telfaire*	Control	Average
Distance for view	61	58	54	58
Stress	90*	62	60	71
Weapon focus	92	84	83	86
Prior view of suspect	52	39	40	44
Duration for view	67	60	61	63
Lineup fairness	61*	41	36	45
Bias to select someone	60	54	60	58
Post-event information	88	79	76	58
Cross-race ID	64*	28	22	38
Confidence	61*	23	21	35
Average	70*	53	51	58

*The difference between instruction conditions was significant at $p < .05$.

The performance data for each item are presented in Table 3 as a function of instruction condition. A majority of the subjects disagreed with the experts on four of the ten items. These items pertained to the effects of prior viewing of the suspect, lineup fairness, cross-race identification, and eyewitness confidence. In contrast, most subjects answered the questions about weapon focus and post-event information in agreement with the expert responses. An item analysis revealed that the revised instructions significantly affected performance on four of the ten items: stress, lineup fairness, cross-race identification, and eyewitness confidence.

DISCUSSION

The results provided a replication of Experiment 1 and Greene (27) in that a) the *Telfaire* instructions had a negative effect on subject-jurors' sensitivity to the eyewitnessing and identification conditions portrayed in the trial testimony, and b) the subjects were able to significantly discriminate good from poor eyewitnessing and identification conditions without any cautionary instructions. The results further indicated that the revised cautionary instructions did not significantly affect sensitivity to the eyewitness evidence. While the revised instructions did not improve sensitivity, they did not result in the adverse effect observed with *Telfaire.* These results tentatively suggest that the present revised instructions are preferable to the linguistically-revised *Telfaire* instructions drafted and evaluated by Greene (27). She found significant skepticism to be created with her revised instructions. In fact, a difference of 61 percent in jury convictions between good and poor eyewitnessing conditions was totally eliminated with her revised instructions (Experiment 2). It is important to note, however, that Greene's instructions were not tested in the present experiment and she used collective jury verdicts as the dependent variable instead of individual juror verdicts.

The null effect of the present revised instructions on jurors' verdicts does not appear to have resulted from a lack of attention on the part of the subjects. Instead, the subjects were able to list more eyewitness factors following the trial when they were presented the revised instructions than either the *Telfaire* instructions or no instructions at all. The revised instructions also led to a significant effect on the subjects' expert knowledge of eyewitness factors. In contrast, no educational effect was found with *Telfaire.* Greene (27) also found no effect of *Telfaire* on expert knowledge, as well as a marginally significant improvement with an alternative revision of *Telfaire.* Finally, the subjects rated the present revised instructions to be more effective and understandable than either *Telfaire* or the judge's general instructions alone, and they listed elements of the revised instructions more often as rationale for the verdicts.

In sum, while the revised instructions did not enhance subject-jurors' sensitivity to the eyewitness evidence in this trial, they did not adversely affect the sensitivity that the subjects brought to the experiment. This outcome is consistent with reported effects of expert testimony itself (19) and suggests that the revised instructions constitute a positive alternative to the current pattern instruction. In addition, it is possible that in cases where the evidence and demeanor of the witnesses is less favorable to the defense than in the present experiment, the revised instructions would enhance the discrimination of good versus poor eyewitnessing conditions. In this experiment, 89 percent of the subject-jurors voted not guilty.

Finally, it is instructive to note that the level of expert knowledge about eyewitness factors could be predicted from the variables collected in this experiment for only 63 percent of the subject-jurors. Those subjects who could list eyewitness elements after the trial and those subjects who gave eyewitness elements as rationale for their verdicts showed greater agreement with the expert answers. However, the relatively low level of discrimination above chance suggests that the subjects' knowledge was largely a function of experiences external to the experiment, including external to the judge's instructions. Only the revised instructions showed a significant effect toward educating the jury about the factors addressed by eyewitness experts. With the revised instructions, the subjects agreed with the expert answers to 70 percent of the questions, whereas with either *Telfaire* or no cautionary instructions, the subjects agreed with the expert answers to only half the questions. This modest increase in comprehension is consistent with the findings of Masson and Waldron (52) who redrafted a set of legal contracts in "plain language." They found that apart from the constraints of language, non-experts have difficulty understanding information presented in legal contexts when they conflict with prior knowledge and beliefs. The modest increase also is consistent with the findings of Lindsay (15) who found little change in jurors' beliefs after receiving expert testimony.

In this regard, Smith (53) found that judges' instructions on a legal definition of the alleged crime were effective as a teaching aid for jurors

only when a supplementary instruction provided "a feature-by-feature attack on jurors' mistaken prior beliefs" (p. 511). Perhaps cautionary instructions on eyewitness testimony will not be maximally effective unless the most common misconceptions held by laypersons are stated and refuted in the instructions. Such a practice seems unlikely given that the "misconceptions" are somewhat open to debate (9) and judges' instructions are not open to cross-examination. Given this state, the cross-examination of expert witnesses may be the best thing going.

However, the revised instructions did significantly affect jurors' responses with respect to four crucial eyewitness factors: stress, lineup fairness, cross-race identification, and eyewitness confidence. This outcome is in contrast with the research of Lindsay (15) in which jurors were found to consider identifications from unfair lineups as accurate regardless of expert testimony, and Luus and Wells (54) who concluded that it is difficult to alter a juror's belief in a confident witness. Thus, as an educational tool, which may have significant carry-over effects in deliberations, the revised instructions appear promising.

In summary, the pattern of results suggests that: a) the *Telfaire* instructions are ineffective in educating subject-jurors about eyewitness factors, b) the revised instructions provide a promising alternative, and c) even the revised instructions do not alter subject-jurors' beliefs so as to perfectly mimic expert opinion.

APPENDIX A: *TELFAIRE* CAUTIONARY INSTRUCTIONS

One of the most important issues in this case is the identification of the defendant as the perpetrator of the crime. The Government has the burden of providing identity beyond a reasonable doubt. It is not essential that the witness himself be free from doubt as to the correctness of his statement. However, you the jury, must be satisfied beyond a reasonable doubt of the accuracy of the identification of the defendant before you may convict him. If you are not convinced beyond a reasonable doubt that the defendant was the person who committed the crime, you must find the defendant not guilty. Identification testimony is an expression of belief or im-

ıture-by-feature at-
aps cautionary in-
ly effective unless
are stated and re-
ely given that the
ıd judges' instruc-
state, the cross-
ɔing.
affect jurors' re-
stress, lineup fair-
e. This outcome is
rors were found to
ɜ regardless of ex-
l that it is difficult
n educational tool,
ations, the revised

a) the *Telfaire* in-
about eyewitness
alternative, and c)
s' beliefs so as to

ɔNS

lentification of the
ıent has the burden
ɔt essential that the
s of his statement.
ınable doubt of the
you may convict
ubt that the defen-
ust find the defen-
on of belief or im-

pression by the witness. Its value depends on the opportunity the witness had to observe the offender at the time of the offense and to make a reliable identification later.

In appraising the identification testimony of a witness, you should consider the following: Are you convinced that the witness had the capacity and an adequate opportunity to observe the offender? Whether the witness had an adequate opportunity to observe the offender at the time of the offense will be affected by such matters as how long or short a time was available, how far or close the witness was, how good were lighting conditions, whether the witness had occasion to see or know the person in the past.

In general, a witness makes any identification he makes on his perception through the use of his senses. Usually the witness identifies an offender by the sense of sight—but this is not necessarily so, and he may use other senses.

Are you satisfied that the identification made by the witness subsequent to the offense was the product of his own recollection? You may take into account both the strength of the identification, and the circumstances under which the identification was made.

If the identification by the witness may have been influenced by the circumstances under which the defendant was presented to him for identification, you should scrutinize the identification with great care. You may also consider the length of time that lapsed between the occurrence of the crime and the next opportunity of the witness to see the defendant, as a factor bearing on the reliability of the identification. You may also take into account that an identification made by picking the defendant from a group of similar individuals is generally more reliable than one which results from the presentation of the defendant alone to the witness. You may take into account any occasions in which the witness failed to make an identification of the defendant, or made an identification that was inconsistent with his identification at trial.

Finally, you must consider the credibility of each identification witness in the same way as any other witness, consider whether he is truthful, and

consider whether he had the capacity and opportunity to make a reliable observation on the matter covered in his testimony.

I again emphasize that the burden of proof on the prosecutor extends to every element of the crime charged, and this specifically includes the burden of proving beyond a reasonable doubt the identity of the defendant as the perpetrator of the crime with which he stands charged. If after examining the testimony, you have a reasonable doubt as to the accuracy of the identification, you must find the defendant not guilty.

APPENDIX B: REVISED CAUTIONARY INSTRUCTIONS

One of the issues in this case is the identification of the defendant as the person who committed the crime. It is the prosecution's responsibility to prove that the defendant is the one who committed this crime. You, the jury, must be convinced beyond a reasonable doubt, that the identification of the defendant as the person who committed the crime is accurate. If you are not convinced beyond a reasonable doubt that the defendant was the person who committed the crime, you must find the defendant not guilty. The testimony given by a witness is an expression of his or her beliefs; that is, their perception may or may not be accurate. You must evaluate the testimony and decide whether the witness had the opportunity to observe the defendant and make a correct identification later. There are two areas you must consider: factors at the time of the crime, and factors leading up to and including the identification.

In considering factors present at the time of the crime, you may ask yourself: Did the witness have the capacity and an adequate opportunity to observe the person who committed the crime? Think about the amount of time the witness had for observation, how far or close the witness was from the perpetrator, and the quality of the lighting.

You may consider the stress the witness may or may not have experienced. Extreme levels of stress sometimes makes identifying a person more difficult. If a witness is extremely afraid or distracted, his or her capacity to observe and remember may be reduced, although this is not always the case. Also, you may consider whether the witness was con-

fronted with a weapon. When a weapon is present, witnesses sometime concentrate on this weapon such that it may be more difficult to remember the person's face, although this is not always the case. You may also consider whether the witness had seen or known the person before the crime. Sometimes prior exposure helps an eyewitness to determine the identity of the personal but sometimes it leads to a mistaken identification.

In considering factors leading up to and including the identification, you may consider: How much time passed between the crime and the witness's identification? Memory loss for an event typically is greatest soon after the crime and then it decreases over time. Therefore, the sooner the identification is made, the more likely that the witness will remember the person accurately.

You must be sure that the witness made the identification on their own, and that they were not influenced by factors present at the time of the identification. Identifying a person from among others who look similar to the accused may be more reliable than identifying the person by seeing them presented by themselves, although this is not always the case.

You may also take into account whether the identification was conducted fairly, and whether or not the witness may have felt compelled to choose someone. You may also consider that memory of the crime can be affected by things a witness is told by other persons after the crime. You should consider the manner in which the witness was questioned, as this may affect their memory of the crime. You may also consider whether the witness was unable to make an identification on a prior occasion, or identified another person before identifying the defendant. You may take into account whether the defendant and the witness are of the same race, as some people have trouble identifying persons of another race, although this is not always the case.

Finally, you must evaluate the testimony of the all the witnesses in the same way. Ask yourself whether each witness had the capacity and opportunity to make a reliable observation. You may take into account that the confidence displayed by a witness does not necessarily indicate that they are accurate. It is possible for a witness to be confident and still be

wrong. Or, a witness may be unsure and still be correct in their identification.

Again, let me remind you that it is the prosecution's responsibility to prove all elements of the crime, specifically proving, beyond a reasonable doubt, the identification of the defendant. If after examining the testimony you have a reasonable doubt that an accurate identification was made, you must find the defendant not guilty.

AUTHOR NOTE

Requests for reprints may be sent to: R. Edward Geiselman, Department of Psychology, University of California at Los Angeles, Los Angeles, CA 90095-1563. Experiment 1 was based on the Honors thesis of the second author, and Experiment 2 was based on the Honors thesis of the first author which received the John Seward Award for excellence in research from the Department of Psychology of UCLA. The authors wish to thank Ana Corral, Makan Delrahim, Lobat Ellis, Elana Evan, Karyne Ghantous, Jennifer Kent, Jeannie Kim, Eric Mayo, Kila O'Neill, Neda Perry, Ronald Plata, and Karen Wang for their valuable assistance with literature review, data collection and data interpretation.

REFERENCES

1. Goldstein AG, Chance JE, Schneller GR: Frequency of eyewitness identification in criminal cases: a survey of prosecutors. Bulletin of the Psychonomic Society 1989; 27:71-74
2. Rand Corporation: The Criminal Investigation Process, Vols. 1-3. Rand Corporation Technical Report R-1777-DOJ. Santa Monica, California, 1975
3. Sanders GS: On increasing the usefulness of eyewitness research. Law and Human Behavior 1986; 10:333-336
4. Visher CA: Juror decision making: the importance of evidence. Law and Human Behavior 1987; 11:1-17
5. Huff R, Rattner A, Sagarin E: Guilty until proven innocent. Crime and Delinquency 1986; 32:518-544

t in their identifica-

l's responsibility to
eyond a reasonable
ining the testimony
ion was made, you

Geiselman, Depart-
ngeles, Los Ange-
lonors thesis of the
onors thesis of the
or excellence in re-
The authors wish to
ana Evan, Karyne
Gila O'Neill, Neda
ole assistance with

ncy of eyewitness
ors. Bulletin of the

ss, Vols. 1-3. Rand
Monica, California,

ness research. Law

of evidence. Law

nocent. Crime and

6. Wells GL: What do we know about eyewitness identification? American Psychologist 1993; 48:553-571
7. Fisher RP, Geiselman RE: Memory enhancing techniques for investigative interviewing. Springfield, Ill, Charles Thomas Publishers, 1992
8. Yarmey AD: The Psychology of Eyewitness Testimony. New York, Free Press, 1979
9. Elliot R: Expert testimony about eyewitness identification: a critique. Law and Human Behavior 1993; 17:423-437
10. Konecni VJ, Ebbesen EB: Courtroom testimony by psychologists on eyewitness identification issues: Critical notes and reflections. Law and Human Behavior 1986; 10:117-126
11. McCloskey M, Egeth H, McKenna J: The experimental psychologist in court: the ethics of expert testimony. Law and Human Behavior 1986; 10:1-14
12. Munsterberg H: On the witness stand. New York, McClure, 1908
13. Monahan J, Loftus EF: The psychology of law. Annual Review of Psychology 1982; 33:441-475
14. Brigham J: Disputed eyewitness identifications: can experts help? The Champion 1988; 8:10-18
15. Lindsay RCL: Expectations of eyewitness performance: jurors' verdicts do not follow from their beliefs, in Adult Eyewitness Testimony. Edited by Ross DF, Read JD, Toglia MP. New York, Cambridge Press, 1994
16. Cutler BL, Penrod SD, Dexter HR: The eyewitness, the expert psychologist, and the jury. Law and Human Behavior 1989; 13:311-331
17. Wells GL, Wright EF: Practical issues in eyewitness research, in The Impact of Social Psychology on Procedural Justice. Edited by Kaplan MF. Springfield, IL, Charles Thomas, 1986; 109-134
18. Cutler BL, Dexter HR, Penrod SD: Nonadversarial methods for sensitizing jurors to eyewitness evidence. Journal of Applied Social Psychology 1990; 20:1197-1207

19. Wells GL, Lindsay RCL, Tousignant JP: Effects of expert psychological advice on juror judgments in eyewitness testimony. Law and Human Behavior 1980; 4:275-28b

20. Lindsay RCL, MacDonald P, McGarry S: Perspectives on the role of the eyewitness expert. Behavioral Sciences and the Law 1989; 8:457-464

21. Devlin Lord: Report to the Secretary of State for the Home Department of the Departmental Committee on Evidence of Identification in Criminal Cases. London, Her Majesty's Stationary Office, 1976

22. Neil v. Biggers, 409 U.S. 188, 93 S. Ct. 375; 34 L. Ed. 2d 401 (1972)

23. State v. Warren, 635 P.2d 1236 (1981)

24. Deffenbacher KA: A maturing research on the behavior of eyewitnesses. Applied Cognitive Psychology 1991; 5:377-402

25. United States v. Telfaire, 469 F.2d 552 (1972)

26. Cutler BL, Penrod SD, Dexter HR: Juror sensitivity to eyewitness identification evidence. Law and Human Behavior 1990; 14:185-191

27. Greene E: Judge's instruction on eyewitness testimony: Evaluation and Revision. Journal of Applied Social Psychology 1988; 18:252-276

28. Holtshouser SE: Eyewitness identification testimony and the need for cautionary jury instructions in criminal cases. Washington University Law Quarterly 1983; 60:1387-1433

29. Woocher FD: Did your eyes deceive you? Expert psychological testimony on the unreliability of eyewitness identification. Stanford Law Review 1977; 29:969-1030

30. Elwork A, Sales BD, Alfini JJ: Juridic decisions: in ignorance of the law or in light of it? Law and Human Behavior 1977; 1:163-189

31. Elwork A, Sales BD, Alfini JJ: Making jury instructions understandable. Charlottesville, VA, The Michie Co., 1982

32. Severance LJ, Greene E, Loftus EF: Toward criminal jury instructions that jurors can understand. The Journal of Criminal Law and Criminology 1984; 75:198-233

expert psychologi-
ıony. Law and Hu-

ives on the role of
Law 1989; 8:457-

the Home Depart-
of Identification in
)ffice, 1976

3d. 2d 401 (1972)

ehavior of eyewit-
402

vity to eyewitness
.990; 14:185-191

ny: Evaluation and
38; 18:252-276

ıy and the need for
hington University

psychological testi-
tion. Stanford Law

in ignorance of the
7; 1:163-189

uctions understand-

ıal jury instructions
ıal Law and Crimi-

33. Loftus EF: Eyewitness Testimony. Cambridge, MA, Harvard University Press, 1979

34. Katzev RD, Wishart SS: The impact of judicial commentary concerning eyewitness identifications on jury decision making. The Journal of Criminal Law and Criminology 1985; 76:733-745

35. Wrightsman LS: Psychology and the Law. Pacific Grove, CA, Brooks/Cole Publishing Co., 1987

36. Prettyman EB: Jury instructions—first or last? American Bar Association Journal 1960; 46:1066

37. Kalven H, Zeisel H: The American Jury. Boston, Little Brown, 1966

38. Heuer L, Penrod SD: Instructing jurors: a field experiment with written and preliminary instructions. Law and Human Behavior 1989; 13:409-431

39. Smith VL: The feasibility and utility of pretrial instruction in the substantive law: a survey of judges. Law and Human Behavior 1990; 14:235-248

40. Cruse D, Browne, BA: Reasoning in a jury trial: the influence of instructions. The Journal of General Psychology 1987; 114:129-133

41. Smith VL: Impact of pretrial instruction on jurors information processing and decision making. Journal of Applied Psychology 1991; 76:220-228

42. Cutler BL, Penrod SD, Stuve TE: Jury decision making in eyewitness identification cases. Law and Human Behavior 1989; 12:41-45

43. Kassin S, Wrightsman LS: The construction and validation of a juror bias scale. Journal of Research in Personality 1983; 17:423-442

44. Leippe MR: Appraisal of eyewitness testimony, in Adult Eyewitness Testimony. Edited by Ross DF, Read JD, Toglia MP. New York, Cambridge Press, 1994

45. Lindsay RCL, Wells GL, Rumpel C: Can people detect eyewitness identification accuracy within and between situations? Journal of Applied Psychology 1981; 66:79-89

46. Geiselman RE: Eyewitness Expert Testimony. Balboa Island, CA, American College of Forensic Psychology Press, 1994

47. Charrow RP, Charrow VR: Making legal language understandable: a psycholinguistic study of jury instructions. Columbia Law Review 1979; 79:1306-1374

48. Goodman J, Greene E: The use of paraphrase analysis in the simplification of jury instructions. Journal of Social Behavior & Personality 1989; 4:237-251

49. Reifman A, Gusick SM, Ellsworth PC: Real jurors' understanding of the law in real cases. Law and Human Behavior 1992; 16:539-554

50. Kassin SM, Barndollar KA: The psychology of eyewitness testimony: a comparison of experts and prospective jurors. Journal of Applied Social Psychology 1992; 22:1241-1249

51. Kassin SM, Ellsworth PC, Smith VL: The "general acceptance" of psychological research on eyewitness testimony. American Psychologist 1989; 44:1089-1098

52. Masson MEJ, Waldron MA: Comprehension of legal contracts by non-experts: effectiveness of plain language redrafting. Applied Cognitive Psychology 1994; 8:67-85

53. Smith VL: When prior knowledge and law collide. Law and Human Behavior 1993; 17:507-536

54. Luus CAE, Wells GL: Eyewitness identification confidence, in Adult Eyewitness Testimony. Edited by Ross DF, Read JD, Toglia MP. New York, Cambridge Press, 1994

ABOUT THE AUTHORS

R. Edward Geiselman is Professor of Psychology at the University of California, Los Angeles. He holds a Ph.D. in experimental psychology from Ohio University and has contributed over eighty articles on memory theory and applied psychology to police science and psychological journals. Gabriella Ramirez and Dennis Zemba received their undergraduate degrees in psychology from UCLA in 1995 and 1994, respectively.

Psychology, Public Policy, and Law
2011, Vol. 17, No. 1, 140–159

© 2010 American Psychological Association
1076-8971/10/$12.00 DOI: 10.1037/a0021311

THE ROLE OF EYEWITNESS IDENTIFICATION EVIDENCE IN FELONY CASE DISPOSITIONS

Heather D. Flowe
University of Leicester

Amrita Mehta
and Ebbe B. Ebbesen
University of California, San Diego

We addressed the question of whether felony case dispositions are associated with eyewitness identification evidence. Toward this end, 725 felony cases (rape, robbery, and assault) were randomly sampled from the archives of a District Attorney's Office in a large south-western city in the United States. A positive identification was present more often in accepted compared to rejected cases, although the association was significant in acquaintance cases (i.e., cases in which one or more of the eyewitnesses was familiar with the defendant), not stranger cases. Additionally, suspect and crime incident factors were associated with case issuing outcomes to a larger extent than eyewitness identification evidence. Analyses further indicated that eyewitness identification evidence was stronger in prosecuted compared to rejected cases in which eyewitness testimony was the sole evidence against the defendant. Neither the presence of multiple identifications nor nonidentifications of the suspect varied across issuing outcomes. The findings are discussed in relation to additional research that is needed at the police and prosecution stages to advance public policy development with respect to the evaluation of eyewitness identification evidence by prosecutors.

Keywords: witnesses, lineups, legal processes, legal evidence, legal testimony

The connection between erroneous eyewitness identification evidence and wrongful convictions has been the motivating force behind much research in forensic psychology (Wells, Memon, & Penrod, 2006). In analyses of cases in which inmates were later exonerated, erroneous eyewitness identification has been cited as the primary reason for the wrongful convictions (e.g., Borchard, 1933; Connors, Lundregan, Miller, & McEwen, 1996; Gross et al., 2005). Wrongful conviction cases may signal that prosecutors are not screening out cases that have weak eyewitness identification evidence. What is more, perhaps cases with eyewitness identification evidence are more likely to be prosecuted than otherwise would be the case, regardless of whether there is other case evidence to implicate the suspect.

The overarching goal of the present study was to examine the role that eyewitness identification evidence plays in felony issuing decisions. The research was carried out in one of the largest District Attorney's Office in the United States, which serves a population equal to about 3 million people. Laboratory research has played a strong and important role in identifying factors that can affect the reliability of memory. This body of research has served as the foundation for the development of best practice guidelines for administering lineups in

This article was published Online First October 18, 2010.

Heather D. Flowe, School of Psychology, Forensic Section, University of Leicester; Amrita Mehta and Ebbe B. Ebbesen, Department of Psychology, University of California, San Diego.

Correspondence concerning this article should be addressed to Heather D. Flowe, School of Psychology, Forensic Section, University of Leicester, 106 New Walk, Leicester, UK, LE1 7EA. E-mail: hf49@le.ac.uk

criminal cases (Wells et al., 2000). We have accumulated limited systematic empirical knowledge, however, of eyewitness identification evidence in actual cases in the United States. Most of what we do know has been gathered from analyses of cases that were not randomly sampled (e.g., Behrman & Davey, 2001). Such cases may not be representative of the types of cases that prosecutors typically evaluate. Consequently, we do not know the types of cases in which eyewitness identification plays an instrumental role in implicating the suspect, whether legal officials assess the strength of eyewitness identification evidence, or whether eyewitness identification evidence affects the decision to prosecute once other case factors are taken into account. Descriptive information of this sort is important because it may profitably lead to additional studies that can further guide public policy development with respect to improving the reliability of eyewitness identification evidence that is used to prosecute defendants. In the sections that follow, we review archival research on prosecutorial decision making and provide an overview of the factors that can affect the accuracy of eyewitness memory. The specific aims of this archival study are then presented.

Archival Research on Felony Issuing Decisions

Prosecutors have the discretion to determine whether a suspect will be charged and what charges the suspect should face (*Bordenkircher v. Hayes*, 1978). Prosecutors also have a legal and ethical obligation to protect felony suspects who are not just innocent-in-fact, but who are also innocent-in-law (California District Attorneys Association, 1996). Charges should not be filed even if the prosecutor has a personal belief in the suspect's guilt. Rather, issuing decisions should be guided by whether the evidence in the case is legally sufficient and admissible. Previous archival research has found that felony charges are more likely to be issued if there is physical evidence to support the allegations (Albonetti, 1987; Feeney, Dill, & Weir, 1983; Jacoby, Mellon, Ratledge, & Turner, 1982; Miller, 1969; Nagel, & Hagan, 1983) and if the crime is serious, such as when a victim has been injured (Kingsnorth, MacIntosh, & Sutherland, 2002). Factors that may lead prosecutors to not file charges include: A primary aggressor has not been identified (e.g., the California Primary Aggressor Law requires a primary aggressor be identified), the suspect is thought to be innocent, or there are "interests of justice" concerns, such as the suspect will provide testimony in a more serious case (Silberman, 1978). Despite the fact that much research has been carried out examining the relationship between evidentiary factors and felony issuing decisions, little is known about the role that eyewitness identification evidence may play in prosecution.

Eyewitness Identification Evidence and Felony Issuing

The prosecution of cases in which eyewitness identification is the sole evidence linking the suspect to the crime may raise serious concerns. Meta-analytic reviews of the large body of laboratory research on eyewitness identification indicates that there are several factors that can reduce the accuracy of face recognition, including: relatively shorter durations of exposure to the culprit (Shapiro & Penrod, 1986), weapon exposure (Steblay, 1992), stress (Deffenbacher, Bornstein, Penrod, & McGorty, 2004), if the culprit is of a different race

than the eyewitness, a factor which is known as *own race bias* or the *cross race effect* (Meissner & Brigham, 2001); relatively longer retention intervals between the crime and the identification test (Shapiro & Penrod, 1986), and the type of procedure that is used to test the eyewitness' memory (Steblay, Dysart, Fulero, & Lindsay, 2003). These results indicate that the validity of the eyewitness identification evidence can be influenced by the circumstances of the crime and the identification test. Arguably, therefore, legal officials should be cautious about prosecuting individuals when there is only eyewitness identification evidence to tie the suspect to the crime.

Survey research suggests that prosecutors may not be sufficiently sensitive to the factors that can affect eyewitness accuracy (Brigham, 1981). Other research finds that the reliability of eyewitness identification evidence is more favorably viewed by jurors, judges, and the police compared to eyewitness memory experts (e.g., Benton et al., 2006; Magnussen, Melinder, Stridbeck, & Raja, 2009). Therefore, felony charging may be more likely in cases that have eyewitness identification evidence because the evidence persuades a prosecutor of the suspect's guilt or because the prosecutor believes the evidence will persuade a jury of the defendant's guilt. Only a single study, to our knowledge, has examined the role that eyewitness identification evidence plays in felony charging. The study surveyed prosecutors asking them to estimate how often they processed cases in which eyewitness identification was the critical piece of evidence linking the defendant to the crime (Goldstein, Chance, & Schneller, 1989). Prosecutors estimated on average that 3% of the time they encountered such cases. More systematic data, however, is needed to further address the association between eyewitness identification evidence and case prosecution.

Once felony charges are issued against a defendant, procedural safeguards that might be available at the trial stage to protect innocent defendants from being convicted on the basis of erroneous eyewitness identification evidence may be ineffective. Previous research suggests that procedural safeguards, such motions to suppress identification evidence at trial, the cross-examination of eyewitnesses, and cautionary instructions to the jury regarding eyewitness evidence (see Devenport, Kimbrough, & Cutler, 2009, for a review), may not adequately protect defendants who have been mistakenly identified. Furthermore, most defendants in the criminal justice system are convicted on the basis of a plea bargain agreement (Piehl & Bushway, 2007; U.S. Department of Justice, 2001), and as such, the reliability of the eyewitness identification evidence will not be evaluated by a judge or a jury in the vast majority of cases. Thus, there is a need to describe the extent to which eyewitness identification evidence plays a role at earlier stages of the criminal justice system, such as at the arrest and prosecution stages.

Goals of the Present Study

In the jurisdiction under study, the Uniform Crime Charging Standards (California District Attorney's Association, 1996), hereafter referred to as UCCS, provide basic criteria to assist prosecutors in determining whether a suspect should face felony charges. The criteria are not legally binding and are meant to assist prosecutors in their exercise of discretion. The criteria recommend prosecutors charge a case only when the following four basic criteria are met: (a) the

evidence shows suspect is guilty of the crime to be charged, (b) there is legally sufficient and admissible evidence *corpus delicti*, whereby the prosecutor can reasonably argue that a substantial crime—affecting significantly the personal or property rights of others—has been committed, (c) there is legally sufficient and admissible evidence of the suspect's identity, and (d) there is a reasonable probability of a conviction in view of the evidence and the foreseeable defense that could be raised.

Given this background, we tested whether charges were issued more often when a case had eyewitness identification evidence, all else being equal. Cases that were forwarded by the police to the District Attorney's Office for prosecution were analyzed to examine the relationship between felony issuing decisions and case characteristics, including suspect, crime, and eyewitness factors. The factors recorded from the case file included: physical evidence, crime severity, prior record, arrest at the scene of the crime, suspect admittance to the crime, and eyewitness identification evidence.

The second aim was to test whether the quality of identification evidence varied in relation to felony issuing outcomes. The U. S. Supreme Court has established criteria, which are commonly referred to as the *Biggers* criteria, for inferring the accuracy of an eyewitness' identification (*Neil v. Biggers*, 1972). The *Biggers* criteria are: the degree to which the eyewitness paid attention to the culprit, the length of time between the crime and the identification test, the eyewitness' identification certainty, the quality of the view that the eyewitness had of the culprit, and the degree to which the suspect matches the eyewitness' description. Previous research has indicated that layperson evaluations regarding the validity of eyewitness evidence relate to the *Biggers* criteria in a summative fashion (Bradfield & Wells, 2000). The present study extended this work by testing whether cases that are issued for felony prosecution are perceived to have stronger eyewitness identification evidence compared to those that are rejected. The degree to which the *Biggers* criteria allow for making a strong inference about the accuracy of an identification has been called into question by psychologists (e.g., Wells & Murray, 1983). Therefore, it is important to point out that we are not endorsing the use of these criteria for establishing eyewitness identification strength. Rather, the aim is to test whether these factors might influence prosecutor perceptions of eyewitness identification evidence strength.

The third aim was to test whether eyewitness identification evidence is stronger in prosecuted versus rejected cases in which eyewitness testimony served as the sole evidence against the suspect. The UCCS provide further guidance to prosecutors regarding cases in which eyewitness testimony is the sole evidence linking the accused to the crime. The UCCS advise that in such cases the prosecutor should generally charge only when: (a) there is no opportunity for mistake because the eyewitness(es) know the suspect, (b) the eyewitness(es) had a substantial opportunity to have observed the culprit, (c) the investigative standards have been satisfied and the eyewitness(es) can furnish an adequate description of the accused, and (d) a line up has been conducted.

In light of these guidelines, we tested whether cases in which eyewitness identification evidence was the sole evidence against the suspect had stronger eyewitness identification evidence than other cases. Additionally, we examined whether photo arrays were more common in eyewitness identification only cases.

In particular, the rate of prosecution was compared for cases that had live show-up compared to photo array identifications. In a live show-up procedure, the police present a single suspect in vivo to the eyewitness for identification at the scene of the crime, whereas in a photo array procedure, the suspect's photograph is presented to the eyewitness imbedded among photographs of distractor persons. The validity of a suspect identification from a live show-up is more difficult to assess compared to a suspect identification from a photo array. A positive identification from a show-up may result from guessing alone, whereas the odds that a suspect is identified from a photo array based on guessing alone are reduced by virtue of there being distractor persons (Steblay et al., 2003). If prosecutors scrutinize the strength of eyewitness identification evidence in deciding whether to issue charges, then perhaps charges are issued more often in cases that have photo array compared live show-up identification evidence.

The fourth and final aim was to test whether prosecuted cases were more likely than rejected cases to have multiple eyewitnesses that identified the police suspect. The likelihood that the suspect is guilty should theoretically increase as the number of positive identifications in a case increases. Additionally, Clark and Wells (2008) demonstrated that nonidentifications of the suspect affect the odds that the suspect is guilty to a larger extent than do multiple positive identifications. In view of their findings, we also tested whether nonidentifications of the suspect covaried with charging decisions.

Finally, the analyses of the cases were conditioned on whether any of the eyewitnesses had been previously acquainted with the suspect. Stranger cases, in which none of the eyewitnesses are acquainted with the perpetrator, arguably carry a greater risk of wrongful conviction because of mistaken eyewitness identification. However, cases that involved an identification of the suspect by an acquaintance were not excluded from the sample because even in acquaintance cases the defendant may have been mistakenly identified. To illustrate, eyewitness experts are called to testify by the defense about eyewitness errors in cases in which the suspect was positively identified from a lineup or a show-up by an acquaintance (e.g., *People v. Aguilar*, 2008; *People v. Cummings*, 2009; *People v. Figueroa*, 2008; *People v. Guerra*, 2009; *People v. Hernandez*, 2004; *People v. Ledesma*, 2006; *People v. Lopez*, 2004; *People v. Magana*, 2003; *People v. Robles*, 2009; *People v. Rountree*, 2005; *People v. Rubalcava*, 2005; *People v. Trujillo*, 2009; *Smith v. Smith*, 2003; *State of Washington v. Riofta*, 2003; *United States v. Burton*, 1998; *United States v. James*, 2001). Although the eyewitness was acquainted with the defendant (e.g., lived in the same neighborhood, went to the same high school, lived in the same building) at issue in these cases was whether the eyewitness' perception of the culprit was affected by variables such as stress, the presence of a weapon, or the witness being of a different race than the defendant. Additionally, inclusion of acquaintance cases is needed to calculate the overall prosecution rate for stranger cases in which eyewitness identification evidence is the sole evidence implicating the suspect.

Method

Case Selection

Felony cases (rape $n = 302$, robbery $n = 239$, and assault $n = 184$) were randomly sampled from the District Attorney's closed (no appeal pending) case

archives (1991–2000), which included cases that the DA accepted or rejected for prosecution (36% of the cases sampled were rejected). The sample was comprised of adult suspects that had been arrested for violating specific sections of the California Penal Code.[1] Basing the sampling scheme on specific sections of the Penal Code ensured that the cases within a given crime category were relatively homogenous with respect to perpetrators and victims. For example, none of the rape cases included child victims or statutory rape allegations. These crime categories were selected because they are violent felonies and were thought likely to have eyewitness identification evidence.

The arrest rates for rapes (45%), robberies (34%), and assaults (72%) for this city during the study period were comparable to national arrest rates (51, 25, and 56%, respectively) (U.S. Department of Justice, Federal Bureau of Investigation, 1996). Moreover, with respect to issuing decisions across the cases in the population from which we sampled, we determined that the prosecutor issued felony charges 56% of the time in rape and 88% of the time in robbery, which is comparable to previous estimates (U.S. Department of Justice, 1998) of prosecutorial charging rates (54 and 73%, respectively); the prosecution rate for assault was 68%, which is a rate that is higher than a previous estimate (34%) (U.S. Department of Justice, 1998).

Among prosecuted cases (n = 465), 73% of defendants pleaded guilty, 16% of defendants went to trial (84% were found guilty on one or more charges), and all charges were eventually dismissed for the remaining 12% of defendants. Defendants were convicted on average of 43% (M = 0.43) of the original charges (SD = 0.35; range: 0.00-.1.00, Median = 0.33).

Materials

A case file typically contained police records, the preliminary trial transcript, the probation officer's assessment of the defendant's personal history and the sentencing agreement. A total of 46 research assistants coded the cases and were provided with detailed (written and verbal) instructions, as well as a number of practice cases to code. The reliability of the coding was examined by having more than one person code a random subset of the cases in the sample, as well as by having the first author on the study check every coding form for consistency and completeness. As might be expected given the basic level of coding, agreement between coders was extremely high (e.g., 100% for crime type, 100% for eyewitness identification procedure type; 100% for both eyewitness and suspect

[1] Rape was defined as "an act of sexual intercourse accomplished with a person not the spouse of the perpetrator" under conditions in which a person is incapable of giving consent or is prevented from resisting, or under conditions in which the sexual intercourse is accomplished "against a person's will by means of force, violence, duress, menace, or fear of immediate and unlawful bodily injury on the person or another," according to California Penal Code, section 261. Cases involving victims under the age of 18 were not sampled. Assault cases were defined as the alleged use of a deadly weapon (that includes firearm or semiautomatic weapons) or force to produce great bodily injury according to the California Penal Code Section 245 (1 and 2, a–d). Robbery was defined as "the felonious taking of personal property in the possession of another, from his person or immediate presence, and against his will, accomplished by means of force or fear" according to section 211 of the California Penal Code.

race; 98% for identification retention interval; 96% for physical evidence). Disagreements were resolved between coders before the data were entered for analysis.

Measures

Felony case issuing served as the primary outcome measure, with felony charges filed by the DA in 68% of the sampled rape cases, 73% of sampled robbery cases, and 45% of sampled assault cases. The predictors coded included suspect, crime incident and eyewitness variables.

The suspect characteristics coded were: Suspect age (continuous variable), gender (man = 1 or woman = 0), race (White = 1 or Non-White = 0) and whether the suspect had any prior felony convictions (yes = 1 or no = 0). Crime incident factors that were coded included: crime type, (rape, robbery, or assault); suspect admittance to the crime (yes = 1 or no = 0); physical evidence implicating the suspect (0 or 1, with 0 indicating no evidence and 1 indicating there was at least one piece of physical evidence, including: stolen property recovered, the weapon was found, biological evidence matched suspect); whether the suspect was arrested at the crime scene (yes = 1 or no = 0); and crime severity (0 or 1, with 1 indicating cases in which the victim was injured and/or the perpetrator used a weapon).

The eyewitness variables that were assessed in relation to charging decisions included: whether any identification procedure had been conducted (yes = 1 or no = 0); whether any positive identification of the suspect had been made by an eyewitness (yes = 1 or no = 0); whether there were any eyewitnesses in the case who did not identify the suspect when given the opportunity (yes = 1 or no = 0); the number of eyewitnesses in the case who made a positive eyewitness identification; the proportion of eyewitnesses who made a positive eyewitness identification; type of identification procedure conducted (live show-up, photo show-up, photo array, both show-up and photo array, or other); and whether the suspect was a stranger to all of the eyewitnesses in the case (yes = 1 or no = 0). We also coded whether all of the eyewitnesses in the case were of a different race than the culprit (yes = 1 or no = 0) because Gross et al. (2005) found that cross race identification errors were common in erroneous convictions (*cf* Flowe, Finklea, & Ebbesen, 2009).

Six measures related to the *Biggers* criteria for evaluating the strength of an eyewitness identification that may have affected charging decisions were derived from the files, including whether the eyewitnesses viewed the culprit from head on (as opposed to having only a side or back view of the culprit), proximity of the eyewitness(es) to the culprit, retention interval length between the crime and the identification test(s), lighting conditions during the crime, eyewitness identification confidence, and the degree to which the eyewitness description(s) matched the suspect. Each of these indicators, which are described in detail below, was derived from the files as a dichotomous variable to reflect the memory strength that might be expected given the circumstances; data were coded as 1 or 0, with 1 reflecting the expectation that memory strength would be relatively better under the given circumstance. Descriptive statistics for the raw data on which the coding of the strength indicators was based are presented in Table 1.

Table 1
Raw Descriptive Statistics and Coding Outcomes for the Eyewitness Strength Indicator Variables

	Raw data descriptive statistics				Code assigned (proportion of *N*)	
Strength indicators	*N*	Median	Mean	Range	0	1
Viewed culprit head-on	257	n/a	n/a	n/a	0.01	0.99
Proximity	272	1 foot	32.75 inches	0–2400 inches	0.36	0.64
Retention interval	270	2.88 days	7 days	0–90 days	0.42	0.58
Lighting	259	n/a	n/a	n/a	0.18	0.82
Confidence	243	n/a	n/a	n/a	0.09	0.91
Number of descriptors	185	10	10.54	0–22	0.35	0.65

If all of the eyewitnesses in the case viewed the culprit head on, the case was assigned a value of 1; otherwise, the case was assigned to a 0 to indicate that the viewing angle was from only the back or the side of the culprit for one or more of the eyewitnesses. For the proximity variable, if more than one eyewitness was tested in a case, the median value across the eyewitnesses in the case was computed to measure proximity; when proximity was averaged across all suspects, the median was 1 foot. Cases in which the identifying eyewitness(es) were a foot or less away from the culprit were assigned a value of 1, indicating that the eyewitnesses were close in proximity to the culprit, whereas the remaining cases were assigned a value of 0 to indicate that the culprit was relatively more remote for one or more of the eyewitnesses. If the information needed to code proximity was not explicitly provided in the report, the coder estimated the value of the variable if there was adequate information for so doing in the eyewitness' statement to the police.

Cases in which the eyewitness(es) attempted an identification of the culprit within 24 hours of the crime were assigned a value of 1, indicating a relatively short retention interval length, and cases in which the identifications took place more than 24 hr after the crime were assigned a value of 0 to indicate a longer retention interval length. The lighting conditions the eyewitness(es) had while viewing the crime were coded based on the police report for every case. If the crime took place either outside in broad daylight or inside in a well-lit room, the case was assigned a value of 1. Otherwise, if the crime took place outside in the dark, under street lamps, or in a dimly lit or darkened room, the case was assigned a value of 0. If all of the eyewitnesses in the case made a statement indicating that they were certain about their identification of the defendant (as opposed to indicating that the defendant resembled the culprit to some degree but they were not certain he or she actually was the culprit), the case was assigned a 1 to indicate that all of the eyewitnesses were confident in their identifications. If any of the witnesses was not positive, the case was assigned a 0 to indicate that the eyewitness(es) were relatively less confident. These factors were measured to examine whether they were associated with charging decisions.

Eyewitness descriptions of the suspect were obtained from the police crime incident report, which has a standard section for collecting 22 descriptors of the culprit's physical appearance (e.g., age, gender, race, height, weight, build, eye color, hair color, hair length, hair type, hair style, facial hair, complexion). The data are collected in a recognition format, as every physical descriptor is accompanied by response options on the form. For instance, with respect to the build of the suspect, eyewitnesses can indicate whether he is stocky, muscular, thin, etc. When the suspect is arrested, the police complete an identical description checklist on the arrest report. The correspondence between the description of the culprit given by the eyewitness and the description of the suspect given by police was determined for every descriptor (match vs. no match[2]) and the average degree of correspondence was determined for every witness. If there was more than one eyewitness in the case who gave a description of the suspect, the correspondence scores were averaged across eyewitnesses to create a single measure of description correspondence for each case. The data across cases were then subjected to a median split (median = 75% correspondence). Cases that were equal to or above the median were scored a value of 1 to indicate a relatively high degree of correspondence between the eyewitness descriptions and the suspect, whereas the remaining cases were scored a value of 0 to indicate a low degree of correspondence.

Overall eyewitness identification evidence strength was computed for every case by summing across the individual strength indices. Scores could range from 0 to 6, with 6 indicating the highest degree of strength possible on the scale.

Cases that pose the greatest risk for erroneous conviction based on mistaken eyewitness identification—and therefore, should be less likely to involve a decision to prosecute—were identified using the following criteria. First, cases in which none of the eyewitnesses were previously acquainted with the defendant (i.e., *stranger cases*) were identified. Second, within the subsample of stranger cases, cases that had eyewitness identification evidence alone and cases that had eyewitness identification evidence with limited corroboration were identified. Cases that had *eyewitness identification evidence alone* were defined as cases that did not have any physical evidence or any suspect behavioral evidence that might be indicative of guilt, including admittance to the current offense and/or a prior felony conviction. Cases that had *eyewitness identification evidence with limited corroboration* were defined as cases that did not have physical evidence but that did have suspect behavioral evidence that might be indicative of guilt, including admittance to the current offense and/or a prior felony conviction.

Results

Preliminary Results

Suspect and crime incident profile. Table 2 provides descriptive characteristics for each crime category and for the overall case sample (see *Total* column). With respect to the overall sample, suspects were typically male (93% of cases), 29

[2] The descriptors were coded as matching using the following criteria: age values within ±5 years, height within ±2 inches, and weight within ±10 pounds, and the remaining descriptors had to correspond exactly to be coded as a match.

Table 2
Descriptive Characteristics of Sample and Zero-Order Correlations With Felony Case Prosecution by Crime Type

Variable	Assault			Rape			Robbery			Total		
	Mean	Prosecution	*N*	Mean	Prosecution	*N*	Mean	Prosecution	*N*	Mean	Prosecution	*N*
Suspect characteristics												
Age[b]	33.1	0.12	154	29.2	−0.08	266	26.9	−0.15*	220	29.4	−0.10*	640
Male[a]	0.90	0.12	154	0.99	0.18	266	0.89	−0.02	223	0.93	0.08*	643
White[a]	0.39	0.00	155	0.28	−0.13*	260	0.25	0.06	217	0.29	−0.06	632
Prior felony record[a]	0.53	0.60***	184	0.49	0.33***	301	0.63	0.35***	239	0.55	0.40***	724
Incident characteristics												
Severity[1]	0.77	0.28***	184	0.75	0.51***	302	0.65	0.28***	239	0.69	0.31***	725
Physical evidence[a]	0.52	0.23**	184	0.46	0.38***	302	0.44	0.17*	239	0.47	0.25***	725
Suspect admit[a]	0.39	0.12	184	0.41	0.47***	302	0.27	0.11	239	0.36	0.24***	725
Suspect arrested at crime scene[a]	0.29	−0.28***	184	0.10	0.06	302	0.21	−0.02	239	0.18	−0.10*	725
Eyewitness characteristics												
All strangers to suspect[a]	0.29	0.11	184	0.12	0.08	302	0.62	0.27***	239	0.33	0.16***	725
Any positive IDs[a]	0.29	0.29***	184	0.22	0.09	302	0.53	0.14*	239	0.34	0.17***	725
Number of positive IDs[b]	1.36	0.19	61	1.11	0.03	70	1.35	0.00	139	1.29	0.05	270
Any non-IDs of suspect[a]	0.21	−0.17	61	0.08	0.16	70	0.21	0.04	139	0.18	0.00	270
Proportion of positive IDs[b]	0.83	0.19	61	0.93	−0.15	70	0.84	−0.02	139	0.86	0.02	270
Cross race eyewitnesses[a]	0.32	0.21*	184	0.30	0.18*	302	0.52	0.20*	239	0.30	0.21***	725

[a] Variable is dichotomous, coded as "1" if present and "0" if absent; therefore, the mean equals the proportion of *N* cases in which the factor was equal to 1. [b] Variable is continuous.
* Correlation is significant at the 0.05 level (two-tailed). ** Correlation is significant at the 0.01 level (two-tailed). *** Correlation is significant at the 0.001 level (two-tailed).

years of age (range: 16–79 years), and were about equally likely to be White (30%), Black (33%), or Hispanic (32%) (6% were classified as Native American, Asian or as belonging to other racial categories). More than half of the suspects had a previous felony conviction (range: 0–22 felony convictions) and the majority (67%) were acquainted with one or more of the eyewitnesses in the case.

Felony charging was positively associated with the suspect having a prior felony conviction. Not surprisingly, given the crime categories coded, most often the criminal action in question was severe, as 69% of the suspects were alleged to have used a weapon and/or to have injured a victim. Cases that were accepted for prosecution were more likely to have severe circumstances compared to rejected cases. About half of the cases had physical evidence that tied the suspect to the crime, and the presence of physical evidence was associated with an increased likelihood of prosecution. In one-third of the cases the suspect admitted to the crime. Admittance to the crime was positively associated with prosecution in only rape cases.

A total of 237 defendants were classified as stranger cases. For 153 of these defendants (64%) there was at least one eyewitness who was given an identification test; 7% of the stranger cases (16/237) were eyewitness identification alone cases, and the suspect was charged with a felony 56% of the time (9/16). Among these defendants, 6 pleaded guilty, 1 was found guilty at trial, and the charges were dropped in 2 cases. An additional 20% of the stranger cases (48/237) were eyewitness identification evidence with limited corroborating evidence cases. Felony charges were issued in 83% of these cases (40/48). Among these defendants, 31 pleaded guilty, 8 were found guilty at trial, and the charges were dropped in 1 case. Therefore, across cases in which the suspect was prosecuted (n = 465), 2% (16/465) were *eyewitness identification alone* cases, and 9% (40/465) were *eyewitness identification evidence with limited corroborating evidence* cases.

Eyewitness identification evidence profile. Table 2 also presents the proportion of cases for which any positive identification of the suspect was made, whether by an acquaintance or a stranger. Across the sample, one out of every three suspects had positive ID evidence in their case. Having had at least one positive identification of the suspect by an eyewitness was positively associated with felony issuing for every crime type.

The association between positive identification evidence and case issuing decisions was examined next. At least one eyewitness made an identification of the suspect in 64% of stranger cases (115/179) and in 50% of rejected cases (29/58); the association between case issuing and positive identification evidence was marginally statistically significant in stranger cases, $\chi^2(1) = 3.16$, $p = .07$. In acquaintance cases, 33% of accepted cases (71/286) and 14% (30/202) of rejected cases had a positive identification of the suspect by at least one eyewitness. The association between case issuing and positive suspect identification was statistically significant in acquaintance cases, $\chi^2(1) = 6.58$, $p < .05$.

An identification test was likelier under the following circumstances: there was physical evidence implicating the suspect ($r = .10$, $p < .05$, two-tailed); the suspect had a prior record ($r = .13$, $p < .0001$, two-tailed); the suspect did not admit to the crime ($r = .13$, $p < .0001$, two-tailed); none of the eyewitnesses knew the suspect ($r = .38$, $p < .0001$, two-tailed); all of the eyewitnesses were

cross race with respect to the suspect ($r = .23$, $p < .0001$, two-tailed); and the suspect was caught at the crime scene ($r = .38$, $p < .0001$, two-tailed). None of the other suspect or crime incident factors were associated with whether an identification test was performed.

Table 3 provides additional information about the characteristics of the identification tests, conditioning the results by acquaintance ($n = 120$) versus stranger ($n = 153$) ID cases. Suspects presented for identification appeared largely in live show-ups followed by photo arrays for both stranger and acquaintance cases.

Eyewitness Evidence and the Decision to Prosecute as a Felony

Table 4 reports the findings from two separate models that examine the likelihood of prosecution in relation to suspect, crime incident, and eyewitness variables. A separate model was performed for stranger and acquaintance cases. Logistic regression analysis was used because the outcome variable is dichotomous (Hosmer & Lemeshow, 1989). The eyewitness evidence variable included in the analyses was whether there was a positive identification in the case because the univariate results indicated that this eyewitness measure was the one that was most strongly related to case issuing outcomes. As the number of cases was not large, the variables included in the analysis were only those that were found to have a strong univariate relationship with case issuing (see Table 2). Crime type was also entered into the models using reference cell coding, with robbery as the reference category. Robbery was designated as the reference category because the descriptive analysis presented earlier (see Table 2) indicated that the characteristics of assault and rape cases were more similar to each other than they were with robbery cases.

Stranger cases. A model with the predictors significantly fit the data better than a model without any predictors, $\chi^2(7) = 73.92$, $p < .01$. The odds that a suspect was charged with a felony were about 8 times higher if the suspect had a prior record. Admittance to the crime and physical evidence also increased the odds that the case was prosecuted. Felony charges were less likely to be brought in assault compared to robbery cases; felony issuing did not differ for rape compared to robbery cases. Positive eyewitness identification evidence was associated with an increase in the odds that a case was prosecuted, but the association was not significant ($p = .14$). The overall model correctly predicted 55% of cases that were rejected and 97% of cases that were prosecuted.

Table 3
Eyewitness Identification Procedure Distributions for Stranger and Acquaintance Cases

ID Procedure	Stranger ($n = 153$)	Acquaintance ($n = 120$)
Live show-up	0.59	0.50
Photo array	0.30	0.41
Multiple	0.05	0.02
Other	0.06	0.08

Table 4
Logit Estimates of Suspect, Incident, and Eyewitness Variables on Likelihood of Prosecution for Stranger and Acquaintance Cases

Variable	Stranger			Acquaintance		
	Coefficient	*SE*	Odds	Coefficient	*SE*	Odds
Suspect characteristics						
Prior felony record	2.08***	0.38	8.01	1.74***	0.25	5.68
Suspect admit	1.41**	0.49	4.12	1.34***	0.26	3.82
Incident characteristics						
Severity	0.71	0.41	2.03	1.43***	0.27	4.21
Physical evidence	0.82*	0.38	2.27	1.16***	0.25	3.20
Crime type						
Assault	−1.75***	0.45	0.17	−1.61***	0.36	0.20
Rape	−0.40	0.52	0.67	0.33	0.32	1.39
Eyewitness characteristics						
Any positive ID	0.57	0.39	1.77	0.63*	0.30	1.89
Constant	−0.98	0.50		−2.31	0.36	
−2 Log likelihood	189.84			435.30		
Cox & Snell R^2	0.27			0.37		

* $p < .05$. ** $p < .01$. *** $p < .001$.

Acquaintance cases. A model with the predictors significantly fit the data better than a model without any predictors, $\chi^2(7) = 224.92, p < .01$. The odds that a suspect was charged with a felony were over 5 times higher if the suspect had a prior record. Severe allegations, admittance to the crime, and physical evidence also significantly increased the odds that the case was prosecuted. Eyewitness identification evidence significantly increased the odds that a case was prosecuted. Felony charges were less likely to be brought in assault compared to robbery cases; rape did not differ from robbery case issuing. This model correctly predicted 67% of cases that were rejected and 85% of cases that were prosecuted.

Eyewitness Identification Evidence Strength

Eyewitness identification strength in accepted versus rejected cases. Table 5 provides eyewitness identification evidence strength by case issuing decision. Overall strength of the eyewitness identification evidence did not dif-

Table 5
Eyewitness Identification Strength by Case Issuing Decision (Issued vs. Rejected)

	Issued (n = 186)	Rejected (n = 59)
Overall strength	4.33	4.19
Viewed culprit head-on	0.99	0.98
Close proximity	0.63	0.63
Short retention interval	0.62	0.59
Good lighting	0.83	0.76
High confidence	0.93	0.89
Description corresponded	0.64	0.58

ferentiate prosecuted from rejected cases ($M = 4.33$ and $M = 4.19$, respectively), nor could case issuing decisions be distinguished on the basis of any of the individual strength indicators. As can be seen in Table 5, the majority of accepted and rejected cases had eyewitnesses with reasonably good viewing conditions and witnesses who expressed high confidence in their identification.

Eyewitness identification strength in eyewitness identification alone cases. Overall eyewitness evidence strength was further examined across prosecution outcomes in the eyewitness identification alone cases. The UCCSs suggest eyewitness identification alone cases should be scrutinized by the prosecutor to assess the validity of the evidence. The overall strength of the eyewitness evidence was higher for charged compared to not charged eyewitness identification alone cases ($M = 4.73$ vs. $M = 3.62$, respectively), $t(25) = 2.76$, $p < .05$, two-tailed. Each of the individual strength indicators was tested to explore whether there was a given factor that best distinguished issued from rejected cases in terms of eyewitness identification strength. No statistically significant differences for any of the individual indicators emerged.

Show-Ups versus lineups. The UCCS advise that charges in eyewitness identification alone cases should be issued when a lineup has been conducted. Accepted and rejected cases, however, could not be differentiated on the basis of identification procedure type. Live show-ups were the most common procedure in both prosecuted (64%) and rejected cases (57%).

Eyewitness identification outcomes and charging decisions. Whether a case had any nonidentifications was unrelated to whether the case was prosecuted in stranger as well as in acquaintance cases. Additionally, charging decisions were not related to either the number of positive eyewitness identifications in the case or the proportion of positive eyewitness identifications in the case. These results are summarized in Table 2.

Discussion

Suspect and case factors that have been identified by previous archival research as influential in felony charging decisions were significantly associated with case issuing decisions in the present study. Eyewitness identification evidence was present more often in cases that were issued compared to cases that were rejected for prosecution, although the association between eyewitness identification evidence and case issuing was statistically significant only in acquaintance cases, not in stranger cases. Defendant prior record, crime severity, admittance to the crime and physical evidence were associated to a larger extent with case outcomes compared to eyewitness identification evidence. Additionally, overall, the strength of the eyewitness identification evidence was not associated with charging decisions. However, in the cases in which eyewitness identification evidence was the sole evidence against the defendant, accepted cases scored higher on the *Biggers* criteria compared to those that were rejected. This result suggests that prosecutors were attempting to evaluate the strength of the eyewitness identification evidence in making issuing decisions when there was only eyewitness identification evidence linking the suspect to the crime. Finally, multiple identifications of the defendant by more than one eyewitness and nonidentifications of the defendant were not associated with charging decisions. The

following sections discuss these findings and the additional research questions they raise.

Association Between Eyewitness ID Evidence and Felony Issuing

In both stranger and acquaintance cases, positive eyewitness identification evidence was associated with increased odds of prosecution. The association between eyewitness identification evidence and case issuing, however, was stronger and statistically significant only in acquaintance cases. In part, this may have resulted because statistical power was greater in the acquaintance case analysis because the sample size was relatively larger compared to stranger cases. The result would also arise, however, if prosecutors tend to scrutinize eyewitness identification evidence in stranger cases. Mistaken eyewitness identification concerns are of course greater in stranger compared to acquaintance cases. Therefore, on one level the fact that eyewitness testimony plays a relatively small role in case issuing decisions is a desirable result. On the other hand, the association between eyewitness identification evidence and case prosecution is likely to vary across jurisdictions depending on the procedures that are in place for determining whether a suspect should be presented in an eyewitness identification test and the weight that should be given to a positive eyewitness identification of the suspect.

Further research is needed to determine the role that eyewitness identification evidence plays in apprehending the suspect. Additional questions that remain include: Is eyewitness identification evidence the "icing on the cake" in determining the guilt of a suspect in the mind of the investigators? Does it play a pivotal role in leading the police to continue (or to stop) collecting evidence against a suspect? At what point during the course of an investigation does eyewitness identification evidence get collected? In the present study, other forms of evidence, such as physical evidence and suspect prior felony conviction, were positively associated with whether there was an identification test conducted in the case. Additional research is needed to determine how crime and suspect factors influence whether the suspect is shown to the eyewitness(es) for identification. This issue is important to investigate because whether eyewitness identification evidence provides additional information regarding the suspect's guilt depends on the extent to which it is independent from other case evidence.

Case Prosecution and the Strength of Eyewitness Identification Evidence

Relatively few cases were prosecuted based solely on eyewitness identification evidence. Results indicated that most of the cases evaluated for prosecution had other forms of evidence in addition to (or instead of) eyewitness identification evidence implicating the suspect. This finding suggests that the police tended to not forward a suspect for prosecution unless there was other evidence in the case. Another factor contributing to the low number of cases that relied solely on eyewitness identification evidence is that quite often the suspects in the sample knew one or more of the eyewitnesses in their case. In acquaintance cases the identity of the suspect may not be much of an issue, although acquaintances can also make incorrect identifications.

There was some evidence to suggest that prosecutors may be scrutinizing the quality of eyewitness identification evidence, albeit in limited circumstances. The

strength of the eyewitness identification evidence did not generally vary across accepted and rejected cases. However, in a subset of cases that were arguably relatively weak from an evidentiary standpoint, the eyewitness identification evidence was stronger in accepted compared to rejected case. In these cases, the suspect was positively identified, none of the eyewitnesses knew the suspect, there was no physical evidence, and the suspect did not admit to the crime and did not have a prior record. In cases in which the suspect admitted to the crime and/or had a prior record, the strength of the identification evidence did not vary across prosecution outcomes. These findings support the conclusion that in the absence of other evidence, such as suspect prior record and physical evidence, prosecutors are taking into account the strength of the eyewitness' testimony as suggested by the UCCS. However, there were few cases in which eyewitness identification evidence was the sole evidence in the case; therefore, the strength of the conclusion would be served well by replication of these findings.

Interestingly, live show-ups were the most often used identification procedure. The prevalence of show-ups versus lineups did not vary in relation to prosecution outcomes, suggesting that prosecutors do not place greater weight on one procedure versus another. Few studies have investigated live show-ups and further research on this topic seems warranted. On the one hand, a positive identification from a live show-up could be strong evidence of the suspect's guilt. An eyewitness who positively identifies a suspect from a live show-up usually does so at the crime scene, immediately after the crime. Memory strength is likely to be higher than if the eyewitness were given a lineup test, which is usually administered on average 5 days later for robbery cases and 18 days later for assault cases (Flowe et al., 2009). Research has demonstrated that identification accuracy usually declines with the passage of time (e.g., Shapiro & Penrod, 1986). Another benefit of live show-ups is that the test is sometimes given in the context in which the crime occurred, which could facilitate memory retrieval. Finally, if the police can test the eyewitness at the crime scene, this obviates the need for having to locate and liaise with a witness for an identification test at a later time. On the other hand, there are no foils in a show-up; therefore, investigators may not be able to distinguish between positive identifications that are made based on the eyewitness' memory for the culprit versus those that are made based on guessing alone (Steblay et al., 2003). Clearly additional research is needed to evaluate the best strategy for identifying suspects at the crime scene, such as the use of mobile lineup devices (e.g., MacLin & Phelan, 2007).

Which Factors Should Prosecutors Take into Account in Evaluating the Strength of Eyewitness Identification Evidence?

The findings raise the possibility that prosecutors take into account a range of indicators to assess the strength of the eyewitness identification evidence. None of the individual indicators of eyewitness identification evidence strength by itself was associated with case outcomes. The strength indicators when considered as a unit were associated with case outcomes in relatively weak cases. These findings are in keeping with previous research that found that participant evaluations of eyewitness evidence strength were related to the *Biggers* criteria in a summative fashion (Bradfield & Wells, 2000). More systematic work that evaluates each of

the strength indicators, however, is needed, as the indicators of strength measured in the present study were not independent of one another. For instance, cases in which the eyewitnesses were relatively confident in their identification probably had better viewing conditions overall. More systematic work on the issue in field settings could lead to the specification of a strength marker(s) this is widely available and that can be reliably measured. Such a marker would assist prosecutors in the evaluation of eyewitness identification strength. The importance of this work is highlighted by the fact that most of the defendants in the eyewitness identification evidence alone cases pleaded guilty. If the defendant pleads guilty, then trial safeguards for protecting defendants from being convicted on the basis of erroneous eyewitness identification are of course rendered impotent.

Multiple Eyewitness Identifications and Nonidentifications and Felony Issuing

Clark and Wells (2008) demonstrated that multiple identifications of the suspect by eyewitnesses increase the odds that the suspect is guilty, whereas nonidentifications decrease the odds that the suspect is guilty. Their analysis further demonstrated that in cases in which there are multiple eyewitnesses who have been presented with a lineup test, the odds that a suspect is guilty are affected to a larger extent by a nonidentification compared to an additional positive identification, all other things being equal. Therefore, multiple identifications and especially nonidentifications should be taken into account by prosecutors. Results in the present study, however, indicated that case issuing outcomes were not associated with whether multiple eyewitnesses identified the suspect or with whether there were any eyewitnesses who did not identify the suspect. Additional research is needed to determine how police and prosecutors evaluate cases that have multiple eyewitnesses. As pointed out by Clark and Wells (2008), the degree to which a nonidentification affects the prior odds of suspect guilt depends on variables such as the nonidentifying eyewitness' memory strength. For example, if an eyewitness with relatively good viewing conditions positively identifies the suspect while an eyewitness with relatively poor viewing conditions does not identify the suspect, then the posterior odds that the suspect is guilty would be reduced little by the nonidentification. In view of this consideration, future research could examine whether police and prosecutors weigh the validity of eyewitness identification evidence in relation to the degree of interaction that the eyewitness had with the culprit.

Limitations

The nature of archival research makes it difficult to draw cause and effect conclusions from the findings. First, our results may be specific to the location and/or the District Attorney's Office that we examined. It is highly plausible that the relationship between eyewitness identification evidence and charging decisions varies across jurisdictions. Additional studies are needed in several other jurisdictions to more fully understand how legal officials use eyewitness identification evidence in deciding whether to prosecute. Second, the degree to which we are able to make inferences about legal decision making using criminal archives depends on the validity and completeness of the information reported by

law enforcement officials in the archival records. Third, we sampled only rape, robbery and assault cases. Eyewitness identification evidence may have a differential effect on case processing depending on crime type. In murder cases, for example, there tends to be enormous pressure to bring a perpetrator to justice. Would charges in murder cases often be issued even if there was only eyewitness identification evidence? Fourth, eyewitness identification evidence in a rejected case may have an effect on case processing if the case is later re-opened. For example, we encountered one case in the sample in which the credibility of a rape victim was called into question because she was not forthright with the police about having seen the perpetrator (who was a stranger) on the bus earlier on the day that she was reportedly raped. The case was rejected partly on this basis. The suspect was later arrested for raping four other women. The first case was re-opened in light of the new victim reports and physical evidence. There were so few of these types of cases that we could not systematically analyze them. However, it is important to point out that just because a case has been rejected does not mean that it will not be prosecuted in the future.

In summary, prosecutorial discretion to bring felony charges against a police suspect is a potential procedural safeguard that could help protect innocent suspects from being convicted on the basis of erroneous eyewitness identification. Results from the current study indicated that physical evidence and suspect factors were more strongly associated with case issuing decisions than eyewitness identification evidence. Additional research is needed to determine markers of identification accuracy that can be used by prosecutors in assessing the strength of eyewitness identification evidence.

References

Albonetti, C. (1987). Prosecutorial discretion: The effects of uncertainty. *Law and Society Review, 21,* 291–313. doi: 10.2307/3053523

Behrman, B. W., & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior, 25,* 475–491. doi: 10.1023/A:1012840831846

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,* 115–130. doi: 10.1002/acp.1171

Borchard, E. M. (1932). *Convicting the innocent.* Garden City, NY: Garden City Publishing Company, Inc.

Bradfield, A., & Wells, G. (2000). The perceived validity of eyewitness identification testimony: A test of the five Biggers criteria. *Law and Human Behavior, 24*(5), 581–594. doi: 10.1023/A:1005523129437

Brigham, J. C. (1981). The accuracy of eyewitness evidence: How do attorneys see it? *The Florida Bar Journal, November,* 714–721.

California District Attorney's Association. (1996). Uniform Crime Charging Standards.

Clark, S. E., & Wells, G. L. (2008). On the diagnosticity of multiple-witness identifications. *Law and Human Behavior, 32,* 406–422. doi: 10.1007/s10979-007-9115-7

Connors, E., Lundregan, T., Miller, N., & McEwen, T. (1996). *Convicted by juries, exonerated by science: Case studies in the use of DNA evidence to establish innocence after trial.* US Department of Justice Office of Justice Programs.

Deffenbacher, K. A., Bornstein, B. H., Penrod. S. D., & McGorty, E. K. (2004). A

meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior, 28,* 687–706. doi: 10.1007/s10979-004-0565-x

Devenport, J. L., Kimbrough, C. D., & Cutler, B. L. (2009). Effectiveness of traditional safeguards against erroneous conviction arising from mistaken eyewitness identification. In B. L. Cutler (Ed.), *Expert testimony on the psychology of eyewitness identification* (pp. 51–68). New York: Oxford University Press (American Psychology-Law Society Series). doi: 10.1093/acprof:oso/9780195331974.003.003

Feeney, F., Dill, F., & Weir, A. (1983). *Arrests without conviction: How often they occur and why.* Washington, DC: Government Printing Office.

Flowe, H. D., Finklea, K. M., & Ebbesen, E. B. (2009). Limitations of expert psychology testimony on eyewitness identification. In B. L. Cutler (Ed.), *Expert testimony on the psychology of eyewitness identification* (pp. 220–221). New York: Oxford University Press (American Psychology-Law Society Series). doi: 10.1093/acprof:oso/9780195331974.003.003

Goldstein, A. G., Chance, J. E., & Schneller, G. R. (1989). Frequency of eyewitness identification in criminal cases: A survey of prosecutors. *Bulletin of the Psychonomic Society, 27,* 71–74.

Gross, S. R., Jacoby, K., Matheson, D. J., Montgomery, N., & Patel, S. (2005). Exonerations in the United States, 1989–2004. *The Journal of Criminal Law & Criminology, 95,* 523–560.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression.* New York: John Wiley & Sons Inc.

Jacoby, J., Mellon, L., Ratledge, E., & Turner, S. (1982). *Prosecutorial decision making: A national study.* Washington, DC: Department of Justice, National Institute of Justice.

Kingsnorth, R., MacIntosh, R., & Sutherland, S. (2002). Criminal charge or probation violation? Prosecutorial discretion and implications for research in criminal court processing. *Criminology: An Interdisciplinary Journal, 40,* 553–578. doi: 10.1111/j.1745-9125.2002.tb00966.x

MacLin, O. H., & Phelan, C. M. (2007). PC_Eyewitness: Evaluating the New Jersey method. *Behavioral Research Methods, 39,* 242–247.

Magnussen, S., Melinder, A., Stridbeck, U., & Raja, A. (2010). Beliefs about factors affecting the reliability of eyewitness testimony: A comparison of judges, jurors and the general public. *Applied Cognitive Psychology, 24,* 122–133. doi: 10.1002/acp.1550

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, and Law, 7,* 3–35. doi: 10.1037/1076-8971.7.1.3

Miller, F. (1969). *Prosecution: The decision to charge a suspect with a crime.* Boston: Little Brown.

Nagel, I. H., & Hagan, J. (1983). Gender and crime: Offense patterns and criminal court sanctions (Vol. 4, pp. 91–144). In M. Tonry & N. Morris (Eds.). *Crime and justice: An annual review of research.* Chicago: University of Chicago Press.

People v. Aguilar, B194509; Cal. App. Unpub. LEXIS 691 (2008).

People v. Cummings, B207053; Cal. App. Unpub. LEXIS 8678 (2009).

People v. Figueroa, B199625; Cal. App. Unpub. LEXIS 3529 (2008).

People v. Guerra, B208460; Cal. App. Unpub. LEXIS 8099 (2009).

People v. Hernandez, B170634; Cal. App. Unpub. LEXIS 9939 (2004).

People v. Ledesma, B178532; Cal. App. Unpub. LEXIS 1522 (2006).

People v. Lopez, B163955; Cal. App. Unpub. LEXIS 22 (2004).

People v. Magana, D039298; Cal. App. Unpub. LEXIS 6614 (2003).

People v. Robles, B209644; Cal. App. Unpub. LEXIS 8099 (2009).

People v. Rountree, D039841; Cal. App. Unpub. LEXIS 2383 (2005).
People v. Rubalcava, H027786; C. A. LEXIS 11916 (2005).
People v. Trujillo, H032260; Cal. App. Unpub. LEXIS 8304 (2009).
Piehl, A. M., & Bushway, S. D. (2007). Measuring and explaining charge bargaining. *Journal of Quantitative Criminology, 23,* 105–125. doi: 10.1007/s10940-006-9023-x
Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100,* 139–156. doi: 10.1037/0033-2909.100.2.139
Silberman, C. E. (1978). *Criminal violence, criminal justice.* New York: Random House.
Smith v. Smith, 02 Civ. 7308 (WHP)(JCF); U.S. Dist. LEXIS 17642 (2003).
State of Washington v. Riofta, No. 28209–7-II; Wash. App. LEXIS 1880 (2003).
Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2003). Eyewitness accuracy rates in police showup and lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 27,* 523–540. doi: 10.1023/A:1025438223608
Steblay, N. M. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior, 16,* 413–424. doi: 10.1007/BF02352267
U.S. Department of Justice. (1998). *German and American prosecutions: An approach to statistical comparison.* Washington, DC: U.S. Department of Justice.
U.S. Department of Justice. (2001). *Felony sentences in state courts, 1998.* BJS Bulletin, October, NCJ190103.
U.S. Department of Justice, Federal Bureau of Investigation. (1996). *Crime in the United States: Uniform Crime Reports 1995.* Washington, DC: U.S. Department of Justice.
United States v. Burton, No. 3:97-cr-154; U.S. Dist. LEXIS 18730 (1998).
United States v. James, NMCM 98 02036; CCA LEXIS 64 (2001).
Wells, G., Memon, A., & Penrod, S. (2006). Eyewitness evidence: Improving its probative value. *Psychological science in the public interest, 7,* 45–75. doi: 10.1111/j.1529-1006.2006.00027.x
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,* 581–598. doi: 10.1037/0003-066X.55.6.581
Wells, G. L., & Murray, D. M. (1983). What can psychology say about the Neil v. Biggers criteria for judging eyewitness accuracy? *Journal of Applied Psychology, 68,* 3, 347–362. doi: 10.1037/0021-9010.68.3.347

Received May 12, 2009
Revision received June 15, 2010
Accepted June 21, 2010 ■

Psychology, Public Policy, and Law
2001, Vol. 7, No. 1, 253–262

Copyright 2001 by the American Psychological Association, Inc.
1076-8971/01/$5.00 DOI: 10.1037//1076-8971.7.1.253

DISCOUNTING THE ERROR COSTS
Cross-Racial False Alarms in the Culture of Contemporary Criminal Justice

James M. Doyle
Carney and Bassil

Research regarding own-race bias (ORB) is examined by focusing on results that indicate that Whites may apply more lenient criteria to the recognition or identification of Blacks, resulting in a higher rate of false-alarm responses. The practical context of the forensic identification task is reviewed to assess whether the more lenient criteria applied by witnesses who are attempting identifications of Blacks resonate with the criminal justice system's tendency to apply more lenient criteria to the conviction and sentencing of Blacks. Practical implications and future research directions are considered.

Are White eyewitnesses simply more willing to guess where the identification of a Black suspect is concerned? Is there an endemic tendency among White American eyewitnesses to discount the *error cost*—defined as "the product of . . . the probability of error, and the cost if an error occurs" (Posner, 1973, p. 399)—implicit in the identification of an African American suspect? Is this what one ultimately means when one says that Whites apply more liberal response criteria to the task of recognizing Blacks (Meissner & Brigham, 2001; Sporer, 2001a)? There are questions of deeper theoretical interest, but from a practical, forensic perspective, the possibility of increased other-race guessing has claims for emergency status on the research agenda.

Is it possible that the cultural situation of contemporary White witnesses has persuaded them that they are less likely to be wrong if they identify a Black suspect? Or convinced them that they can be less concerned by the harm that results if they do misidentify a Black suspect? Or both? Do Whites see the probability of error as lower in cross-racial situations and the cost of error as easier to accept? Nothing in the data gathered in this issue establishes the truth of these conjectures, but nothing has yet ruled them out either, and there are more than a few findings that suggest that an inquiry is in order. The picture is a depressing one.

The Identification in Context

In the Innocence Project's latest study of wrongful convictions uncovered by DNA analysis, 36% of the wrongful convictions derived from Whites misidentifying Blacks (Dwyer, Neufeld, & Scheck, 2000). Researchers have reported that White Europeans demonstrated a significantly larger own-race bias (ORB) when

Correspondence concerning this article should be addressed to James M. Doyle, Attorney at Law, Carney and Bassil, 20 Park Plaza, Suite 800, Boston, Massachusetts 02116. Electronic mail may be sent to jamesdoyle@sprintmail.com.

compared with Black African participants (Meissner & Brigham, 2001). Further analysis has indicated that this ORB effect among Whites stems largely from false-alarm responses (Meissner & Brigham, 2001). Moreover, while other measures of ORB have been decreasing over time, measures of false-alarm responses have been increasing. Meissner and Brigham's (2001) meticulous meta-analysis of past studies concluded that "over the past 3 decades, the ORB effect appears to have become most prominent in false-alarm responses" (p. 22). That conclusion is difficult to reject when 11 of 14 samples, drawn from six studies, indicated that more liberal response criteria were applied in cross-race situations. False alarms, remember, result in prison sentences, even death sentences.

In the contemporary American criminal justice system, an eyewitness guess (if that is what it is) occupies a central place in a criminal process composed of a cascading series of events, each augmenting the dangers of any guess that went before or comes after and all of which have been shown by one study or another to be racially sensitive (Morris, 1988). The error cost calculations I have described can be applied to a whole catalogue of racially weighted decisions that extend from the street cop's decision to stop a suspect to the Supreme Court's decision to affirm his death sentence (*McCleskey v. Kemp,* 1987; Morris, 1988). The practice of racial profiling, for example, is rational if one believes that stopping Black drivers to search their cars for drugs is less likely to be a mistake or if one believes that the cost of the mistake is less serious. It is especially rational if one believes both of these things. Racial profiling is not rational—although it could be a powerful expression of irrational bigotry—unless one believes these things.

The startlingly elevated rates of incarceration for Black adolescents could indicate that throwing a young Black life away is perceived to be less likely to be a mistake or that the mistake matters less than it might with a White adolescent (Bridges & Steen, 1998). Providing inadequate defense counsel and investigative services to indigent Black defendants, incarcerating the young Black male population at a lavish rate for exaggerated terms (Mauer & Huling, 1995), imposing a draconian system of mandatory minimum drug sentences with disparate impact on Blacks, executing Black murderers of Whites at a rate 13 times the rate of White murderers of Black victims (*McCleskey v. Kemp,* 1987)—all of these inequities continue to mount long after claw-and-fang racism in the Ku Klux Klan style has gone out of fashion.

Cross-racial eyewitness cases provide a window on these issues because, in eyewitness cases, racism of the old-fashioned sort is beside the point by definition. An eyewitness to a crime never faces the question "Do I identify a White or a Black?" but rather "Which Black (or which White) do I identify?" Paradoxically, analysis of the eyewitness situation—because race is apparently not salient in the most flagrant way (Sommers & Ellsworth, 2001)—stands an excellent chance of illuminating the litany of racially weighted choices I have described. The eyewitness research can provide an opportunity to test whether the racially imbalanced results in the justice system flow from the fatalistic acceptance of the cost of error in crimes where the perpetrator was undoubtedly Black and the universe of choices was entirely Black too. Are White discounts of error costs a necessary permissive factor? Do they inflict as much damage in their way as old-school racism?

I am guilty here of pulling one thread from the impressive fabric of research displayed in this issue and perhaps of placing more weight on that thread than it can bear. As Wells and Olson (2001) observed, I cannot point to anything resembling proof of this hypothesis. Still, since a false-alarm identification is an indispensable predicate to a wrongful conviction in an eyewitness-based prosecution, it does seem worth asking whether conscientious research from a system-variable perspective might counteract racially driven criterion shifts, reduce false alarms, and show valuable results in the real world. Before that can happen, the focus must be tightened. Even the tightened focus reveals daunting challenges.

Attitudes and Contact

The role of racial attitudes and interracial contacts in ORB is endlessly intriguing, but perhaps the point should be made that it is less immediately important to the forensic setting than it might seem. In part, this is a matter of sequence. Improved perceptual learning and more reliable encoding resulting from increased interracial contact will produce only evanescent improvements if the actual identification task follows the encoding—as it always will in the forensic context—and the identification task is approached with criteria so relaxed that they aggravate the number of false alarms.

Among the more provocative findings assessed in this issue is that although cross-racial contacts are rising and racial attitudes are generally felt to be improving, the number of false alarm responses from White Europeans attempting to recognize other race individuals is rising (Meissner & Brigham, 2001). Putting aside for a moment the direction of the change, the fact that there is a change in any direction across this relatively brief historical period argues that one should hesitate to embrace explanations intrinsic to human memory before ruling out dynamic explanations rooted in the culture. It seems, at least to amateur eyes, that the existing data on attitudes and contact fail to take adequate account of key aspects of the problem.

This is especially important to a distinct group that is at acute risk for misidentification: the young, urban, African American men depicted in police mug shots, who are likely to be displayed to eyewitnesses in a photo array.

The demise of Jim Crow segregation laws means that Blacks and Whites experience more face-to-face contact, but it does not necessarily follow that those face-to-face contacts are the most influential element of the White witnesses' view of the other race. The fact that general racial attitudes have improved since the days of *Uncle Tom's Cabin,* or even *To Kill A Mockingbird,* does not mean that attitudes toward that particular, young male subgroup are consistent with the general trend.

Face-to-face contact with African Americans constitutes a small fraction of White America's contact with African American faces—or at least of White America's contact with representations of African American faces. One of the central features of the African American experience has always been that of being represented by others. Every American, Black and White, has seen, as Henry Louis Gates (1994) put it,

> Tens of thousands of the most heinous representations of Black people, on children's games, portable savings banks, trade cards, postcards, calendars, tea cosies, napkins . . . nothing but images of Black people devoid of reason, simian or satanic in appearance, and slothful, lustful, or lascivious in nature. (p. 11)

This endemic societal tendency is particularly noticeable in contemporary portrayals of the young men enmeshed in the forensic identification process. As Tricia Rose (1994) pointed out,

> The White American public, many of who only tangentially know any Black men personally, has been inundated with images of young Black men who appear fully invested in a life of violent crime, who have participated in drug-related gang shootings and other acts of violence for "no apparent reason." (p. 153)

It might be argued that the shift in stereotypes from the 19th century's grinning minstrel show "coons" to today's glowering gangsta' rappers shows progress of a kind, but, if so, it is not progress that is likely to heighten White witnesses' sense of the cost of potential errors in identification. The qualitative change in stereotypes raises the question whether data about racial attitudes generated when the older stereotype prevailed have any usefulness in understanding the current situation or whether older studies (when for example, they are included in a meta-analysis) simply obscure the issue. Whatever these media-created contacts do with one hand to enhance perceptual learning and identification accuracy—and studies showing improved performance by hardcore NBA basketball fans indicate that they can (Li, Dunning, & Malpass, 1988)—they may take away with the other hand when they decrease the identifying witnesses' fears of misidentification. To put it another way, the research must recognize that the ability to identify faces of another race is a function not simply of the absolute amount of exposure one has had to members of that race but also of the quality of that exposure. (MacLin, Malpass, & MacLin, 2001). Our concept of quality of the contact must encompass the substance of media portrayal, not simply the duration of exposure or the degree of attention.

If there is a cohort effect in action here, as Chance and Goldstein (1996) posited, the effect on the cohort that has raised the number of false alarms is more likely to have its source in contemporary media images than in the legacy of the civil rights movement. Face-to-face contact with the mature African American professional who sits in the next office cubicle is not likely to ameliorate this situation. The Blacks that Whites encounter in photo-arrays will look more like rappers Public Enemy than Bill Cosby.

The Photo-Array Moment

One recurrent forensic event, a witness's exposure to a photo-array composed of mug shots of young Black men, stands at the confluence of a startling number of psychological forces, all tending to depress a White European's assessment of the likelihood of an error in making an identification or of that error's cost, or both.

A White eyewitness presented with a photographic array of mug shots of young Black men confronts a uniform, semantic category. This category evokes

all of the in-group/out-group dynamics fully discussed in a number of contributions to this issue. (Sporer, 2001a, 2001b). It is also aggravated by the fact that the particular out-group under scrutiny here has chosen for reasons of its own to adopt a "hard" pose by dressing, talking, and looking as much alike as they can (Majors & Billson, 1992; Doyle, 1996). There is no sinister design behind this: Many of these features of the photo-array process are natural consequences of the approved match-to-description procedure for lineup construction (Technical Working Group for Eyewitness Evidence, 1999).

All of this brings to the stage the traditional elements of Europe's encounter with the Other. At least to Whites, "they" do sometimes all look the same. George Orwell (1968) commented on his own reactions with characteristic honesty in describing a visit to Marrakesh:

> When you walk through a town like this—two hundred thousand inhabitants, of whom at least twenty thousand own literally nothing except the rags they stand up in—when you see how the people live, and still more, how easily they die, it is always difficult to believe that you are walking among human beings. . . . The people have brown faces—besides, there are so many of them! Are they really the same flesh as yourself? Do they even have names? Or are they merely a kind of undifferentiated brown stuff, about as individual as bees or coral insects? (p. 387)

This attitude—irresistibly evoking Nickerson and Adams's (1979) demonstration that although most people distinguish pennies from other coins, few can recognize the correct features of the penny in a penny lineup—is only the beginning of the problem.

The fact that the suspect's criminal history is revealed by the mug-shot format shows the witness not only that the suspect is a member of an out-group but also that he is a member of a smaller out-group within that out-group—one more likely to be guilty of a crime (Sporer, 2001a). Moreover, the format also indicates that because each of the men displayed in the mug-shot array was guilty in the past, each is also likely to be guilty of something else in the future; the format therefore intensifies the spectre of a new error cost associated with a false negative or a miss. On the one hand, it seems to the witness that an error in making an identification choice is less likely; on the other hand, it seems that making no identification at all dictates an error cost in the form of a future crime that might have been prevented. At the same time, the witness is encouraged by the format to feel that a mistake in identifying a veteran criminal for the wrong crime is simply the price society has to pay for security: Maybe he didn't do this one; probably he did another.

The cultural forces driving down White estimates of the error cost in an identification case will not be captured by self-reports of contact or generalized assessments of attitude. Research into attitudes and contacts must begin to take account of what John Thompson (1990) has called "the mediazation of modern culture" (p. 16). People rely less and less on face-to-face contact, more and more on the products of complex commercial mass media. As Thompson put it,

> The deployment of technical media should not be seen as a mere supplement to pre-existing social relations: rather, we should see this deployment as serving to

> create new social relations, new ways of acting and interacting, new ways of representing oneself and responding to the self-presentation of others. (pp. 16)

I have argued elsewhere (Doyle, 1992, 1996) that contemporary media products in both their news and entertainment incarnations have a specific, dehumanizing impact on contemporary Whites' views of who lives in the inner city and how they live. I do not say that these arguments are in any sense proven by the data collected in this issue, but it does seem to me that the false alarm virus that this issue has begun to isolate would find a particularly receptive host in a cultural context weighted in that direction.

The Problems With Solutions

What does it mean for the relationship between psychological science and the legal system if the heart of the ORB problem does lie in the criterion shifts experienced by White eyewitnesses? Can the legal system adjust to the findings collected here? It is very difficult to be optimistic: Problems of principle and problems of practice stand in the way.

Opponents of expert psychological testimony in eyewitness cases have always worried that the effect of expert testimony is to lower the number of mistaken convictions at the cost of raising the number of mistaken acquittals. In this view, expert testimony, by artificially enhancing the jury's general skepticism, may do more harm than good and certainly involves scientific psychology in a utilitarian calculation that, to put it bluntly, is none of science's business (Egeth & McCloskey, 1984). As the body of eyewitness research has developed and deepened, proponents of expert testimony have been able to respond that, on the contrary, expert testimony effectively draws the jurors' attention to factors—for example, decision time—that are diagnostic of accuracy and away from others—for example, trial testimony confidence—that can mislead (Leippe, 1995). Where thoughtful courts have accepted expert psychological testimony on eyewitness identification, it has been because the courts have been persuaded that there is a fit (*United States v. Downing,* 1985) between particular, diagnostically significant elements of the research and the case on trial.

If, however, ORB turns out to be rooted in discounted error costs and relaxed criteria, the question arises whether general skepticism is not what ORB is all about. There is, after all, no preexisting set of ideal criteria. Does one actually know that Whites' criteria in regard to choosing Blacks are too low and their criteria for choosing Whites are just right? Or does one know only that they are different? Perhaps both sets of criteria are wrong. Although there is widespread agreement that ORB exists, can anyone actually say that the racial difference between a White witness and a Black suspect is in any meaningful sense diagnostic of reliability? That racial difference "postdicts" error? If not, isn't error cost being discussed as a judicial question—even a profound political question—rather than a scientific one? Everyone can agree that in trials of Black and White defendants, jurors should see the error costs as equivalent and apply the same criteria, but how to bring that about is a complex question.

Contributors to this issue, for example, find cause for guarded optimism in *State v. Cromedy* (1999), a recent decision of the New Jersey Supreme Court that

mandated a jury instruction in cross-racial identification cases (Meissner & Brigham, 2001). However, the *Cromedy* decision merely illustrates the danger of using the existing legal toolbox to catch eyewitness errors after they have occurred (Bartolomey, 2001).

The *Cromedy* opinion requires New Jersey judges to draw the jury's attention to the fact that the identification is cross-racial; it requires nothing more. The *Cromedy* court apparently believed that this highlighting will activate the jury's preexisting understanding that cross-racial identifications are difficult to accomplish. (Because the *Cromedy* court believed that this difficulty was common knowledge, it also forbade expert testimony on the point.)

It is difficult to say whether the *Cromedy* instruction amounts to nothing or whether it amounts to worse than nothing. Even assuming that the laconic instruction that the court approved does remind the jurors that cross-racial identifications are more difficult, has anything been achieved? After all, if the dangers in ORB lie in relaxed criteria, then cross-racial identifications are not more difficult for White witnesses; cross-racial identifications are more facile. Correct cross-racial identifications are more difficult, but the act of choosing is apparently experienced as less demanding.

Besides, there is the pronounced danger that once a jury is persuaded that cross-racial identifications are more difficult in the sense that they impose heightened challenges for eyewitnesses, the jury will rely more heavily on misleading factors (such as witness confidence) in determining whether the witness has surmounted those challenges or not. At best, using jury instructions to manipulate the jurors' criteria in an effort to compensate for artificially relaxed criteria used by White eyewitnesses will be an unwieldy business. Conceivably, a cross-racial identification instruction given at the beginning of a trial and framing the jurors' response to what is to come may prove more effective, but it is hard to be optimistic that the effect will be dramatic. Expert testimony will be preferable in that it can attempt to steer jurors away from resort to false criteria of their own (e.g., testimonial confidence) to buttress the relaxed criteria employed by the witness, but it will be no panacea. The current state of the science precludes psychology achieving any immediate transforming impact on the legal system's post hoc evaluation of cross-racial identifications.

The Best and the Good

Still, there is the possibility of preventing errors before they occur, and the research into cross-racial identification error does suggest some avenues of exploration that tend in that direction. Although the best approach to the problem might be to secure a comprehensive understanding of the fundamental nature of the ORB, there are systems-variable inquiries that may shed light on the fundamentals while at the same time ameliorating the symptoms.

To begin with, if the ORB false alarm increase is the result of lowered barriers to identification choices (and "lowered" may not be the word; "more permeable" might better capture the situation), then a set of marginal identifications that call for particularly rigorous application of the system-variable procedures already uncovered may be at issue. The National Institute of Justice's (NIJ's) *Eyewitness*

Evidence: A Guide For Law Enforcement (Technical Working Group for Eyewitness Evidence, 1999) takes on a heightened importance in cross-racial situations. Indeed, the elements of psychologically endorsed identification practice that were eliminated in the NIJ guide—most notably, sequential lineup administration and blind lineup procedures—seem particularly important once one has reviewed the cross-racial studies (Wells et al., 2000). For the marginal White eyewitness tempted to make an identification by a lower perceived error cost, eliminating the relative judgment aspects of a simultaneous lineup may make the difference. For a marginal White witness who has allowed an identification to slip past more liberal criteria, inadvertent confidence boosting by a lineup administrator who is aware of the suspect's identify may prove particularly seductive.

Research may also uncover new system-variable methods of tweaking existing procedures. Wells and Olson's (2001) ingenious suggestion that cross-racial cases use blank screening lineups or arrays is a promising example. Research may also continue to investigate the causes while attacking the symptoms. Perhaps, for example, the error cost calculations could be tested in staged crime events by presenting target/present and target/absent cross-racial arrays in differing formats: say, mug-shot, yearbook, and family picture settings. Perhaps computer technology could be used to "de-mug" the mug shot. Perhaps a foil could ask the witness, "Can you believe all these guys are from the medical school?" Perhaps instructions to lineup witnesses that include a cross-racial warning and make race salient would lessen the social pressure on witnesses by making them feel free to say no. Perhaps, as Sommers and Ellsworth's (2001) jury study indicated, White decision makers will do better when race is made more salient.

It may be that continued conscientious research into the psychological roots of ORB will result in a diagnosis to which experimental psychology cannot be applied as a cure. It maybe that the source of ORB is impacted in our culture. That is no reason to fail to pursue research along the lines collected in this issue. One may as well know, even if there is little one can do about what one learns.

References

Bartolomey, D. (2001). Cross-racial identification testimony and what not to do about it. A comment on the cross-racial jury charge and cross-racial expert identification testimony. *Psychology, Public Policy, and Law, 7,* 247–252.

Bridges, G., & Steen, S. (1998). Racial disparities in official assessments of juvenile offenders. *American Sociological Review, 63,* 554–571.

Chance, J. E., & Goldstein, A. G. (1996). The other-race effect and eyewitness identification. In S. L. Sporer, R. S. Malpass, & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 153–176). Mahwah, NJ: Erlbaum.

Doyle, J. M. (1992). "It's the third world down there!": The colonialist vocation and American criminal justice. *Harvard Civil Rights–Civil Liberties Law Review, 27,* 71–112.

Doyle, J. M. (1996). The lawyers' art: "Representation" in capital cases. *Yale Journal of Law and Humanities, 8,* 417–450.

Dwyer, J., Neufeld, P., & Scheck, B. (2000). *Actual innocence: Five days to execution and other dispatches from the wrongly convicted.* New York: Doubleday.

Egeth, H. E., & McCloskey, M. (1984). Expert testimony about eyewitness behavior: Is it safe and effective? In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony:*

Psychological perspectives (pp. 127–136). Cambridge, England: Cambridge University Press.

Gates, H. L. (1994). Preface. In T. Golden (Ed.), *Black male: Representations of masculinity in contemporary American art* (pp. 11–14). New York: Abrams.

Leippe, M. (1995). The case for expert testimony about eyewitness memory. *Psychology, Public Policy, and Law, 1,* 909–959.

Li, J. C., Dunning, D., & Malpass, R. S. (1988, March). *Basketball fandom and cross-racial identification among European-Americans: Another look at the contact hypothesis.* Paper presented at the biennial meeting of the American Psychology–Law Society, Redondo Beach, CA.

MacLin, O. H., & Malpass, R. S. (2001). Racial categorization of faces: The ambiguous race face effect. *Psychology, Public Policy, and Law, 7,* 98–118.

MacLin, O. H., MacLin, M. K., & Malpass, R. S. (2001). Race, arousal, attention, exposure, and delay: An examination of factors moderating face recognition. *Psychology, Public Policy, and Law, 7,* 134–152.

Majors, R., & Billson, J. M. (1992). *Cool pose: The dilemmas of Black manhood in America.* New York: Lexington Books.

Mauer, M., & Huling, T. (1995). *Young Black Americans and the criminal justice system: Five years later.* Washington, DC: The Sentencing Project.

McCleskey v. Kemp, 481 U.S. 297 (1987).

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, and Law, 7,* 3–35.

Morris, N. (1988). Race and crime: What evidence is there that race influences result in the criminal justice system? *Judicature, 72,* 111–134.

Nickerson, R. S., & Adams, M. J. (1979). Long-term memory for a common object. *Cognitive Psychology, 11,* 287–307.

Orwell, G. (1968). Marrakesh. In S. Orwell & I. Angus (Eds.), *The collected essays, journalism, and letters of George Orwell* (vol. 1, pp. 287–313). New York: Harcourt, Brace & World.

Posner, R. (1973). An economic approach to legal procedure and judicial administration. *Journal of Legal Studies, 2,* 399–426.

Rose, T. (1994). Rap music and the demonization of young Black males. In T. Golden (Ed.), *Black male: Representations of masculinity in contemporary American art* (pp. 149–158). New York: Abrams.

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, and Law, 7,* 153–169.

Sommers, S. R., & Ellsworth, P. E. (2001). White juror bias: An investigation of prejudice against Black defendants in the American courtroom. *Psychology, Public Policy, and Law, 7,* 201–229.

Sporer, S. L. (2001a). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law, 7,* 36–97.

Sporer, S. L. (2001b). The cross-race effect: Beyond recognition of faces in the laboratory. *Psychology, Public Policy, and Law, 7,* 170–200.

State v. Cromedy, 158 N.J. 112, 727 A.2d 457 (1999).

Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement.* Washington, DC: United States Department of Justice, Office of Justice Programs.

Thompson, J. B. (1990). *Ideology and modern culture.* Stanford, CA: Stanford University Press.

United States v. Downing, 609 F. Supp. 784 (E.D. Pa. 1985) (opinion after remand).
Wells, G. L., Malpass, R. S., Lindsay, R. C. L., Fisher, R. P., Turtle, J. W., & Fulero, S. (2000). From the lab to the police station: A successful application of eyewitness research. *American Psychologist, 55,* 581–598.
Wells, G. L., & Olson, E. A. (2001). The other-race effect in eyewitness identification: What do we do about it? *Psychology, Public Policy, and Law, 7,* 230–246.

Received October 3, 2000
Revision received December 20, 2000
Accepted December 21, 2000 ■

APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **20**: 1009–1023 (2006)
Published online 21 June 2006 in Wiley InterScience
(www.interscience.wiley.com). DOI: 10.1002/acp.1241

Show-ups: The Critical Issue of Clothing Bias

JENNIFER E. DYSART[1]*, R. C. L. LINDSAY[2] and PAUL R. DUPUIS[3]

[1]*John Jay College of Criminal Justice, City University of New York, New York, NY, USA*
[2]*Queen's University, Kingston, Ontario, Canada*
[3]*Red Deer College, Red Deer Alberta, Canada*

SUMMARY

A field study (N = 379) investigated the effects of clothing bias on show-up identifications using variations in type of clothing (distinct and common), the similarity of clothing between the event and the identification procedure, target-present and two target-absent show-ups (high similarity and low similarity innocent suspects), and time delay. Results showed a significant clothing bias by clothing type interaction on identification accuracy; however, no overall effects of delay or common clothing on identification accuracy were found. With distinct clothing, significant effects of clothing bias and suspect similarity emerged. Implications for police use of show-ups are discussed. Copyright © 2006 John Wiley & Sons, Ltd.

In 1990, witnesses to a gang shooting in New Jersey described one of the gang members as wearing a Day-Glo reversible jacket (orange and burgundy) with the orange side turned out. Shortly after the shooting, the police apprehended Luis Kevin Rojas wearing an orange and burgundy reversible jacket, with the burgundy side out. After being asked to turn the jacket inside out to expose the orange side, Mr. Rojas was identified by seven eyewitnesses as being a gang member who had been involved in the shooting (Flynn, 1998). Was Mr. Rojas identified because he was presented to the witness in a show-up rather than a lineup? Would he have been identified from the show-up even if he had not been forced to wear his coat orange side out? In other words, was this a mistaken identification resulting from accumulated biases? A brief review of the show-up literature will demonstrate that show-ups are a risky method of eyewitness identification. Further, a review of the (very limited) literature suggests that clothing bias is a serious concern. We will argue that the show-up is an identification procedure prone to clothing bias and then explore the extent and some limiting conditions, of the clothing bias effect on show-ups in a field study.

The show-up involves exposing a witness to one person, the suspect, either live or in a photograph. A potential problem with this procedure is that with only one person shown to the eyewitness, the identity of the police suspect is obvious. In fact, when the identity of a suspect is obvious in a lineup, the lineup is considered unfair (e.g. Lindsay & Wells, 1980; Phillips, McAuliff, Kovera, & Cutler, 1999). Thus, the inherent suggestiveness of the show-up may lead a witness to identify an innocent suspect as the perpetrator. Despite the

*Correspondence to: J. E. Dysart, Department of Psychology, John Jay College of Criminal Justice, The City University of New York, 445 West 59th Street, New York, NY, 10019-1199, USA. E-mail: jdysart@jjay.cuny.edu

Contract/grant sponsor: Social Sciences and Humanities Research Council of Canada.

Copyright © 2006 John Wiley & Sons, Ltd.

suggestiveness of the show-up, past research on show-up identifications has reported a reduced tendency for witnesses to choose anyone from a show-up, including the perpetrator (Gonzalez, Ellsworth, & Pembroke, 1993; Wagenaar & Veefkind, 1992). In addition, a recent meta-analysis of show-up studies found that show-ups and simultaneous lineups result in virtually identical rates of correct identification (Steblay, Dysart, Fulero, & Lindsay, 2003). The fact that guilty suspects are no more likely to be selected from show-ups than lineups is puzzling. The complete absence of foils (or distractors) in the show-up would seem to make the task easier for the witness and thus, one would expect the show-up to result in higher correct identification rates of the perpetrator. One possible explanation for this finding is that witnesses respond to the perceived pressure to choose with psychological reactance (Brehm & Brehm, 1981) and employ a more stringent decision criterion. Regardless of the reason, show-ups appear to provide neither an advantage nor disadvantage over lineups in terms of correct identifications.

Historically, the concerns expressed about show-ups have centred on the potential for mistaken identification. Some studies support biasing effects of show-ups (e.g. Dysart & Lindsay, in press; Yarmey, Yarmey, & Yarmey, 1996) but others report false identification rates from show-ups that are comparable to those from simultaneous lineups (Lindsay, Pozzulo, Craig, Lee, & Corber, 1997). Finally, some studies have found that show-ups yield more accurate identification decisions overall than lineups (Dekle, Beal, Elliott, & Huneycutt, 1996; Yarmey, Yarmey, & Yarmey, 1994). This last finding, however, can be misleading. Show-ups generate higher rates of correct decisions because they produce higher rates of correct rejection in target-absent conditions as compared to simultaneous lineups. The concern with show-ups, however, is that they do not afford the same protection to innocent suspects, that is provided by using foils (distracters) in lineups. In lineups, incorrect foil selections are known errors and do not result in any legal action against the foil that is selected. With show-ups, all false positive choices are false identifications of innocent suspects. In support of this concern, the Steblay et al. (2003) meta-analysis indicated that false identifications of the suspect were substantially more likely from show-ups than lineups.

Eyewitness research may also be underestimating the rate of false identifications with show-ups in the real world. The reason for this is because lineups usually require that police have a specific suspect in custody and an arrest usually requires some source of evidence leading police to have probable cause to believe the suspect is guilty prior to the identification procedure. Show-ups, however, frequently are based on minimal information, usually that the suspect resembles a description provided by a witness and that the suspect was apprehended near the scene of the crime within a generally brief period following the crime. Given that eyewitness descriptions are frequently vague (Lindsay, Nosworthy, Martin, & Martynuck, 1994), show-ups may commonly contain innocent suspects. Furthermore, police and the courts justify the use of show-ups, at least in part, by arguing that they provide a means of quickly and efficiently exonerating the innocent. If this logic is used, police would be expected to have a more lenient criterion for presenting suspects in show-ups. As a result, the base rate of suspect innocence may be substantially higher for show-ups than lineups. Combined with an equal or higher propensity for false identification, a higher base rate of innocence could lead to a much higher rate of false identification from show-ups than lineups.

Despite the intuitive and empirically demonstrated problems with the show-up procedure described thus far, it is a common method of identification used by police in the United States. Flowe, Ebbesen, Burke, and Chivabunditt (2001) reported that 55% of

Copyright © 2006 John Wiley & Sons, Ltd.

identifications conducted in 488 cases between 1991 and 1995 in a large metropolitan area in the western US used show-ups. McQuiston and Malpass (2001) reported that show-ups were used by police in El Paso County, Texas, for 30% of identification attempts. Gonzalez et al. (1993) enlisted the help of a detective in Illinois to record all identifications (line-ups and show-ups) he was involved in for a period of time. Results from this field study indicated that 77% of identifications conducted were show-ups. Because of these sizeable percentages, and the limited number of studies that have investigated factors that may affect the show-up procedure, more studies are needed which demonstrate the potential limitations or advantages of this identification technique.

Some researchers have attributed lineup identification errors to the use of relative judgement strategies (i.e. the witness treating the lineup like a multiple choice test and choosing the lineup member most similar to the memory trace of the criminal) (Lindsay & Bellinger, 1999; Lindsay & Wells, 1985; Wells, 1984; for an alternative view see Clark & Davey, 2005). Lineup biases may function by making the suspect easy to select relative to other lineup members (Lindsay, Wallbridge, & Drennan, 1987; Lindsay & Wells, 1980) or by encouraging witnesses to employ relative judgements (Malpass & Devine, 1981). Sequential lineups have been shown to reduce reliance on relative judgements (Lindsay & Bellinger, 1999) and consistently reduce false positive choices of innocent suspects (Steblay, Dysart, Fulero, & Lindsay, 2001). Lineup members in sequential lineups are presented individually to prevent witnesses from using a comparative strategy. Since, by definition, show-ups involve exposing the witness to a single person, relative judgements are not possible with show-ups. As a result, the show-up procedure may produce fewer false identifications than what would be predicted based on the suggestive nature of the procedure.

Lineup biases have been shown to have less impact with sequential lineups (Lindsay, Lea, & Fulford, 1991; Lindsay, Lea, Nosworthy, Fulford, Hector, LeVan, & Seabrook, 1991). Clothing bias, where only the suspect appears in the identification procedure wearing clothing similar to that described by the eyewitness, has been shown to increase false identifications from lineups (Lindsay & Wells, 1980; Lindsay et al., 1987; Malpass & Devine, 1981), mug shots (Lindsay et al., 1994), and show-ups (Yarmey et al., 1996). However, the impact of clothing bias has not sufficiently been tested in association with show-ups.

Clothing bias is of particular concern with the show-up procedure because suspects presented in show-ups often are apprehended because they were near the scene of the crime and matched the overall description given by the eyewitness. This concern is supported by the fact that approximately 50% of all information reported in eyewitness descriptions is clothing information (Lindsay et al., 1987, 1994). Furthermore, clothing descriptions may be more distinctive than person descriptions. If a witness describes a robber as a clean-shaven white male in his early 20s, the police are unlikely to stop everyone they see on the street who fits that description. However, if the description also includes information that the robber was wearing blue jeans, a white T-shirt, and a baseball cap, police may rely heavily on these clothing cues to obtain a suspect near the scene of the crime who fits the description of the clothing and is not inconsistent with the vague person description. This scenario, which is likely to be common with show-up identifications, leads to potential clothing bias because the suspect is likely to be wearing clothing similar to that remembered by the eyewitness.

To date, one study (Yarmey et al., 1996) has investigated the effects of clothing bias on identification accuracy from show-ups. Yarmey et al. found that when a similar looking

Copyright © 2006 John Wiley & Sons, Ltd.

(innocent) suspect was presented in a show-up wearing the same clothing that was worn by the target, false identifications were higher than when the innocent suspects wore different clothing than that worn by the target during the event. Yarmey et al. also presented participants with implausible innocent suspects who did not match the description of the targets in any manner. In this condition, clothing bias had no significant effect on the false identification rate. In addition, Yarmey et al. found that correct identifications of the target were not affected by the clothing manipulation, supporting the notion that clothing bias only serves to decrease identification accuracy (Lindsay et al., 1987).

There are many unanswered questions with respect to clothing bias that need to be addressed, particularly when combined with the show-up procedure. One question relates to the probability that a similar looking innocent person, wearing exactly the same clothing as the perpetrator, will be in the vicinity of the crime immediately after the event. Although this scenario is possible, it seems rather unlikely and, thus, the manipulation used by Yarmey et al. (1996), where the innocent person both resembled the target (her sister) and was presented in the exact clothing worn by the perpetrator, may not accurately reflect most potentially dangerous clothing bias situations in the real-world. Addressing this issue with lineups, Lindsay et al. (1987) investigated the effects of identical, similar, and different clothing on identification accuracy. There was no difference in the impact of identical and similar clothing, both produced the clothing bias effect. Clothing similarity was enough to make the innocent target stand out in the lineup and result in selections of him. The current study investigated the effects of dissimilar, similar, and identical clothing on identification accuracy from show-ups.

A second question that has yet to be addressed in clothing bias investigations is whether the type of clothing worn by the perpetrator has an effect on the magnitude of the biasing effect. For example, if a perpetrator is described as wearing a white T-shirt, is a witness more or less likely to be affected by clothing bias than if the perpetrator was described as wearing a distinctive shirt (e.g. a shirt with a distinctive logo, photograph, or words)? The former 'common' clothing could have less of an impact on identification accuracy due to the increased likelihood that others could be wearing this type of clothing (and thus that the police could easily find an innocent person matching this description). The latter 'distinct' clothing is unlikely to be as typical and thus, its biasing impact may be greater because of the low probability of finding an innocent person who matches this distinct description. If a witness was to engage in the reasoning described above (e.g. 'What are the chances that the police could find an *innocent* person matching my entire description?'), it is likely that he or she would use the distinct clothing as evidence of suspect guilt. Alternatively, common clothing may be less likely to draw the attention of witnesses, less likely to be encoded, and thus less likely to produce any biasing effect. The clothing used in previous clothing bias investigations has clearly been distinctive. Yarmey et al. (1996) used a university T-shirt or sweater (depending on the weather) with a 'Guelph XXL Gryphons' logo on the front (A. D. Yarmey, personal communication, April 30, 2001); Lindsay et al. (1987) used a university sweatshirt with a distinctive logo; Lindsay, et al. (1991) used a T-shirt with stick figures and writing; and Lindsay et al. (1994) used a colourful Hawaiian patterned shirt to demonstrate clothing bias in a mug shot search. Thus, one issue that has not been addressed in the previous literature is the effect of less distinctive clothing on identification bias.

Just as the similarity of the clothing is an issue of some importance, so is the match between the true perpetrator and the innocent suspect. The majority of studies conducted to date have employed as innocent suspects people selected because they resembled the confederates who staged the events. In real cases, the descriptions provided by witnesses

Copyright © 2006 John Wiley & Sons, Ltd.

are often vague and, indeed, the description of the perpetrator's clothing may be more detailed and distinctive than the description of the person. As a result, when police apprehend an innocent suspect based on these descriptions, the clothing may match the description of the clothes worn by the perpetrator more closely than the suspect resembles the perpetrator. An important question not addressed in previous studies is the impact of clothing bias on identification of innocent people who are reasonable suspects (i.e. match the general description of the perpetrator) but not particularly similar to the true culprit. If an innocent suspect must closely resemble the perpetrator as well as being dressed similarly for clothing bias to present a problem, the laws of probability suggest that clothing bias may be a real but relatively rare problem for the criminal justice system. On the other hand, if clothing bias produces mistaken identification of people who would not be likely to be identified in the absence of the clothing cues, clothing biased show-ups are a serious threat to innocent suspects and false identifications are likely to occur frequently. The fact that show-ups are recommended in official, government sponsored guidelines for police practice (Technical Working Group for Eyewitness Evidence, 1999) and frequently used by police (Flowe, Ebbesen, Burke, & Chivabunditt, 2001; Gonzalez et al., 1993; McQuiston & Malpass, 2001) make it imperative that the effects of clothing bias on show-ups be investigated further.

Finally, there is no theory concerning the impact of clothing similarity and suspect similarity on the accuracy of show-up identification. One explanation is that clothing provides a memory cue to witnesses assisting them in the search for the face of the criminal and misleading them when the suspect is innocent (literally, a form of guilt by association). This seems unlikely given the consistent reports from lineup studies that correct identification rates are not increased by clothing cues. A second possibility was illustrated above. Clothing information may be used by witnesses in an inferential process where they estimate the likelihood that the person presented is the criminal. A third possibility is that clothing acts as a piece of information in an inferential process but only if the memory trace is weak. Although the current research was not designed to test theoretical explanations for the clothing bias, it is worthwhile to begin discussing the issue.

To explore the issues raised thus far, participants in the current study were exposed to a target person in a field setting. Then, following a delay period, participants were shown either a photograph of the target, a high similarity innocent suspect, or a low similarity (but still plausible) innocent suspect.

METHOD

Participants

A male confederate (hereafter target) approached store and business employees in various shopping malls in southern Ontario, Canada. Only employees who were not occupied (e.g. serving customers or stocking shelves) were approached and asked if they would volunteer to participate in a study for the Department of Psychology at Queen's University. Of the 430 people who agreed to participate (approximately 90% of all persons approached), there were 99 men and 331 women. Some participants ($N = 51$) were dropped from the analysis because they were no longer available (e.g. gone home for the day or in a meeting) when the experimenter returned to conduct the identification phase of the study.

Copyright © 2006 John Wiley & Sons, Ltd.

The final sample (N = 379) consisted of 88 men (*M* age = 28.1 years) and 291 women (*M* age = 27.9 years).

Design

The study employed a 3 (show-up individual: target, high similarity innocent suspect, low similarity innocent suspect) × 2 (type of clothing worn at original event: distinct, common) × 2 (clothing bias: same clothing worn at event and show-up, different clothing at original event and show-up) between-subjects design with time delay as a continuous independent variable. Two additional conditions were used to test a specific issue. The two innocent suspects also were shown in clothing similar, but not identical, to the distinct clothing to investigate the effects of showing innocent people in similar, yet incorrect, clothing. These two conditions will be referred to as 'similar-distinct' conditions and these data were only included in the 'Similar-distinct clothing' section in the Results. The primary dependent measure was identification accuracy. Post-decision confidence also was measured and some description data were obtained.

Materials

Materials were 7.6 × 10.2 cm frontal, head and chest, colour photographs of the target and the high and the low similarity innocent suspects. The target was white, in his early 30s, 1.83 m tall, weighed 113 kg, was clean-shaven, had short brown hair, and no distinguishing or outstanding features. The two innocent suspects were selected based on pilot work where 60 participants interacted with the target for approximately 45 seconds, gave a description of the target, and then ranked from most to least similar the similarity to the target of five foils used in previous lineup research (Pryke, 2001). The foils that were ranked first (i.e. most similar) and fourth overall were selected as the high similarity and low similarity innocent suspects, respectively (Dupuis, 2001). The pilot study results also showed that a few participants ranked the low similarity innocent suspect as most similar to the target, indicating that, although he was dissimilar to the target, he was not an implausible replacement. On the other hand, in lineup research, the low similarity suspect had never been selected from either the target-present or target-absent lineup.

Two shirts were used for the primary clothing bias manipulations. One shirt was a man's blue-gray, short sleeved, button-up plaid shirt (hereafter common clothing). This style of shirt was typical of men's fashion at the time the data were collected and was considered to be common men's apparel. The second shirt was a black Harley-Davidson T-shirt with a motorcycle, blue eagle wings and the Harley-Davidson logo on the front (hereafter distinct clothing). While recruiting participants for the study, the target either wore the common clothing or the distinct clothing. A third shirt was used to specifically test the possibility that similar but not identical clothing may be sufficient to produce a clothing bias effect. The third shirt was a black Harley-Davidson T-shirt with a motorcycle, howling wolf, and the Harley-Davidson logo on the front (hereafter similar-distinct clothing). The size of the design on both Harley-Davidson T-shirts was approximately equal. Photographs of the innocent suspects were taken in each of the three shirts to be used in the identification phase of the study. Photographs of the target in the common and the distinct shirts worn at the event also were used. The two innocent suspects appeared in one of the three shirts at test.

Copyright © 2006 John Wiley & Sons, Ltd.

Procedure

Employees (i.e. prospective participants) of stores in shopping malls in southern Ontario, Canada were approached by the target, who introduced himself as a graduate student in psychology. He then asked if they would participate in an experiment for the Department of Psychology at Queen's University. Employees (in groups of one to three people) were told that, if they agreed to participate, an (female) experimenter would return later that day to ask them a few innocuous questions. Because we wanted to avoid telling participants that the study was investigating their ability to identify the target (in case this type of knowledge affected the participants' accuracy) and we needed to give participants enough information to decide if they wanted to participate in the study, the target explained that he could not tell participants what the questions were because his job was only to recruit participants. However, the target assured employees that no personal or embarrassing questions would be asked of them and that no questions would be asked regarding their place of employment. At this point, if store employees agreed to take part in the study, the target thanked them for their time and reminded them that a second person would return later that day. If employees did not want to participate or felt that they needed permission from management prior to participating in the study, they were told of the purpose of the study and were not included as participants. Immediately upon leaving each store where employees had agreed to participate, the target recorded the time, the name of the store, and the sex of the participant, approximate age, and a brief description of the participant to facilitate the identification of participants by the returning experimenter. This was an important step because not all store employees who were present when the experimenter returned were present when the target had earlier recruited participants.

The experimenter entered the stores and recorded the time which, when compared with the confederate's recorded time, indicated delays varying from 10 to 272 minutes (to the nearest minute). The experimenter assessed whether the participant(s) were unoccupied and, therefore, available to complete the study. At this time, if participants were occupied, the experimenter returned later or waited until the participant(s) were available (recording that time). If the participant(s) were unoccupied, they were told that they were going to be shown a photograph and asked to make a decision as to whether or not the photograph was of the person who asked them to participate in the study. Participants were told that the photograph they were going to be shown may or may not be of the person who recruited them to be in the study (standard lineup warning). A photograph of the target, the high similarity innocent suspect, or the low similarity innocent suspect was then shown to the participant. The person in the photograph was either wearing the common, distinct, or similar-distinct (innocent suspects only) clothing. Participants then made an identification decision and rated their confidence in that decision on a 10-point scale (1 being 'Not at all confident', 10 being 'Extremely confident'). If customers/clients had entered the store during the identification and participants needed to be dismissed, the study ended immediately after obtaining their confidence rating. However, if no customers awaited service, participants were asked, while viewing the photograph, if they could remember and describe what the target was wearing when he came into the store.

Participants were randomly assigned to conditions where possible. It was logistically impossible for the confederate to change shirts from one store to the next. As a result, the confederate would recruit participants wearing either the common or the distinct shirt for a

Copyright © 2006 John Wiley & Sons, Ltd.

period of time and then change to the other shirt. Within groups of participants recruited during each block of time, show-up conditions were assigned randomly. Differences in delay were haphazard rather than random but mean delay was approximately equal (not significantly different) across clothing and show-up conditions (all F values <1).

RESULTS

The results presented below tested a 3 (show-up individual) $\times$ 2 (clothing bias) $\times$ 2 (clothing worn at event) between-subjects design with identification accuracy from show-ups as the primary dependent variable. The data from two additional conditions, where the innocent suspects were presented in similar looking Harley-Davidson (distinct) attire were not included in these analyses and are presented separately in the section titled 'Similar-distinct clothing'.

Identification accuracy

Delay

Logistic regression analyses revealed no effects of delay on identification accuracy from show-ups or any significant interactions with delay in any condition. Participants were equally likely to make a correct identification decision from 10 minutes to 4 hours after the event. Therefore, the remainder of the analyses were conducted by collapsing over this variable.

Overall

Using logistic regression analysis with accuracy as the dependent variable and all of the variables in the equation, a three-way interaction was tested between clothing bias, type of clothing at event and photograph shown. Results showed no significant interaction ($Wald = 1.68$, $p = 0.195$) of these variables. Overall, there was a significant main effect of bias on identification accuracy, with the biased conditions yielding a lower correct decision rate (74%) than the unbiased conditions (87%; $Z = 3.22$, $p = 0.001$). However, logistic regression analysis, with clothing type and clothing bias in the equation, indicated a significant clothing type by clothing bias interaction ($Wald = 10.29$, $p = 0.001$). Participants in the distinct clothing condition were significantly less likely to make a correct identification decision from the biased (64%) than from the unbiased (90%) condition ($Z = 4.03$, $p < 0.001$), whereas participants in the common clothing condition were as likely to make a correct identification decision from the biased (89%) and unbiased conditions (84%; $Z = 0.96$, $p = 0.34$).

Additional interactions were tested by separating the target-present and two target-absent conditions. The clothing bias by type of clothing interaction (with both variables in the equation), for the target-present and low similarity innocent suspect were non-significant ($Wald = 2.72$, $p = 0.10$, $Wald = 0.01$, $p = 0.99$, respectively). The result of the same analysis with the high similarity innocent suspect, however, was significant ($Z = 3.71$, $p = 0.05$). Participants in the unbiased conditions were more accurate in the distinct clothing condition (92.6%) over the common clothing condition (84.0%), whereas participants in the biased conditions were more accurate in the common clothing condition (96.0%) than in the distinct clothing condition (83.3%).

Copyright © 2006 John Wiley & Sons, Ltd.

Table 1. Proportion of 'Yes' identification decisions for common and distinct clothing as a function of clothing bias conditions and show-up individual

Clothing at event	Clothing bias condition		
	Unbiased	Biased	Biased similar-distinct
Common clothing			
Show-up individual			
Target	0.84 (21/25)	0.96 (24/25)	N/A
High similarity innocent suspect	0.25 (6/24)	0.16 (4/25)	N/A
Low similarity innocent suspect	0.08 (2/25)	0.13 (3/24)	N/A
Distinct clothing			
Show-up individual			
Target	0.93 (25/27)	0.83 (20/24)	N/A
High similarity innocent suspect	0.23 (7/30)	0.50 (17/34)	0.50 (14/28)
Low similarity innocent suspect	0.00 (0/29)	0.37 (11/30)	0.14 (4/29)

Note: Frequencies shown in parentheses.

Distinct clothing

A significant bias by photograph interaction was found within the distinct clothing conditions ($Wald = 16.64$, $p < 0.001$). Correct identifications of the target were unaffected by the clothing bias manipulation in the distinct clothing conditions, with 83% of participants identifying him in the biased condition and 93% identifying him from the unbiased condition (see Table 1; $Z = 1.01$, $p = 0.31$). On the other hand, identification accuracy for the high similarity and low similarity innocent suspects was significantly affected by the clothing bias manipulations. When the high similarity innocent suspect was presented in the unbiased condition, 23% of participants incorrectly identified him, whereas 50% of participants identified him from the biased condition ($Z = 2.21$, $p = 0.01$). Similar results were obtained for the low similarity innocent suspect, where 0% and 37% of participants identified him from the unbiased and biased conditions, respectively ($Z = 3.61$, $p = 0.001$).

Common clothing

No interaction was found between the clothing bias conditions and photograph shown ($Wald = 0.15$, $p = 0.70$) when the target appeared in the common clothing. Participants who viewed the biased target-present show-up were no more accurate (96%) than those who viewed the unbiased target-present show-up (84%; $Z = 1.41$, $p = 0.16$; see Table 1). Participants who were shown a photograph of the high similarity innocent suspect were unaffected by the biased clothing manipulation, with 16% and 25% identifying the innocent suspect in the biased and unbiased conditions, respectively ($Z = 0.78$, $p = 0.44$). Finally, participants who viewed the low similarity innocent suspect were also unaffected by innocent suspect attire, with 13% and 8% of participants identifying the innocent suspect from the biased and unbiased conditions, respectively ($Z = 0.52$, $p = 0.60$).

Similar-distinct clothing

Two additional conditions were run in which the high and low similarity innocent suspects were presented in a Harley-Davidson T-shirt that was similar to, but not the same as, the T-shirt worn by the target. As explained earlier, this was done to explore the effects of

Copyright © 2006 John Wiley & Sons, Ltd.

presenting an innocent person in clothing that matches the description provided by the witness to some extent, but is not the exact clothing worn by the target. When the high similarity innocent suspect was presented in the similar-distinct clothing, 50% of participants identified him as the target. This proportion of identifications was not significantly different from that obtained in the biased (i.e. same Harley-Davidson) T-shirt condition (50%; $Z = 0.00$, $p = 1.00$). When the low similarity innocent suspect was presented in the similar-distinct clothing, 14% of participants identified him as the target, a significantly greater percentage than those who viewed the unbiased photograph (0%; $Z = 2.07$, $p = 0.02$), and a significantly lower percentage than those who viewed him in the biased photograph condition (37%; $Z = 2.02$, $p = 0.02$).

Clothing descriptions

The primary issue of interest for the accuracy of clothing descriptions is whether participants who gave a correct description of the target's clothing were more likely to make a correct identification decision. It should be noted again, however, that not all participants provided a description of the target. Participants who clearly needed to be excused from the study after the primary dependent measures had been collected (e.g. customers had entered the store during data collection and needed assistance) were dismissed after their confidence judgement had been given. In total, 66.2% of participants in the distinct conditions (N = 153) and 48.6% of participants in the common clothing conditions (n = 72) provided a description of the target's clothing.

Common clothing

A description of the common clothing was considered correct if the participant said that the target wore a blue/grey button-up or plaid shirt. Overall, 52.7% of participants correctly described the common clothing. An additional 31.9% of participants gave an incorrect description (e.g. light T-shirt), 6.9% gave an insufficient description (e.g. dark shirt), and 8.3% stated that they could not remember what the target was wearing. Identification accuracy was unaffected by whether or not a participant gave a correct (89%) or incorrect (78%) clothing description ($Z = 1.11$, $p = 0.13$).

Distinct clothing

A description of the distinct clothing was considered correct if the participant stated that the target wore a black T-shirt. Overall, 62% of participants correctly described the distinct clothing, 28% of participants gave an incorrect description (e.g. red T-shirt), 4% gave a correct but very limited description (e.g. T-shirt), and 7% stated that they could not remember what the target was wearing. Participants who provided a correct description of the distinct clothing were no more likely to make a correct identification decision (83%) than were participants who gave an incorrect description (74%; $Z = 1.17$, $p = 0.12$).

Confidence

To examine the relationship between eyewitness confidence and accuracy, we distinguished between eyewitnesses who chose the suspect from the showup (choosers) and those who did not choose the suspect from the showup (non-choosers). For choosers, accurate eyewitnesses were more confident ($M = 8.88$) than inaccurate eyewitnesses ($M = 7.87$; $F(1, 156) = 21.28$, $p < 0.001$). The confidence-accuracy correlation was

Copyright © 2006 John Wiley & Sons, Ltd.

0.32. Similarly, accurate non-choosers were more confident ($M = 7.88$) than inaccurate non-choosers ($M = 5.96$; $F(1, 219) = 12.29$, $p = 0.001$). The confidence-accuracy correlation was 0.19. Breaking the conditions down further (biased vs. unbiased showups, distinctive vs. common clothing, etc.) produced correlations ranging from 0.09 to 0.42. The size and range of the obtained correlations is consistent with the results of lineup research.

DISCUSSION AND CONCLUSIONS

Several interesting results were obtained in the current study. First, there was a significant difference in overall identification accuracy from common and distinct clothing bias manipulations. When the target wore what was considered to be common clothing, the effects of clothing bias were non-existent. However, when the target wore clothing that was considered to be distinct, the effect of clothing bias on identification accuracy was profound. A second intriguing finding from the current study pertained to the similar-distinct clothing manipulation, where the clothing worn during the event and that shown in the identification procedure were different but similar. The results from these conditions suggest that if a person who resembles the perpetrator is apprehended near the scene of the crime, and is wearing distinct clothing similar to that described by the eyewitness, the likelihood of false identification is considerable. Furthermore, these false identifications are as likely whether an innocent person is wearing identical distinct clothing to that worn by the perpetrator or merely similar clothing. A similar pattern of results was found for the dissimilar, yet plausible looking innocent suspect. In this situation, the false identification rate with the similar-distinct clothing was significantly higher than that obtained in the non-biased condition. Together, these results suggest that the effect of distinct clothing bias on identification accuracy from show-ups is substantial. The fact that the dissimilar innocent suspect was never identified unless he was wearing similar or identical clothing is a clear indication of how powerful the clothing bias can be in identification from show-ups.

Future research needs to be done which addresses issues not yet raised in this paper, including the ability of witnesses to identify clothing from other similar looking items. For example, would it be possible for a witness to distinguish between a lineup of white T-shirts? Or short-sleeved plaid dress-shirts? This could be tested for in any investigation using clothing lineups to help determine if the item is unique enough to be distinguishable from other similar items.

Another important issue that remains unexplored is whether a clothing bias will exist when the witness' description of the clothing is incorrect. In all bias research to date, clothing bias has been manipulated by presenting innocent suspects in attire that matches the clothing **actually** worn during the event (either an exact match or highly similar). When dissimilar clothing is used (control conditions), it is selected to differ considerably from the attire worn during the event. In real cases, police must rely on the witness' description of clothing. As the results in this paper indicate, witnesses are not perfect in their descriptions of clothing. In real world cases this may lead to occasions when police detain a suspect because his clothing matches an inaccurate description provided by an eyewitness. Only further research will determine if the clothing bias effect is present in these situations.

The salience of the clothing worn during the event also raises several interesting issues. Context may influence how salient clothing is to the witness. Strikingly salient clothing in one context (e.g. a military uniform on a university campus) may not be salient at all in another context (e.g. on a military base). This may present serious problems for police and

Copyright © 2006 John Wiley & Sons, Ltd.

courts, as it may be impossible to determine if clothing bias is likely to occur in a given case because it will be difficult to decide if the clothing worn by the criminal was distinctive. As well, salience may vary with witness characteristics. Using the previous example, a military uniform may be either more or less salient to a member of the military than to a civilian.

As with all research, there are limitations that must be acknowledged. The current study relied on a single confederate and thus lacks stimulus generalizability (Wells & Windschitl, 1999). Identification procedures may interact with the appearance of confederates and criminals in non-obvious ways. Our confederate was a large man. His size may have influenced the witnesses either as a memory cue or via conscious inferences. Thus, the fact that such a person is less common than men of slighter build may have increased the impact of the clothing bias. Replication and extension of the findings with other confederates will be necessary to alleviate this concern. Similarly, the high and low similarity innocent suspects were specific individuals and the results may have differed if others had taken their place. Again, only further research can determine how important the degree of similarity to the criminal is in this situation.

In addition, the identification procedure in the current study used photographs to present the targets and innocent suspects, whereas, Flowe et al.'s (2001) results showed that the majority of real-world show-up identifications were conducted live. Live presentation may assist witnesses, helping them to realize that the clothing is not identical to the clothing they saw and thus leading to an inference of innocence. Alternatively, live show-ups may produce a more powerful effect if the reasoning process described earlier was being used by real witnesses. If witnesses ask themselves 'What are the chances that the police could find someone so quickly who matched my description?', they may conclude that the probability is lower when the person is presented live leading to a stronger inference of guilt.

Another limitation of the current study is that the exposure to the target did not involve a 'criminal act', like the majority of laboratory studies that have been conducted over the past 20 years. However, the live, staged-crime thefts that are conducted in eyewitness laboratories cannot be conducted safely in field settings. The likelihood that witnesses in field settings would phone the police or alert security after a 'theft', or even chase after a 'thief' is great, and thus not a wise practice. It also could be argued that witnesses in laboratory studies are usually made aware immediately after a staged event that no real crime has occurred. Thus, their interactions with the target person are effectively no different from the field setting witnesses, with the exception that there is a speedy exit of the target person in laboratory studies. Although there may be a small trade-off between using laboratory staged crimes and field studies, the exposure to the target is, for all practical purposes, similar. In addition, not all crimes that occur in the real-world involve a quick exit of the criminal or knowledge on the part of the witness that a crime has actually occurred. For example, victims of fraud are not aware that their interaction with the person committing fraud is actually a crime. In fact, they may only come to realize this weeks or months after the event has occurred. This type of crime is very similar to the situation used in this study.

Live crimes also can be actually or potentially violent and thus arousing or threatening to the witness. Very few studies mimic this aspect of real crimes. On the other hand, the available data do not support the conclusion that eyewitness accuracy is enhanced by threat or violence (Steblay, 1992). Even potentially violent crimes can produce witnesses who were neither threatened nor aroused. For example, a few years ago a colleague opened a

Copyright © 2006 John Wiley & Sons, Ltd.

door for a man who was limping. Only later, after the man had left the scene, did he learn that the man was limping because he had a shotgun concealed in his pants. He had used the shotgun to rob the bank they had entered together. Although low arousal identification situations are very common and worthy of study and potentially arousing crimes may not generate arousal in all witnesses, the fact remains that the results of this research need not generalize to witnesses who are aroused.

Finally, the distinction between common and distinct clothing is ill defined, relying exclusively on the judgement of the authors. To the extent that this reflects a concern about the salience of the clothing within the context of the event and not just a property of the clothing itself, further research on this issue may be difficult. Salience has a tendency to be defined circularly; stimuli that are salient are those that have an effect and stimuli that do not have an effect were not salient (e.g. Taylor & Fiske, 1978). On the other hand, the different shirts did result in different patterns of correct and incorrect identification decisions and the pattern of effects reported seem to be interpretable in terms of salience or distinctiveness.

In real world cases it will be difficult or impossible to be certain that a clothing description is accurate and to know if the clothing was salient to the witness (though spontaneously providing a clothing description would suggest that it was). The results from the current study suggest that clothing bias can have a powerful impact on the accuracy of show-up identification decisions. Although future studies should investigate the most effective means of avoiding clothing bias, this problem could potentially be avoided by not exposing the witness to the suspect's clothing during an identification procedure. One way of doing this may be to cover the suspect from the neck down with a blanket or jacket. If a photo is used, the picture could be restricted to the face only or by covering the clothing before showing the photo to the witness. However, these suggestions are tentative, and one must be cautious in conducting such a procedure prior to empirical investigations being conducted which show that covering clothing information, and thus effectively covering body cues (e.g. stature, weight), does not reduce identification accuracy. Until a satisfactory procedure is found and successfully tested, any showup procedure conducted where clothing cues are visible may potentially fall prey to the clothing bias effect.

ACKNOWLEDGEMENTS

This research was supported by separate grants to Jennifer E. Dysart, Paul R. Dupuis, and R. C. L. Lindsay from the Social Sciences and Humanities Research Council of Canada.

REFERENCES

Brehm, S., & Brehm, J. W. (1981). *Psychological reactance: A theory of freedom and control.* New York: Academic Press.

Clark, S. E., & Davey, S. L. (2005). The target-to-foils shift in simultaneous and sequential lineups. *Law and Human Behavior, 29*, 151–172.

Dekle, D. J., Beal, C. R., Elliott, R., & Huneycutt, D. (1996). Children as witnesses: A comparison of lineup versus show-up identification methods. *Applied Cognitive Psychology, 10*, 1–12.

Copyright © 2006 John Wiley & Sons, Ltd.

Dupuis, P. R. (2001). Calibrating eyewitness identification accuracy using rank-combined lineups. Unpublished Master's thesis, Queen's University, Kingston, Ontario, Canada.

Dysart, J. E., & Lindsay, R. C. L. (in press). Suggestive technique or reliable method: What do we really know about show-up identifications? In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. Toglia (Eds.), *Handbook of eyewitness psychology: Memory for people*.

Flowe, H., Ebbesen, E., Burke, C., & Chivabunditt, P. (2001 June). At the scene of the crime: An examination of the external validity of published studies on lineup identification accuracy. *Paper Presented at the Annual Meeting of the American-Psychological Society*, Toronto, Ontario, Canada.

Flynn, K. (1998, October 23). Persevering woman helps free stranger in '90 murder case. *New York Times*, p. 1.

Gonzalez, R., Ellsworth, P. C., & Pembroke, M. (1993). Response biases in lineups and show-ups. *Journal of Personality and Social Psychology, 64*, 525–537.

Lindsay, R. C. L., & Bellinger, K. (1999). Alternatives to the sequential lineup: The importance of controlling the pictures. *Journal of Applied Psychology, 84*, 315–321.

Lindsay, R. C. L., & Wells, G. L. (1980). What price justice? Exploring the relationship of lineup fairness to identification accuracy. *Law and Human Behavior, 4*, 303–313.

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

Lindsay, R. C. L., Lea, J., & Fulford, J. (1991). Sequential lineup presentation: Technique matters. *Journal of Applied Psychology, 76*, 741–745.

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 Behavioral Science, Special Issue: Forensic Psychology, 19*, 463–478.

Lindsay, R. C. L., Nosworthy, G. J., Martin, R., & Martynuck, C. (1994). Using mug shots to find suspects. *Journal of Applied Psychology, 79*, 121–130.

Lindsay, R. C. L., Pozzulo, J. D., Craig, W., Lee, K., & Corber, S. (1997). Simultaneous lineups, sequential lineups, and show-ups: Eyewitness identification decisions of adults and children. *Law and Human Behavior, 21*, 391–404.

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.

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

McQuiston, D. E., & Malpass, R. S. (2001). An archival review of eyewitness identifications in criminal felony cases. *Paper Presented at the 4th Biennial Conference for the Society for Applied Research in Memory and Cognition*, Kingston, Ontario, Canada.

Phillips, M. R., McAuliff, B. D., Kovera, M. B., & Cutler, B. L. (1999). Double-blind photoarray administration as a safeguard against investigator bias. *Journal of Applied Psychology, 84*, 940–951.

Pryke, S. (2001). Multiple independent identification decisions: A method of calibrating eyewitness identifications. Doctoral dissertation, Queen's University, Kingston, Ontario, Canada.

Steblay, N. M. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior, 16*, 413–424.

Steblay, N., Dysart, J. E., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 25*, 459–473.

Steblay, N., Dysart, J. E., Fulero, S., & Lindsay, R. C. L. (2003). Eyewitness accuracy rates in police showup and lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 27*, 523–540.

Taylor, S. E., & Fiske, S. T. (1978). Salience, attention, and attributions: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11). New York: Academic Press.

Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement*. Washington, DC: U.S. Department of Justice, Office of Justice Programs.

Copyright © 2006 John Wiley & Sons, Ltd.

Wagenaar, W. A., & Veefkind, N. (1992). Comparison of one-person and many-person lineups: A warning against unsafe practices. In F. Losel, D. Bender, & T. Bliesener (Eds.), *Psychology and law: International perspectives* (pp. 275–285). Berlin: de Gruyter.

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

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25*, 1115–1125.

Yarmey, A. D., Yarmey, A. L., & Yarmey, M. J. (1994). Face and voice identifications in show-ups and lineups. *Applied Cognitive Psychology, 8*, 453–464.

Yarmey, A. D., Yarmey, M. J., & Yarmey, A. L. (1996). Accuracy of eyewitness identifications in show-ups and lineups. *Law and Human Behavior, 20*, 459–477.

Copyright © 2006 John Wiley & Sons, Ltd.

Journal of Applied Psychology
2004, Vol. 89, No. 3, 524–541

Copyright 2004 by the American Psychological Association
0021-9010/04/$12.00 DOI: 10.1037/0021-9010.89.3.524

Timing of Eyewitness Expert Testimony, Jurors' Need for Cognition, and Case Strength as Determinants of Trial Verdicts

Michael R. Leippe, Donna Eisenstadt, Shannon M. Rauch, and Hope M. Seib
Saint Louis University

In 2 experiments, college students read a murder-trial transcript that included or did not include court-appointed expert testimony about eyewitness memory. The testimony either preceded or followed the evidence, and the judge's final instructions reminded or did not remind jurors about the expert's testimony. Expert testimony decreased perceptions of guilt and eyewitness believability when it followed the evidence and preceded the judge's reminder. This effect occurred whether the prosecution case was moderately weak or moderately strong. Jurors' need for cognition (NC) was curvilinearly related to convictions in a strong case. Low and high NC jurors convicted less than did moderate NC jurors. Greater scrutiny by high NC jurors may make them more likely to consider evidence for the weaker side.

Jurors are asked to perform the difficult tasks of understanding, evaluating, and integrating evidence on both sides of the important issue of whether a criminal defendant is guilty. The defendant's liberty (and sometimes life), as well as the service of justice, depends on whether jurors are more persuaded by the incriminating evidence and arguments presented by the prosecution or the exculpatory evidence and arguments advanced by the defense. Psycho-legal researchers have devoted considerable attention to the nature of jury decision making and the factors that influence it (cf. Hastie, 1993a; Kerr & Bray, 1982). The research reported here sought to contribute to this literature by gaining new insights into two important aspects of evidence evaluation. The first of these involves eyewitness evidence and whether expert testimony about eyewitness evidence alters jurors' reactions to it. Research suggests that eyewitness identifications strongly influence jurors, perhaps more strongly than they should under some circumstances (see, e.g., Leippe, 1994, 1995; Wells, 1984). Because eyewitness testimony is most commonly presented as incriminating evidence, its potency may represent an asymmetry in the comparison of prosecution and defense evidence, one that favors the prosecution. The present study examined how and under what conditions expert testimony that seeks to educate jurors about human memory in eyewitness contexts might alter the incriminating value of eyewitness evidence (Leippe, 1995).

The second aspect of interest concerns the individual-differences variable of need for cognition (NC; Cacioppo & Petty, 1982). This characteristic of jurors may have implications for how prosecution and defense evidence is interpreted and weighted because NC correlates with how carefully and systematically jurors will consider the evidence. Because they will very carefully scrutinize the evidence, for example, very high NC jurors may more readily find a reasonable doubt in a reasonably strong prosecution case—an asymmetry in evidence evaluation that, in this case, would favor the defense.

Expert Testimony About Eyewitness Memory

Scientific knowledge continues to accumulate about eyewitness memory, factors that affect eyewitness reports and identifications, and the degree to which people appreciate these influences. The reasons for the high level of sustained research attention to eyewitness testimony are manifold and certainly include fascination with the myriad psychological processes involved in forming a memory, reporting it, and having it believed. Equally or more influential is a set of reasons that converge into a recipe for miscarriages of justice. First, eyewitness testimony powerfully influences legal decisions, from charging someone with a crime to finding a criminal defendant guilty. In the courtroom, eyewitness accounts are among the most common and influential forms of testimony (Leippe, 1995; Loftus, 1983; Woocher, 1977), and frequently (i.e., perhaps in close to 80,000 trials in the United States per year; Goldstein, Chance, & Schneller, 1989) eyewitness evidence is the primary evidence against the defendant. Second, at least some of the impact of eyewitness testimony may be undue, as there is evidence that jurors tend to overbelieve eyewitnesses (e.g., Brigham & Bothwell, 1983; Wells, 1984), have insufficient understanding of the factors that affect memory (e.g., Kassin & Barndollar, 1992; Lindsay, 1994), and are overly swayed by eyewitness confidence, which is not very diagnostic of accuracy and apt to be inflated by the time the eyewitness reaches the courtroom (cf. Leippe, 1980, 1995; Wells & Murray, 1984; Whitley & Greenberg, 1986). Third, eyewitness identification and memory may be highly prone to error (e.g., Buckhout, 1974; Haber & Haber, 2001; Leippe, Wells, & Ostrom, 1978). These factors combine to create realistic concern about convictions derived from false identifications. The legitimacy of this concern has been advanced signifi-

Michael R. Leippe, Donna Eisenstadt, Shannon M. Rauch, and Hope M. Seib, Department of Psychology, Saint Louis University.

Portions of this research were presented at the annual meeting of the American Psychology-Law Society in Austin, TX in March 2002 and at the Eyewitness Experts in Court Symposium at the International, Interdisciplinary Conference on Psychology and Law in Edinburgh, Scotland in July 2003.

We thank Deanna Brady, John Hicks, and Lisa Votraw for their assistance with this research.

Correspondence concerning this article should be addressed to Michael R. Leippe, Department of Psychology, Saint Louis University, 3511 Laclede Avenue, Shannon Hall, St. Louis, MO 63103. E-mail: leippemr@slu.edu

cantly with the advent of DNA testing. As of 2000, of the 62 cases in which individuals convicted and imprisoned for felonies were later exonerated by DNA evidence, 52 (including 8 with death sentences) involved mistaken eyewitness identifications (Scheck, Neufeld, & Dwyer, 2000). Mistaken identifications, resulting in loss of liberty or even life, are thus a realistic worry and underscore the practical importance of research that may be informative about how to reduce their commission or convictions based on them.

One approach to reducing false identifications is to educate law enforcement officials about how to collect eyewitness evidence in ways that do not "pull" for false identifications, inflated confidence, or both in identifications (e.g., suggestive questioning, repeated questions, biased lineup instructions, and poorly constructed lineups). An important step along these lines has been made with the 1999 publication sponsored by the Department of Justice of *Eyewitness Evidence: A Guide for Law Enforcement,* a "how-to" manual based on scientific research and the deliberations of research psychologists, attorneys, and police officers (Wells et al., 2000). Widespread adoption of the recommended procedures for collecting eyewitness evidence is quite likely to reduce the prevalence of false identifications that result, largely, from how eyewitnesses are treated and tested. There are other causes of eyewitness error, however, and a substantial likelihood that even unreliable eyewitness testimony will be greeted with acceptance and belief by jurors. Accordingly, expert testimony about eyewitness memory has also been promoted as a "remedy" and has been admitted into courtrooms with increasing frequency in the past decade (Kassin, Tubb, Hosch, & Memon, 2001; Leippe, 1995).

Leippe (1995) defined eyewitness expert testimony as "the delivery to a jury by a qualified research psychologist of information about research and theory on eyewitness behavior" (p. 910). The major arguments in favor of expert testimony are that jurors have insufficient knowledge about how memory works and the factors that affect it, that jurors' naive theories of memory lead them to overbelieve eyewitness accounts, especially those that are confidently delivered, and that scientific research has provided reliable insights into eyewitness memory that can help jurors better judge the veracity of eyewitness testimony. On the other hand, critics of eyewitness expert testimony argue that, among other problems, expert testimony may cause jurors to become overly skeptical of eyewitness evidence, "overcorrecting" to the point of *under*belief.

A number of mock-juror and trial-simulation studies have been conducted to assess the likely impact of eyewitness experts in the courtroom. In a review of this research, Leippe (1995) counted 12 experiments that compared the presence versus absence of eyewitness expert testimony in one of three combinations of form and context. One method has been to present a brief written summary of a trial, including the main evidence, witnesses on both sides, and a description of a psychologist's expert testimony (e.g., Loftus, 1980; Maass, Brigham, & West, 1985). A second approach involves presenting a "minimal trial" consisting of a written summary of the crime and the case, as well as videotaped eyewitness testimony (under direct and cross-examination) and expert testimony (e.g., Fox & Walters, 1986; Wells, 1986). Finally, expert testimony has also been studied as part of a videotaped mock trial (e.g., Cutler, Penrod, & Dexter, 1989; Lindsay, 1994). The expert testimony in most of the studies, regardless of method, was limited to a general overview of what scientists know about eyewitness memory and behavior, typically with a gentle implication that eyewitnesses err more often than one would suspect. A few studies had expert testimony that spoke directly to eyewitness factors in the case at hand, without offering an ultimate opinion about accuracy. Across all of these variations, however, the dominant result—evident in 10 of the 12 studies—was heightened *skepticism* about the eyewitness testimony, leading to a decrease in guilty verdicts, decreased belief in the eyewitness, or both. Eight of the experiments allowed some test of sensitivity—whether the expert testimony made mock jurors more (or less) likely to take into account what the expert testimony suggested were valid (or invalid) cues to accuracy. Heightened sensitivity was found in two of the eight tests.

Thus, the most common finding is that expert testimony about eyewitness memory heightens jurors' skepticism about eyewitness evidence. Given the apparent tendency to overbelieve eyewitnesses, this may not be an inappropriate outcome, especially if the prosecution's case relies primarily on eyewitness evidence, which itself is limited (e.g., there is only one eyewitness). In the present research, we used a trial simulation to seek further insight into the skepticism effect, particularly about some of the trial conditions under which it might be especially likely to occur.

The Strength of the Prosecution's Case

One such condition of interest was the strength of the prosecution case in the absence of an identifying eyewitness. When the prosecution has considerable circumstantial or physical evidence in addition to an eyewitness identification, a decrease in guilty verdicts that is due to increased skepticism regarding the eyewitness may not be an especially desirable effect of expert testimony. Among the 64 eyewitness experts surveyed by Kassin et al. (2001), 19 indicated that one of the reasons they had declined a request to testify was because they personally believed the defendant was guilty. This may reflect, in part, a worry that they might cause skepticism about the eyewitness evidence, which might spread somehow to the rest of the case. For example, research on the impact of a positive eyewitness identification has found that, in some cases, it continues to have incriminating influence even when it is discredited by assailing the visual capability of the eyewitness (Loftus, 1974; Saunders, Vidmar, & Hewitt, 1983). One interpretation of this persistence is that identifying testimony creates an interpretational bias such that more weight or attention is given to other incriminating evidence, and this greater receptivity to the prosecution's case persists even if doubt is cast on the eyewitness (Saunders et al., 1983). Perhaps a similar, but directionally reversed, effect can occur for expert testimony that implicitly or explicitly calls eyewitness evidence into question. Jurors may reason that, if a highly credible expert suggests that one piece of evidence presented by the prosecution (the eyewitness) can be doubted, perhaps other prosecution evidence can be doubted as well. Consistent with this idea, Leippe, Seib, Eisenstadt, and Alia-Klein (1998) obtained suggestive evidence that, indeed, expert testimony could "take down" a fairly strong prosecution case.

On the other hand, there is reason to believe that the impact of eyewitness expert testimony will decrease as the strength of extra-eyewitness, incriminating evidence increases. As the strength of evidence increases, the ability to interpret it in a fashion that fosters disbelief should decrease. Even if the weight given to the eyewitness is decreased by expert testimony, an averaging model of information integration (e.g., Kaplan & Kemmerick, 1974;

Moore & Gump, 1995) would suggest that the impact of this discounting of the eyewitness would be increasingly diluted as the amount of less assailable other evidence increases. Given the strong likelihood that eyewitness evidence carries disproportionately larger weight, however, it might take a great deal of other evidence to make up for the removal of the eyewitness from the equation. Thus, unless the prosecution case is extremely strong, eyewitness expert testimony that discredits the eyewitness should reduce, somewhat, perceptions of guilt.

The Timing of the Expert Testimony

The impact of eyewitness expert testimony on evaluations of both the eyewitness and other evidence may depend on when the expert takes the stand. Expert testimony, theoretically, could be presented either before or after the eyewitness testimony (and other evidence) is presented. Among the dozen studies of expert testimony reviewed earlier, the expert has preceded the eyewitness in only two studies (Wells, Lindsay, & Tousignant, 1980; Wells & Wright, 1983, reported in Wells, 1986). In neither of these studies were the expert and the eyewitness embedded in a trial or trial summary. Rather, the expert, presenting "psychological advice," and an actual eyewitness to an earlier staged crime, were presented on a videotaped recording. The results were mixed. Wells et al. (1980) found the typical skepticism effect of reduced belief in the eyewitness, but no sensitivity effect was found in that both accurate and inaccurate eyewitnesses were equally believed, even with expert testimony. Wells and Wright, in contrast, found that expert testimony promoted sensitivity but no skepticism effect—both rare findings in this domain. Unlike previous research, the present study examined the effects of preevidence eyewitness expert testimony in a "whole-trial" simulation and directly compared the effects of pre- and postevidence expert testimony within the same experiment.

In most North American trials, eyewitness expert testimony is introduced toward the end of the trial, after the eyewitness has testified and after the prosecution has presented its case. This is because eyewitness experts most commonly are called by the defense, which presents its case after the prosecution has rested. Compared with an earlier preevidence presentation, such postevidence expert testimony has the advantage of recency. Coming late, the testimony may be more likely to be spontaneously recalled than other evidence when jurors deliberate and may be subject to fewer compelling counterarguments, to the extent that it cannot be followed by an opposing expert. These recency advantages would be especially likely if the case is complex and tedious and jurors' attention and comprehension suffer or they lapse into more heuristic (and less systematic) processing (Cooper, Bennett, & Sukel, 1996; Horowitz, Bordens, Victor, Bourgeois, & ForsterLee, 2001). Expert testimony under these conditions might serve as a cue to discount the already-heard eyewitness account. The advantage of recency might also surface if memories and evaluations of the eyewitness testimony can be retroactively influenced by new data (i.e., the expert testimony; see, e.g., Snyder & Uranowitz, 1978).

On the other hand, the typical postevidence testimony of an eyewitness expert might have diminished ability to influence evaluations of either the eyewitness or the other evidence. Jurors will already have forged a story or theory of the crime at hand (Pennington & Hastie, 1992). Because eyewitnesses have come and gone off the witness stand, jurors are not able to apply online any expert advice about diagnostic cues (e.g., reports of witnessing conditions) or miscues (e.g., confidence) as eyewitnesses testify (Leippe, 1995). From this perspective, postevidence expert testimony would seem unlikely to heighten either skepticism or sensitivity to credibility-relevant aspects of the eyewitness testimony.

The pros and cons of postevidence expert testimony, of course, are reversed for expert testimony that precedes the evidence. If the expert testimony precedes the evidence, it has the potential to influence the crime theory developed by jurors as they receive the early evidence. For example, if the expert's advice or overview compels jurors to be more skeptical of the eyewitness evidence, the eyewitness(es)' account of the crime and the perpetrator may play less of a role in shaping the story. Instead, jurors might focus more on how the conditions of witnessing or testing may have hindered accurate identifications. In addition, if the expert precedes the eyewitness, jurors might be able to apply any advice about cues to memory accuracy as the eyewitness testifies. These advantages of preevidence expert testimony, however, would seem to require that expert testimony be explicit in terms of what is and is not diagnostic of accuracy in the eyewitness's statements and behavior and specific in terms of what aspects of the eyewitness testimony are at issue. When the expert testimony is a more general overview about what is known scientifically about eyewitness memory, as in the present study (and in some actual trials), jurors may not be able to apply much of what the expert says to their appraisal of the eyewitness, and their theories may not be altered. In addition, of course, the availability and discounting-cue advantages associated with recency are lost if the expert testimony comes early and an entire trial fills the interval between the expert's information and juror deliberation.

To compare the impact of pre- and postevidence testimony in the present study, the eyewitness expert was presented as a court-appointed friend of the court in all conditions, as this would seem to be the only way in which an eyewitness expert could advise a jury in an actual trial before evidence was heard. Moreover, because it is not permissible to discuss evidence before it is presented, the expert gave only a general overview of what is known about eyewitness memory, never mentioning the specific case or singling out the factors relevant to the case at trial.

Mere General-Knowledge Expert Testimony

In fact, as noted earlier, most previous studies of eyewitness expert testimony have presented mock jurors primarily with such "general-knowledge testimony," typically a "minilecture" that overviews the methods and findings of the scientific study of eyewitness memory (Leippe, 1995). Only a few studies have presented expert testimony specifically tailored to the case at hand. A moderate skepticism effect has been the rule, regardless of the specificity of the testimony. By presenting expert testimony that was completely limited to a general overview and that did not single out any problematic witnessing or testing factors in the case at hand, we sought to provide a stringent test of the capacity of eyewitness expert testimony to create skepticism and reduce perceptions of guilt. That is, especially when the case for guilt is relatively strong, in that it includes multiple pieces of physical and circumstantial evidence, the finding that even a general discussion by a psychologist about eyewitness memory reduces convictions would strongly suggest that the mere presence of an eyewitness expert implies to jurors that there is some reason to doubt the

eyewitness and, as a result, may reduce the persuasive impact of the prosecution's case, even when, arguably, it should not.

The Impact of a Judge's Reminder

A traditional "safeguard" against the inappropriate impact of eyewitness evidence is for judges to include in their end-of-trial instructions to the jury guidelines about how to judge the credibility of eyewitness testimony. Perhaps the most common of these are the pattern instructions laid down in *United States v. Telfaire* (1972; cf. Greene, 1988). However, judges also can and do prepare and present their own instructions, which may include their own intuitions as well as observations about the case at hand. Indeed, the senior author of this article, Michael R. Leippe, was once asked by a judge in a federal case to help him write the instructions he would provide to the jury in a specific case (in lieu of the author being allowed to give expert testimony at trial). In general, mock-jury research suggests that judges' instructions, if they have any effect at all, increase juror skepticism about the eyewitness evidence (cf. Leippe, 1995). They do not appear to increase sensitivity to eyewitness factors that might be diagnostic, perhaps partly because instructions modeled after *Telfaire* and the *Neil v. Biggers* (1972) criteria present some incorrect information about eyewitness memory (Brigham, Wasserman, & Meissner, 1999; Wells & Murray, 1983). A different role for judges' instructions was examined in the present research. Because of the discretion they enjoy, and especially if they call the expert themselves, judges could use their final instructions to remind jurors of the *expert's* testimony, summarizing the expert's points and suggesting that jurors apply them in their evaluation of eyewitness credibility. Such a reminder and summary would be close to the "last words" heard about the evidence before deliberation and would come from a trustworthy and impartial authority figure. Thus, to the extent that the mere presence of eyewitness expert testimony serves as an implicit suggestion to distrust the eyewitness evidence against the defendant, a judge's reminder is likely to enhance this exonerating influence for two reasons: The reminder is temporally close to deliberation (recency) and enhances the credibility of the cue (expert) to discount the eyewitness. In effect, a judge may be *authorizing* or *legitimating* heightened skepticism about an eyewitness.

Predictions for Expert-Testimony Effects

Given the pervasiveness of the skepticism effect of eyewitness expert testimony in past research, at least weak predictions regarding two factors can be made for the present study. First, expert testimony should lead to decreased perceptions of guilt and decreased believability of the eyewitness no matter when it is presented, regardless of whether there is also a judge's reminder to consider what the expert said (Hypothesis 1). Second, this heightened skepticism effect should occur whether the prosecution case in addition to the eyewitness evidence is relatively weak or relatively strong (Hypothesis 2).

Should the impact of expert testimony be especially large in any specific expert testimony condition? We believe so, given the content of the expert testimony and the nature of the eyewitness evidence in the trial scenario under study. As noted previously, the expert testimony in the present research was general-knowledge expert testimony, in which no reference was made to any specific factors that might be relevant to the case at trial. In addition, the single eyewitness to the crime in the case was not a remarkably strong or weak eyewitness. There were no outstanding aspects of his testimony that would automatically constitute a "fatal flaw" to his credibility. As in many actual trials that involve expert testimony, the issue that brought forth the eyewitness expert was the fact that he was the sole eyewitness, not that he was an especially "bad" eyewitness. Under these circumstances of general-knowledge testimony, and what we might call "reasonable-but-limited eyewitness evidence," it seems most likely that the skepticism-arousing effects of expert testimony would be strongest when the expert testimony is made salient after the evidence is presented, and as near as possible to the end of the trial, and when the late-breaking expert testimony is bolstered by a judge's reminder to use the expert testimony. In other words, the largest skepticism effect should occur in a postevidence-testimony-with-reminder condition (Hypothesis 3). The reasoning behind this specific hypothesis is threefold.

1. General-knowledge testimony should not facilitate the primacy advantages described earlier, especially when the eyewitness evidence has no fatal flaw. General-knowledge testimony provides few guidelines for evaluating and comparing specifics about the eyewitness. Thus, the amount of a priori skepticism created about the eyewitness may be limited and, to the extent that the expert does not point to specific problems with the forthcoming eyewitness testimony and the testimony itself has no fatal flaw, the eyewitness may hold up reasonably well on dimensions described in general terms by the expert. As a result, jurors may form a not-too-skeptical impression of the eyewitness early, and this impression may persist through to the end of the trial, while the expert testimony may lose its salience over time.

2. If it is not facilitative of primacy influences, the major remaining role of general-knowledge expert testimony should be to serve as a general cue to be skeptical of eyewitness evidence. This role is best served when the expert testimony comes late in the trial, after all evidence is presented and just before jurors must decide the case. Under these circumstances, the expert testimony should be maximally salient and accessible as jurors begin their deliberation. In addition, these circumstances permit a backward application of the skepticism cue to the eyewitness evidence. Newly alerted that there may be cause to be skeptical of the eyewitness testimony, jurors may worry about whether they initially overbelieved the eyewitness. As a result, they may attempt to "correct for bias" (Wegener & Petty, 1997), the bias in this case of overbelieving the eyewitness that is implied by the expert's general overview. It should be easier to make this "correction" and be more skeptical when the expert testimony follows (vs. precedes) the evidence, in that it is probably easier to "see" reasons to doubt eyewitness testimony when it is being reviewed in memory than when it is being reviewed online. Interpretations of remembered events (i.e., the eyewitness's statements) are less bound by reality constraints than are interpretations of the events as they occur.

3. Exercising such enhanced skepticism requires authorization. Receiving the expert testimony after the trial evidence may facilitate the ability to apply it retrospectively to the eyewitness evidence. Actually doing so in an especially skeptical manner, however, should depend on the judge's reminder, which, in such close temporal proximity to the expert testimony, communicates a tacit suggestion to be skeptical about the eyewitness evidence and serves to authorize the application of the expert testimony in this skepticism-enhancing way.

In summary, we hypothesize that the exonerating effects of eyewitness expert testimony will be especially strong when the trial situation includes two circumstances that facilitate interpretation and use of general expert testimony as a cue to be skeptical of the guilt of the eyewitness. The recency of the expert testimony will help mock jurors remember to apply it in judging retrospectively the eyewitness evidence and enhance the likelihood that the application will turn up reasons for skepticism. The judge's reminder will provide authorization to apply the expert testimony in a critical manner.

Jurors' Need for Cognition

An additional focus in this research was whether jurors would be differentially influenced by case evidence as a function of their level of NC (Cacioppo & Petty, 1982; Cacioppo, Petty, & Kao, 1984). NC refers to the extent to which people enjoy and spontaneously exert effort in thinking. Past research has found relatively wide variability on this dimension and established that high NC individuals are more likely to exert cognitive effort in evaluating complex stimuli they encounter, including persuasive communications about social issues, videotaped events, social-interaction vignettes, advertisements, and, in fact, evidence from courtroom trials (see Cacioppo, Petty, Feinstein, & Jarvis, 1996, for a review). For example, a number of studies have demonstrated that high NC individuals have a more differentiated reaction to strong-argument and weak-argument messages than do low NC individuals, suggesting that they processed the messages more systematically so as to discern their strength (e.g., Cacioppo, Petty, & Morris, 1983; Smith & Petty, 1996).

On the face of it, these findings suggest that, in a courtroom trial, NC should moderate the overall impact of prosecution case strength. Higher NC jurors should be more persuaded to convict than their lower NC counterparts when the prosecution case for guilt is strong and less likely to convict when the prosecution case is weak. However, the nature of the decision faced by jurors may complicate this picture. In a trial, there are two sides to consider, even though one side (e.g., the prosecution) may have the stronger case. Exceptionally high cognitive elaboration may result in greater recognition of the merits of the weaker side, leading some very careful scrutinizers to perceive lower likelihood of guilt in a strong case for guilt and greater likelihood of guilt in a weak case than they would have, had they elaborated less. Thus, at extremely high levels of elaboration, the discrimination of moderately strong and moderately weak cases for guilt may be less in terms of differences in conviction rates because the side with less evidence (either the prosecution or the defense) has a stronger impact than it would if elaboration had been less. Because NC is correlated with elaboration likelihood and high effort at integrating information from opposing communications (Petty, Tormala, Hawkins, & Wegener, 2001), this suggests a *curvilinear* relationship between NC and the degree of discrimination of strong and weak cases for guilt, at least when the strong case is not overwhelmingly incriminating and the weak case is not entirely without merit (Hypothesis 4). Compared with those who are low in NC, jurors with moderately high NC should discriminate more (in terms of judgments of guilt) between a relatively strong and a relatively weak prosecution case because their level of elaboration makes them more aware of the relative strength of evidence of the two sides. However, jurors with extremely high NC should discriminate less than jurors with moderately high NC because their even higher level of elaboration will make them more aware that, despite relative differences in strength of evidence, the side with less evidence, nevertheless, has evidence.

The moderation of guilt judgments among those very high in NC would seem especially likely when the prosecution case for guilt is relatively strong. In North American criminal trials, jurors are asked to convict only if they find the defendant to be guilty beyond a reasonable doubt. The reasonable doubt standard creates a high burden of proof for the prosecution and a decision criterion in which, in theory, a single credible argument to doubt guilt should be sufficient to acquit the defendant. It is not surprising, then, that jurors have been found to exercise a leniency bias (i.e., a bias toward acquittal) in cases in which reasonable evidence exists on both sides (MacCoun & Kerr, 1988). Given the high burden of proof associated with the beyond-a-reasonable-doubt standard, exceptionally careful scrutiny by jurors in a trial in which the prosecution evidence is strong but not wholly unimpeachable is likely to lead many such jurors to find some basis for reasonably doubting the guilt of the defendant and, in turn, voting for acquittal. Thus, very high NC individuals may be more likely than moderately high NC individuals to uncover defense evidence or arguments that configure into a reasonable doubt, and therefore to show, in essence, a leniency effect.

Note that we are not suggesting here that, in the moderately-strong-case-for-guilt scenario, very high NC jurors will assign a different criterion or value in defining what constitutes *reasonable doubt*. High NC jurors, for example, might require a higher degree of certainty that the defendant committed the crime than lower NC jurors (e.g., 70% vs. 50%) and, thus, be more likely to acquit in a prosecution case that conveys a high, but not high enough, level of certainty for them (e.g., 65%). Instead, our hypothesis is that, unless the prosecution case is wholly unassailable, high NC jurors will be more likely to see a basis for doubt in the evidence because they will analyze it more closely. Thus, even if their criterion of reasonable doubt is the same as that of less contemplative jurors, their interpretation of the same evidence will lead to a lower level of certainty of guilt that may fall below the criterion. It is well established that the same evidence can be argued and structured for presentation in different ways by lawyers, with resultant differences in how jurors evaluate the evidence (Schum, 1993). Similarly, it would seem, the same evidence can be structured differently in the active, interpretational minds of jurors, as research on how certain attitudinal and personality differences are related to juror decisions suggests (see, e.g., Ellsworth, 1993, and, as an example involving NC, Kassin, Reddy, & Tulloch, 1990).

Experiment 1

In Experiment 1, college students were exposed to a courtroom trial of a robbery/murder case in the form of a transcript that included the attorneys' arguments, judge's instructions, and direct and cross-examination of the key witnesses, including the single eyewitness to the crime. The strength of the prosecution's case, in terms of amount and clarity of evidence, was manipulated such that some mock jurors read a trial containing a stronger prosecution case and others read a trial with a weaker prosecution case. The trial also included or did not include the question-and-answer testimony of a psychologist called by the court to "describe what science can tell us about the factors that affect the reliability of eyewitness reports," and, when included, this expert testimony was introduced either before or after any evidence or witnesses were presented. In addition, the judge, in his final instructions, either did or did not include a reminder and summary of what the expert said.

The hypotheses tested in Experiment 1 may be summarized as follows:

Hypothesis 1: Eyewitness expert testimony should lead to decreased perceptions of the defendant's guilt and of the eyewitness's believability (i.e., a skepticism effect) whether it precedes or follows the evidence and whether it is accompanied by a judge's reminder. That is, compared with the no-expert control condition, perceptions of guilt and believability should be lower in all expert-testimony conditions.

Hypothesis 2: The skepticism effect should occur similarly whether the noneyewitness prosecution case is relatively strong or relatively weak. Thus, there should be no interaction between case strength and type of expert testimony on perceptions of guilt and eyewitness believability.

Hypothesis 3: The largest skepticism effect should occur when expert testimony follows the evidence and is accompanied by a judge's reminder. Compared with the other three expert-testimony conditions (preevidence/reminder, preevidence/no reminder, and postevidence/no reminder), as well as, of course, the no-expert condition, perceptions of guilt and eyewitness believability should be lower in the postevidence/reminder condition.

Hypothesis 4: In judging guilt, mock jurors with moderate NC should discriminate between the relatively strong and relatively weak cases more than either low NC or high NC mock jurors. Thus, an interaction between NC and prosecution case strength should be found on measures of perceived guilt: When the case is relatively strong, perceived guilt will be higher among moderate NC mock jurors than among mock jurors with low or high NC; when the case is relatively weak, perceived guilt will be lower among mock jurors with moderate NC, compared with their low and high NC counterparts.

Method

Participants and Design

Three hundred and eighty-five male (n = 156) and female (n = 229) undergraduates at a large Midwestern university participated to earn credit toward a psychology course requirement. They were assigned randomly to 1 of 10 conditions corresponding to a 2 (case strength: stronger or weaker prosecution case) × 5 (type of expert testimony about eyewitness memory: preevidence/no-judge's-reminder, preevidence/judge's-reminder, postevidence/no-judge's-reminder, postevidence/judge's-reminder, or no-expert) factorial design.[1,2] Fourteen participants, relatively evenly distributed across conditions, did not properly complete the dependent measures; their data were not analyzed.

Procedure

Participants privately read the transcript of a robbery/murder case and completed dependent measures in group sessions ranging in size from 5 to 50 participants. Participants were seated far enough apart so that they could not see other participants' materials or interact with them in any manner. The experimenter distributed booklets containing instructions and a transcript of the case of *State of Missouri v. David Lucas* (a fictitious case with elements drawn from actual cases familiar to us). The booklets were distributed from a shuffled pile of booklets that included transcripts for all conditions. The instructions indicated that the study was concerned with how people react to evidence and testimony in criminal court cases. Participants were asked to "take the role of a juror who is receiving this information in court and will later deliver a verdict concerning the guilt or innocence of the defendant" and to "try to evaluate all of the evidence and consider how it operates together." Reading of the transcript was self-paced. They were implored to "attend to the case as if you were actually serving as a juror." Per instructions, after participants had read the trial transcript, they closed their booklet and placed it aside on their desk. The experimenter came to a given participant's desk, collected the transcript booklet, and handed the participant a second booklet containing the dependent measures and, on the last page, the NC Scale. Participants were debriefed, thanked, sworn to secrecy, and dismissed after they completed all measures.

Core trial and transcript. The transcript concerned the trial of a robbery/murder case involving a holdup at knife point at night on a deserted street that led to a fatal stabbing. The sole eyewitness was an adult male who observed the crimes from a second-floor bedroom window. Participant jurors initially learned what allegedly happened through reading the first sentences of the prosecution's opening arguments. The transcript consisted of critical excerpts from the trial, including (a) opening prosecution arguments; (b) opening defense arguments (which were waived until after the prosecution presented and rested its case); (c) closing arguments by the defense followed by the prosecution; (d) direct and cross-examination of the eyewitness (the prosecution's first witness) and several other witnesses; and (e) the judge's opening statements, introduction of the eyewitness expert, and final instructions to the jury. The trial transcripts were approximately 6,500–8,000 words, depending on condition.

Prosecution case strength. Whereas the eyewitness evidence was identical in all conditions, circumstantial evidence was varied to create a stronger and a weaker prosecution case. Evidence presented in the *stronger case* implicated the defendant on several scores. For example, the money

[1] The percentage of participants who were women did not differ significantly across the conditions of the experiment, χ^2 (9, N = 371) < 1, ranging from 51% to 66%.

[2] The eight conditions in which expert testimony was presented also conform to a 2 (case strength) × 2 (timing of expert testimony: pre- or postevidence) × 2 (judge's reminder: present or absent) between-subjects factorial design. Because a data analysis plan centered on this three-way factorial would have difficulty incorporating the no-expert level of type of testimony, which is involved in major hypotheses of the study, we opted for a data-analytic approach based on the 2 × 5 design. Analyses of the major dependent variables that included the timing and judge's reminder as separate, crossed factors, revealed no outcomes that added to or altered the conclusions drawn from the 2 × 5 analyses presented later.

found in the defendant's clothing nearly equaled what was stolen, the blood on the defendant's clothing was a type shared by only 20% of the population yet it matched the victim's blood type, there was clear bodily evidence on the defendant of a struggle (e.g., swollen eye, scraped knuckles), and the defendant's alibi witness was vague about the time interval he claimed to have known the defendant's likely whereabouts. In contrast, evidence in the *weaker case* was more questionable on each score in terms of how incriminating it was. The money found in the defendant's clothing was less than half the amount stolen, the blood type found on the defendant's clothing was shared by up to 65% of the population, no evidence of a bodily struggle was presented, and the defendant's alibi witness claimed times of knowing the defendant's whereabouts that would have left only a barely sufficient amount of time to go to the crime scene, commit the crime, and return to the location specified by the alibi witness.

Timing and nature of expert testimony. In the *preevidence expert* conditions, eyewitness expert testimony was delivered from the witness stand immediately following the prosecution's opening arguments, before any evidence, including the eyewitness, was heard. In the *postevidence expert* conditions, the testimony was delivered immediately after the final defense witness, after all evidence, including the eyewitness, and both sides of the case had been presented. In the *no-expert* control conditions (for both the stronger and the weaker case), no expert was called at all; instead, before closing arguments, the judge briefly noted that jurors are the sole judges of witness credibility. In both the preevidence and postevidence expert conditions, the eyewitness expert, a male experimental psychologist, employed as a professor at a major research university, was called by the judge as a friend of the court, who introduced him with the following instructions:

> Because an eyewitness's report is a crucial part of the evidence in this case, I have called forth an expert on eyewitness memory to describe what science can tell us about the factors that affect the reliability of memory reports. The jury is advised that the professor will be discussing a general theory of memory and research findings that illustrate the theory. He will not, nor can he, speak directly to the credibility of the witnesses in this case. Nor has he personally interviewed the witnesses. The jury is solely responsible for judging the credibility of eyewitnesses. The court asks that you use this scientific information provided by the expert as you see fit while considering the eyewitness reports that you will have heard.

The judge proceeded to question the expert, first asking him to describe his educational and occupational background and areas of professional expertise, as well as, in a series of questions, requesting that the expert describe "the general theory of memory and the important factors in eyewitness identification." In his testimony, the expert did not make any reference to the case at trial, nor did he focus his testimony explicitly on any eyewitness variable at issue in the case. Instead, the expert summarized the active, integrative nature of human memory and the stages of perception, encoding, storage, and retrieval. He also briefly described the scientific procedures used to study eyewitness memory, the greater problems of remembering once-seen faces versus familiar ones, the weak link between eyewitness confidence and accuracy, the curvilinear relationship between stress and memory accuracy, weapon focus, the forgetting curve, effects of physical conditions of witnessing, and the importance of proper and unbiased lineup construction and instructions. The expert pointed out that staged crime studies have "shown that, depending on conditions, anywhere from 15% to 85% of the witnesses may choose a wrong person from a lineup." He concluded his testimony by recommending that

> a person attempting to judge whether an eyewitness had identified the criminal or an innocent suspect should avoid placing too much faith in the eyewitness's confidence . . . and, instead focus more on the situational factors present during witnessing, questioning, and lineup administration that may have helped or hindered memory.

The judge's final instructions and reminder about the eyewitness expert. In all conditions, the transcript ended with final instructions from the judge, which included definitions of the charge (murder in the second degree) and of reasonable doubt, as well as the admonition that jurors are the "sole judges of the credibility of each witness and of the importance of his or her testimony." In the *reminder* condition, the judge supplemented these instructions by reminding jurors of the eyewitness expert, stating the following:

> Among other things, I believe it is important for you to take into consideration the following things which [the expert] noted: (a) memories are not necessarily permanent and may change as new information is combined with older information in memory; (b) recognition of faces seen only once and briefly is difficult; (c) an eyewitness's level of certainty in the courtroom is not always a good indicator of the eyewitness's accuracy; (d) most important, as jurors, you should pay the most attention to how good the viewing conditions were when the memory was formed and how good and fair a test of memory was provided when the eyewitness made his or her identification. Please consider these points in judging the credibility of the eyewitness identification.

Neither this statement nor any mention of the eyewitness expert appeared in the no-reminder condition.

Dependent measures. Participants were first asked to indicate "the verdict you would cast your vote for" by checking either "guilty" or "not guilty." They were then asked to rate their certainty about their verdict on an 11-point scale (1, 4 [*extremely, moderately uncertain*]; 8, 11 [*moderately, extremely certain*]) and to write down the "most important reason" for their verdict. Then, on 11-point scales, participants rated the likelihood that the defendant committed the crime (1, 4 [*extremely, moderately unlikely*]; 8, 11 [*moderately, extremely likely*]); the strength of the evidence against the defendant (1 [*not at all*]; 6 [*somewhat*]; 11 [*extremely strong*]); and the believability of the testimony of the eyewitness, the defendant, and two witnesses who did not observe the crime (the eyewitness's wife and the defendant's neighbor, who was an alibi witness; 1 [*not at all*]; 6 [*somewhat*]; 11 [*extremely believable*]).[3]

NC Scale. Finally, participants completed the 18-item NC Scale (Cacioppo et al., 1984). The items on the NC Scale are statements (e.g., "I would prefer complex to simple problems") that respondents rate on a 5-point scale of self-descriptiveness (1 [*extremely uncharacteristic of me*];

[3] A subset of the participants in the experimental conditions were additionally asked to rate the credibility of the eyewitness expert (1 [*not at all credible*]; 11 [*extremely credible*]) and the extent to which their verdicts were influenced by the expert (1 [*not at all*]; 11 [*extremely*]). The booklet page containing these items was inadvertently omitted in several data collection sessions. Analysis of the responses to these items for those who did receive them yielded few findings of interest. In analyses in which case strength, timing, reminder, and trichotimized NC scores were independent variables, significant results were largely limited to a few higher order interactions that were not readily interpretable. In general, the expert was rated as highly credible in all conditions (overall $M = 9.60$) and moderately influential (overall $M = 6.99$).

In addition, after providing the scalar ratings relevant to guilt and the evidence, all participants were asked to list in writing, first, thoughts about the evidence that supported a guilty verdict and, second, thoughts about the evidence supporting a not-guilty verdict. Examination of the thought listings (as counted and coded by two raters whose ratings were averaged) of 100 participants (10 randomly selected from each condition) revealed few effects of interest in the present article. These are reported in a footnote to the results.

5 [*extremely characteristic of me*]). Items are scored or reverse scored so that higher scores indicate greater preference for thinking. Examples of items are as follows: "Thinking is not my idea of fun" (reverse scored); "I find satisfaction in deliberating hard and for long hours"; "I usually end up deliberating about issues even when they do not affect me personally"; and "I only think as hard as I have to" (reverse scored).[4]

Results

Manipulation Checks and Analysis Plan

A check was first made of two aspects of the experimental design and procedure. First, guilty verdicts and strength of evidence (against the defendant) ratings of participants who read transcripts depicting the stronger case were compared with those of participants who read transcripts depicting the weaker case. As anticipated, those who read the stronger (vs. weaker) case were more likely to vote guilty (56.3% vs. 33.5%), $\chi^2(1, N = 371) = 19.32$, $p < .001$, $\phi = .23$, and rated the strength of evidence against the defendant as stronger (Ms = 7.02 vs. 5.66), $t(368) = 5.53$, $p < .001$, $d = .58$.[5,6] Thus, the manipulation of the strength of the prosecution's extra-eyewitness case was successful in creating perceptually different cases: one perceived as relatively strong (rated on the higher side of the 11-point scale) and one perceived as relatively weak (on the lower side of the 11-point scale).

The second preliminary consideration was whether NC, presumably a trait (vs. state) variable, was affected by trial variations, which preceded its assessment. A 2 (case strength) × 5 (type of expert testimony: preevidence/no-reminder, preevidence/reminder, postevidence/no-reminder, postevidence/reminder, or none) analysis of variance (ANOVA) of NC scores revealed no significant effects (all ps > .12 and ηs < .15).

All of the major dependent variables were subjected initially to a 2 (case strength) × 5 (type of expert testimony) × 3 (NC: low, moderate, or high) between-subjects ANOVA. The third variable was based on a tertiary split of NC scores, which divided participants as closely as possible into three groups separated by the 33rd and 67th percentiles. NC scores lower than 59 were classified as low, and NC scores higher than 67 were classified as high. Scores in-between these values were classified as moderate. The mean NC score was 62.9, the median was 64, and scores ranged from 27 to 86.

Effects of Type of Expert Testimony and Case Strength

Perceptions of guilt. Table 1 presents the percentage of guilty votes in the 10 trial conditions, as well as condition means for guilt-certainty scores (see below) and ratings of the likelihood that the defendant committed the crime. As can be seen, the greater likelihood of a guilty verdict when the prosecution case was stronger (vs. weaker) was evident across all types of expert testimony. In addition, for both the stronger and the weaker cases, the lowest percentages of guilty verdicts and the only ones that differed appreciably from the percentages in no-expert conditions occurred in the postevidence/reminder conditions (stronger case: 41% vs. 58%; weaker case: 21% vs. 42%.) The same pattern is evident for the other two measures.

To examine more formally the effects of expert testimony, a measure of "guilt-certainty" was created by multiplying certainty-in-verdict ratings by +1 when the verdict was guilty and by −1 when it was not guilty (cf. Kassin et al., 1990). These guilt-certainty scores and ratings of likelihood of commission were very highly correlated, $r(369) = .78$, and were, therefore, combined to form a composite measure of perceived guilt by averaging the standard scores on each measure (the means of this composite measure are presented in Table 1).[7] The perceived guilt measure was then subjected to a 2 (case strength) × 5 (type of expert testimony) × 3 (NC) ANOVA, which revealed a significant main effect of case strength ($M_{\text{STRONGER}} = .35$, $M_{\text{WEAKER}} = -.37$), $F(1, 341) = 27.79$, $p < .001$, $\eta = .27$, a near-significant effect of expert testimony, $F(4, 341) = 2.26$, $p = .063$, and a significant Strength × NC interaction, $F(2, 341) = 4.58$, $p < .02$, $\eta = .26$. This interaction effect is discussed later, when attention shifts to the NC variable. Of immediate interest, first, is the utter absence of a Strength × Type-of-Expert Testimony interaction ($F < 1$, $\eta = .0$), which is consistent with Hypothesis 2, that the effects of expert testimony would not vary as a function of the strength of the prosecution case. Second, there is the overall main effect of type of expert testimony, which appears to be due, basically, to decreased perceptions of guilt in the postevidence/reminder condition relative essentially to all other types-of-expert-testimony conditions. Examination of Table 1 indicates that perceptions of guilt were quite similar to the no-expert condition in the preevidence expert conditions and the postevidence/no-reminder condition. Thus, Hy-

[4] The NC Scale is a highly reliable instrument. In seven studies reviewed by Cacioppo et al. (1996), Cronbach alphas exceeded .84. Multiple studies have found split-half and test–retest (intervals > 6 weeks) reliabilities averaging .83 and .77, respectively. The construct validity of the scale is evidenced by findings that NC scores predict level of effortful engagement on many tasks, including information recall in memory, discernment of argument quality in persuasion studies, and knowledge gained from exposure to current issues and other everyday stimuli (see Cacioppo et al., 1996, for a review of more than 100 empirical studies involving measurement of NC).

[5] Estimates of the effect size of ANOVA tests involving scaled variables were estimated by eta, where η = the square root of $[df_N(F)/df_N(F) + df_D]$. Effect size estimates for pairwise contrasts of two means on scaled variables were obtained using Cohen's (1988) d statistic, which is the ratio of the numerical difference between the means and the pooled standard deviation of scores in the two groups being compared. In the case of comparisons of two proportions (i.e., chi-square tests), effect sizes were estimated by the phi-coefficient (ϕ = square root of χ^2/n). Friedman (1968) described effect sizes of less than .30 as small effects, between .30 and .50 as moderate effects, and greater than .50 as large effects.

[6] Small differences in degrees of freedom across statistical tests of different variables reflect the fact that a few participants failed to circle a response to one or two questions.

[7] The Pearson correlations among all of the scaled dependent measures were computed for the entire sample ($N = 371$). Perceived guilt (composite score) and ratings of likelihood of commission, strength of evidence, and eyewitness believability were all highly positively correlated with each other ($.56 \le r \le .90$). These variables were moderately negatively correlated with ratings of the defendant's believability ($-.66 \le r \le -.44$), moderately positively correlated with ratings of the believability of the eyewitness's wife ($.27 \le r \le .48$), and not significantly correlated with ratings of the believability of the alibi witness ($.04 \le r \le .10$). Certainty in verdict had a small but significant correlation with perceived guilt ($r = .16$) and strength of evidence ($r = .12$), and a small negative relationship with believability of the defendant ($r = -.11$). The only other significant correlation was positive between believability of the eyewitness's wife and of the alibi witness ($r = .34$).

Table 1
Verdicts, Guilt-Certainty, Ratings of Likelihood the Defendant Committed the Crime, and Composite Perceived Guilt in the Conditions of Experiment 1

	Type of expert testimony				
	Preevidence		Postevidence		
Dependent measures	No reminder	Reminder	No reminder	Reminder	None
Percent guilty verdicts					
Stronger case	.61	.64	.56	.41	.58
Weaker case	.37	.32	.35	.21	.42
Guilt-certainty (a) [Verdict (−1 or +1) × Certainty Rating (11-point scale)]					
Stronger case	2.28	2.56	2.28	−1.02	1.26
Weaker case	−2.00	−2.38	−2.00	−4.56	−0.97
Likelihood of commission (b)					
Stronger case	7.83	7.84	7.62	7.07	7.97
Weaker case	6.77	6.38	6.32	5.71	7.00
Perceived guilt composite [average of (a) and (b) *z* scores]					
Stronger case	.49	.52	.44	−.08	.39
Weaker case	−.26	−.39	−.35	−.80	−.09
Condition *n*					
Stronger case	36	45	39	41	31
Weaker case	35	37	37	34	36

pothesis 1, that perceptions of guilt would be depressed in all expert-testimony conditions, was not borne out. However, in support of Hypothesis 3, perceptions of guilt were lowest in the postevidence/reminder condition. Two contrast tests to evaluate Hypothesis 3 revealed that, as predicted, the perception of guilt in the postevidence-testimony/reminder condition ($M = -.41$) was significantly lower than the average perceived guilt in the other three expert-testimony conditions (Ms = .12, .11, and .05), $t(366) = 2.65, p = .008, d = .36$, and also significantly lower than that in the no-expert control condition ($M = .13$), $t(366) = 2.25$, $p = .025$, $d = .39$.[8]

The eyewitness's believability. Overall, the eyewitness was somewhat believed ($M = 6.98$ on the 11-point scale). A 2 × 5 × 3 ANOVA of believability ratings of the eyewitness revealed only one significant effect, a Strength × NC interaction, $F(1, 341) = 4.90$, $p < .01$, $\eta = .17$; all other ps > .12 and ηs < .15, to be described and discussed later. All eyewitness believability was not affected overall by type-of-eyewitness testimony, a test of the postevidence/reminder versus no-expert-control contrast proved significant, $t(366) = 1.99$, $p < .05$, $d = .33$, with the eyewitness rated as less believable when there was an expert (postevidence) and a reminder ($M = 6.64$) than when there was no expert ($M = 7.40$). This is noted here because it is consistent with hypothesized effects of expert testimony (Hypotheses 1 and 3) and replicates a significant outcome in Experiment 2.

The believability of the defendant and other witnesses. It is of interest to examine whether type of expert testimony about eyewitness evidence influenced perceptions of individuals other than the eyewitness who took the stand. In fact, it did not. A 2 × 5 × 3 ANOVA of believability ratings of the defendant revealed only a main effect of case strength—the defendant was rated as more believable in the weaker (vs. stronger) case condition, Ms = 6.08 and 5.32, $F(1, 341) = 15.63$, $p < .001$, $\eta = .21$. ANOVAs revealed no significant or interpretable effects on the believability ratings of two other witnesses: the eyewitness's wife and a neighbor of the defendant who provided alibi-relevant information.[9] These witnesses, overall, were considered moderately believable (Ms = 8.43 and 8.38).

The eyewitness as the "most important reason" for the verdict. To gain further insight into how perceptions of the eyewitness were affected by expert testimony and case strength, we examined the evidence mock jurors cited as most important to their verdicts.

[8] Analyses of the thoughts that 100 mock jurors listed about evidence supporting guilt and evidence supporting innocence (see footnote 3) included a 2 (case strength) × 5 (type of expert testimony pre/no-reminder, pre/reminder, post/no-reminder, post/reminder, or none) ANOVA of the proportion of all evidence thoughts that supported a guilty (vs. a not-guilty) verdict. (A three-way ANOVA that included NC as a variable would not be interpretable owing to extremely small cell sizes.) This revealed that the proportion was higher in the strong case ($M = .68$) than in the weak case ($M = .51$) condition, $F(1, 90) = 15.15$, $p < .001$, $\eta = .38$. In addition, thoughts about the eyewitness comprised an average of 32.5% of all evidence thoughts listed. A 2 × 5 ANOVA of the proportion of evidence thoughts that involved the eyewitness revealed no main effect of either case strength or type-of-expert testimony, ps > .28, ηs < .18. The interaction was significant, $F(1, 90) = 2.67$, $p = .037$, ηs = .33, but largely uninterpretable (i.e., relative to the other nine conditions, a larger proportion of evidence thoughts concerned the eyewitness in the weak-case/prereminder condition; $M = .46$ vs. a range of .25 to .36 in the other conditions).

[9] The one significant effect on these measures was an interaction of type of testimony and NC on the ratings of the alibi witness, $F(8, 341) = 2.07$, $p = .038$, $\underline{0} = .21$. This interaction is not interpretable, as it involved small, seemingly random differences between the NC groups in the ordering of mean ratings across the five type-of-testimony conditions (e.g., low NC participants rated the wife numerically highest in the no-expert condition, whereas the numerically highest ratings were in the postevidence-expert/no-reminder and preevidence- expert/no-reminder conditions among moderate and high NC participants, respectively).

The eyewitness was listed as the most important reason or as among the most important reasons for their verdict by almost exactly half of the mock jurors (50.7%).[10] A dichotomous score was created by assigning a one to reasons that listed the eyewitness and a zero to reasons that did not. A preliminary 2 × 5 × 3 ANOVA of these dichotomous scores was conducted, which revealed significant main effects of case strength and type of expert testimony, but no other effects. The main effects were subsequently analyzed using chi-square tests. The eyewitness was mentioned as a most important reason more often when the prosecution case was weaker (57.0%) than when it was stronger (44.8%), $\chi^2(1, N = 371) = 5.51, p < .02, \phi = .16$. Across the five types-of-testimony conditions, overall $\chi^2(4, N = 371) = 9.58, p < .05, \phi = .16$, the eyewitness was mentioned most frequently in the postevidence/reminder condition (64.0%). Chi-square tests corresponding to the contrast tests associated with Hypothesis 3 revealed that the frequency in this condition was significantly higher than in the other three expert-testimony conditions combined (range: 40.9%–47.4%), $\chi^2(1, N = 304) = 8.18, p < .01, \phi = .16$, and higher, but not significantly so, than in the no-expert control condition (55.2%), $\chi^2(1, N = 142) = 1.13, p = .28, \phi = .09$. This suggests that when it followed the evidence and was highlighted by the judge, the expert testimony made the eyewitness evidence seem more salient and important.

It was also the case that verdicts were significantly related to whether the eyewitness was cited as the main reason for the verdict—but primarily in the expert-testimony conditions. In the expert-testimony conditions, the eyewitness was the most important reason for 38.8% of the guilty votes, as opposed to 59.2% of the not guilty votes, $\chi^2(1, N = 304) = 14.39, p < .001, \phi = .20$.[11] In the no-expert conditions, the percentages were 48.4% and 61.8%, respectively.

Only 6.9% of the mock jurors in the expert-testimony conditions mentioned the eyewitness expert as the most important reason for their verdicts. A preliminary 2 × 4 ANOVA on the dichotomous variable (0 = did not mention; 1 = mentioned) revealed only a marginal effect of type of testimony. Chi-square tests confirmed that the overall effect of type-of-expert testimony approached significance, $\chi^2(3, N = 304) = 6.40, p = .09, \phi = .14$. Consistent with Hypothesis 3, that the expert in the postevidence/reminder condition would have the greatest impact—this overall effect reflected a significantly higher frequency of mentioning the expert in the postevidence/reminder condition (12.0%) relative to the other three expert-testimony conditions combined (range: 1.4%–7.3%), $\chi^2(1, N = 304) = 4.02, p < .05, \phi = .12$.

Mock jurors who mentioned the eyewitness were significantly more likely to mention the eyewitness expert than were mock jurors who did not claim the eyewitness testimony as the primary reason for their verdict (9.9% vs. 3.9%), $\chi^2(1, N = 304) = 4.27, p < .04, \phi = .12$.

Effects of Strength of Prosecution Case and NC

Perceptions of guilt. As noted previously, the 2 × 5 × 3 ANOVA of perceived guilt revealed, in addition to main effects of strength and type of expert testimony, a significant Strength × NC interaction. This interaction is depicted in Figures 1 and 2 for the percentages of guilty verdicts and for the perceived guilt variable (statistical analyses both of which yielded the same results), respectively. This interaction is most clearly revealed by looking at the percentages of guilty verdicts. As can be seen in Figure 2, moderate NC, in support of Hypothesis 4, was associated with stronger discrimination of stronger and weaker cases (66% vs. 22% guilty) than was either low (57% vs. 41%) or high (44% vs. 36%) NC. With perceived guilt as the dependent variable, tests of the simple effects of case strength at each level of NC revealed that, whereas case strength did not significantly influence either low or high NC mock jurors, Fs(1, 341) = 3.50 and 2.29, ps = .07 and .15, ds = .33 and .25, respectively, the stronger (vs. weaker)

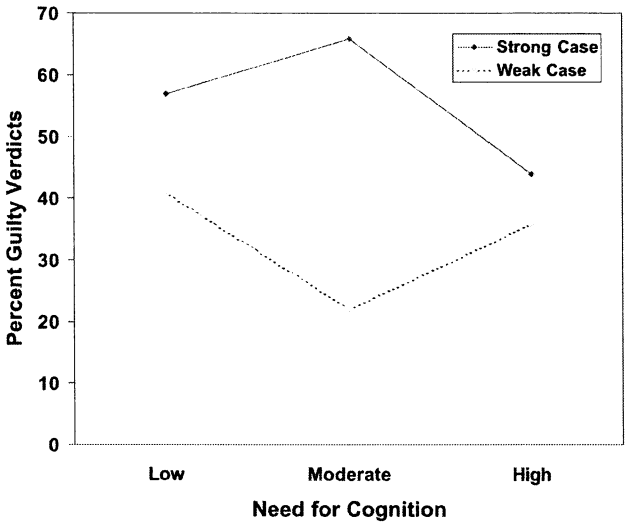

Figure 1. Percentages of mock jurors who made an individual verdict of guilty as a function of juror need for cognition (tertiary split) and strength of the prosecution's circumstantial/physical-evidence case.

[10] The questions posed to mock jurors, specifically, was "Why? [referring to the just previously checked choice of guilty or not guilty]. Please write down the MOST IMPORTANT REASON for your verdict [caps in questionnaire booklet]." Despite this instruction that was meant to ask for a single reason, a substantial number of mock jurors listed several items.

[11] When they mentioned the eyewitness, most of the mock jurors who voted not guilty noted that they doubted the eyewitness's credibility because of poor witnessing conditions or perceived inconsistencies in his testimony (e.g., "It was too dark and the witness was too far away to clearly see David Lucas" and "He didn't mention anything about tattoos. He said the murderer was wearing a sleeveless shirt which would have allowed anyone to see the tattoos"). Mock jurors who voted guilty, in contrast, most often cited the eyewitness as one of several pieces of incriminating evidence or noted that his testimony was consistent with the other evidence (e.g., "There was lots of good evidence against him; for example, the blood on his jacket was the same as Perez's and Willie's testimony" and "The eyewitness may not have had the most accurate account of what happened but it made sense—he recalled more and more as he testified that fit the evidence"). About one third of the guilty voters noted that the witnessing conditions were favorable (e.g., "Willie witnessed the crime at close distance and in good lighting") or that the eyewitness's identification was impressive (e.g., "The eyewitness account seems reliable—all his recountings of the story matched up and he did not take long to identify the man in the lineup").

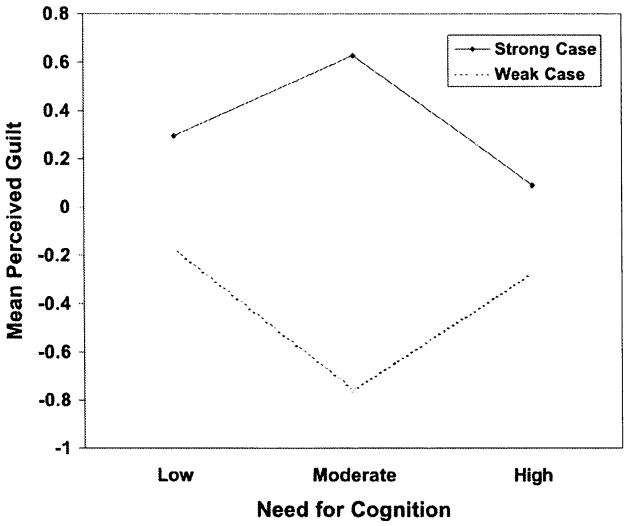

Figure 2. Mean scores on the perceived guilt index as a function of juror need for cognition (tertiary split) and strength of the prosecution's circumstantial/physical-evidence case. Higher numbers reflect greater perceived guilt.

case led to greater perceptions of guilt by moderate NC mock jurors, $F(1, 341) = 29.51$, $p < .001$, $d = .97$.[12]

As a further test of Hypothesis 4, that the perceptions of guilt of high NC mock jurors would be less sensitive to case strength, we examined the interaction of NC and strength when low NC participants were excluded and found it to be significant, $F(1, 341) = 8.54$, $p < .01$, $\eta = .18$. Dissection of this 2 × 2 interaction revealed that, as expected, compared with their moderate NC counterparts, high NC mock jurors had significantly lower perceptions of guilt when the prosecution case was relatively strong, $t(365) = 2.23$, $p < .03$, $d = .40$, and significantly higher perceptions of guilt when the case was weaker, $t(365) = 1.92$, $p = .05$, $d = .36$.

As noted previously, the 2 × 5 × 3 ANOVA revealed no Type of Expert Testimony × NC interaction ($F < 1$, $\eta = .12$), suggesting that variations in expert testimony did not differentially affect mock jurors at different levels of NC.

Perceived strength of evidence. A 2 × 5 × 3 ANOVA of ratings of the strength of the evidence against the defendant revealed a significant main effect of case strength, $F(1, 340) = 32.13$, $p < .001$, $\underline{0} = .29$, a significant Strength × NC interaction, $F(2, 340) = 5.81$, $p < .01$, $\eta = .18$, but no other reliable effects. Overall, the evidence was rated as stronger in the strong (vs. weaker) case condition. However, in a fashion that closely parallels the perceived guilt results and confirms Hypothesis 4, the stronger-versus-weaker difference in ratings was larger among moderate NC participants (Ms = 7.82 vs. 5.28) than among either low NC (Ms = 6.76 vs. 5.92) or high NC (Ms = 6.37 vs. 5.73) participants.

The eyewitness's believability. The significant Strength × NC interaction on the believability of the eyewitness noted earlier also closely parallels the same interaction on perceived guilt that was consistent with Hypothesis 4. Whereas the believability ratings of moderate NC participants were strongly affected by case strength ($M_{\text{STRONGER}} = 7.53$, $M_{\text{WEAKER}} = 6.12$), the believability ratings of both low (Ms = 7.08 and 7.02) and high (Ms = 6.75 and 7.14) NC participants did not differ as a function of case strength.

Discussion

The results of Experiment 1 suggest some interesting answers to the questions of asymmetry in evidence evaluation posed at the outset of this article. First, it does indeed appear that having individuals on a jury who have especially high need for cognition is an advantage to the defense when it faces a relatively strong case for guilt. Up to a point, increasing NC was associated with greater discernment of case strength and a corresponding increase in perceptions of guilt in the strong case condition. Beyond that point, however, the relationship reversed, and the mock jurors highest in NC evinced lower perceptions of guilt in the stronger case condition than did mock jurors with moderate NC. Presumably, their very careful consideration of the evidence led to recognition of whatever reasonable doubts there were in the case. When the case for guilt was weaker, the opposite occurred and very high (vs. moderate) NC mock jurors saw greater evidence of guilt—an advantage for the prosecution.

Second, whatever advantage the prosecution has in terms of jurors' trust in their eyewitnesses, it appears somewhat negated by the testimony of an eyewitness expert, under some circumstances. Eyewitness expert testimony served to significantly lower perceptions of guilt and believability of the eyewitness, but only when the expert's testimony came after the other trial evidence had been presented and the judge included a reminder and summary of the expert's testimony in instructions to the jury. This pattern confirms our prediction that postevidence expert testimony coupled with a judge's reminder would create a greater skepticism effect and be more exonerating than the other combinations of timing and presence or absence of a reminder. However, the absence of any effect of the other combinations represents a greater dampening of the

[12] An alternative approach to analyzing the effects of NC in conjunction with the manipulated variable of case strength would be to treat NC as a continuous variable in a multiple regression analysis. Doing so with the present data does not change the results and their interpretation. A regression analysis was conducted in which an interaction factor was created by multiplying each participant's case strength condition (dummy coded −1 and +1) and his or her specific NC score. When this factor was regressed on perceived guilt, the beta was highly significant ($\beta = .254$), $t(369) = 5.04$, $p < .001$, reflecting the interaction posited in Hypothesis 4. In addition, we dissected this interaction by conducting two additional regressions that break the interaction effect into two linear components. In both analyses, case strength, NC score, and the interaction factor were simultaneously regressed onto perceived guilt. In Analysis 1 ($N = 236$), only the low and moderate NC participants (NC score < 67) were included; in Analysis 2 ($N = 245$), only moderate and high NC participants (NC score > 58) were included. In both analyses, case strength (ts > 2.01, ps < .05) and the Strength × NC interaction (ts > 2.40, ps < .015), but not NC score (ts < 1), significantly predicted perceived guilt. NC, of course, did not predict perceived guilt overall, because the relationship, as hypothesized, depends on case strength. More important, a positive beta (1.24) was obtained for the interaction term in Analysis 1, indicating that (as is evident in on the left side of the Figure 1 presentation based on trichotomized NC scores) the difference in perceived guilt as a function of case strength increased as NC scores increased from low to moderate. A negative beta (−1.64) was observed for the interaction term in Analysis 2, reflecting the decrease in the difference as a function of case strength as NC increased from moderate to high (right side of Figure 1).

expert's influence than we had anticipated. We return to this dampening effect later. Nevertheless, the skepticism/exonerating effect did occur under the conditions that we argued were most favorable to it. Two important aspects of the exonerating impact of expert testimony under these conditions are that it (a) occurred even though the testimony was quite general and did not speak specifically to the eyewitness evidence in the case and (b) was more or less equally evident whether the prosecution case was relatively weak or relatively strong. Taken together, these aspects are somewhat disconcerting in that they suggest that general, educational testimony about eyewitness psychology may work in the defense's favor (i.e., a prodefense asymmetry), even when it perhaps should not (i.e., an otherwise strong prosecution case).

The inference that the postevidence-with-judge's-reminder expert testimony reduced perceptions of guilt in the strong case was based on the significant overall effect of this type of expert testimony collapsing across the two levels of case strength and on the fact that there was absolutely no hint of an interaction between case strength and type of expert testimony. Specific contrast tests that compared perceived guilt and percentages of guilty verdicts within the stronger-case/postevidence-expert/reminder condition against those in the stronger/no-expert condition were only marginally significant (ps = .15 and .16, respectively). This may reflect a lack of power in Experiment 1 to detect as reliable a difference of the size we obtained. Indeed, if we assume that, for the stronger case, the true size of the specific effect (i.e., strong/postevidence/reminder vs. strong/no-expert control) of having a postevidence expert coupled with a reminder corresponds to the effect we obtained (i.e., a decrease of 17 percentage points in guilty verdicts from the no-expert control baseline of 58%), the power of Experiment 1 to detect this difference as significant at $\alpha = .05$ was only 0.27. Yet, as argued later, the obtained difference in rates of voting guilty indeed may be a meaningful one. Moreover, the likelihood that the difference in perceptions of guilt is merely a chance finding seems slim, given that, as revealed by a contrast test, the postevidence-with-reminder expert's effect of decreasing the eyewitness's believability was not only significant overall (as reported earlier) but also significant within the strong-case conditions only (Ms = 7.74 vs. 6.66), $F(1, 361) = 3.94$, $p < .04$, $d = .47$. Accordingly, one goal of Experiment 2 was to replicate the conditions of this comparison, in an effort to increase power (when the replication is combined with the Experiment 1 data) and to search for firmer statistical evidence for the expert effect in a strong prosecution case.

A second observation about the stronger case condition is that, apparently, the case was not especially strong. In the no-expert control condition, only a relatively small majority (58%) of the mock jurors voted guilty. Apparently, in spite of the numerous pieces of evidence that implicated the defendant, there was plenty of reasonable doubt about his guilt. As a result, Experiment 1 may not have provided a truly good test of the important practical question of whether eyewitness expert testimony can reduce perceptions of guilt even when there is compelling noneyewitness evidence that incriminates the defendant. In Experiment 2, the stronger case was strengthened further by adding more incriminating evidence to the arsenal of the prosecution. Would the trend for the eyewitness expert to reduce perceptions of guilt be found when the prosecution case was especially strong?

Experiment 2

The second experiment was limited to two conditions: a strong-case/no-expert control condition and the strong-case version of the expert condition that produced the lowest perceptions of guilt in Experiment 1 (viz., the postevidence/reminder condition). We hypothesized that, as found in Experiment 1, perceptions of defendant guilt and of eyewitness believability would be lower in the expert condition (Hypothesis 5). In both conditions, the prosecution's case was bolstered through the addition of circumstantial and physical evidence that pointed squarely at the guilt of the defendant.

Method

Participants and Design

Sixty-eight male and female undergraduates at a large Midwestern university participated to earn credit toward a psychology course requirement. They were assigned randomly to either an expert testimony (53% men) or a no-expert testimony (59% men) condition.

Procedure and Materials

Participants in two group sessions privately read and provided impressions of the same basic trial as in Experiment 1. The two conditions studied in Experiment 2 consisted of the stronger prosecution case that included a postevidence expert witness and a judge's reminder (expert condition) and the stronger prosecution case in which there was no expert witness (no-expert condition). Additional evidence and another witness for the prosecution were incorporated into both conditions with the intent of making the prosecution's case even stronger than it was in Experiment 1. Through the questioning of witnesses and commentary in the closing arguments of both sides, it was communicated that, in addition to the other incriminating evidence, the victim's wallet was found in a trash can on the same block that the defendant lived and that DNA test results, presented by a forensic scientist, indicated a 94% certainty that the blood sample found on the defendant's jacket was that of the victim. Other than these changes, the case materials and the dependent measures (scaled measures only) were exactly the same as those used in the corresponding stronger case conditions of Experiment 1.

Results

Consistent with our intentions, the prosecution case appears to have been successfully strengthened. Overall, the mean strength-of-evidence rating was 8.02 on the 11-point scale. Among the 34 mock jurors in the no-expert condition, 74% voted guilty (compared with 58% in the corresponding condition—no expert/stronger case—of Experiment 1). In the expert condition of Experiment 2, 59% voted guilty. This reduction of guilty verdicts from 74% to 59% owing to expert testimony was directionally consistent with Hypothesis 5, but did not reach statistical significance, $\chi^2(1, N = 68) = 1.64$, $p = .20$, $\phi = .16$.

The remaining dependent variables were subjected to 2 (presence or absence of expert testimony) × 3 (NC: low, moderate, or high) ANOVAs, with levels of the latter variable determined by a tertiary split of NC scores ($M = 63.6$, $Mdn = 64$, range = 44–85) into low (NC score less than 60), moderate (NC score between 61

and 68), and high (NC score higher than 68).[13] The only significant effect obtained was on ratings of the eyewitness's believability. Consistent with Hypothesis 5 and with the results of Experiment 1, mock jurors who received expert testimony judged the eyewitness as less believable than did those who did not receive expert testimony (Ms = 6.56 vs. 7.79), $F(1, 61) = 6.92$, $p < .02$, $d = .63$.

Relevant to the relationship between NC and the tendency to find reasonable doubt in a relatively strong case, an interesting pattern was evident in guilty votes in the no-expert condition. Even though the prosecution case was quite strong, high NC mock jurors were significantly less likely to vote guilty (50%) than were mock jurors who were either moderate or low in NC (83%), $\chi^2(1, N = 34) = 4.03$, $p < .05$, $\phi = .34$. Thus, as in Experiment 1 (i.e., Hypothesis 4), those mock jurors with the highest NC scores were the least swayed toward guilt by a relatively strong case for guilt.

Combined Results

It is noteworthy that a difference in verdicts between the no-expert and postevidence-expert/reminder conditions of close to the same size in percentage points was found for both a moderately (Experiment 1) and more than moderately (Experiment 2) strong prosecution case. The difference fell short of conventional significance levels in both experiments taken singly, but given the consistency of the difference across the two replications, this would seem to reflect a lack of statistical power and not a case of a merely chance outcome. Not only were the differences in the two experiments remarkably similar, but they co-occurred with effects of expert testimony on eyewitness believability that were significant in both experiments. Moreover, the observed differences in guilty verdicts are not trivial ones. The presence of postevidence eyewitness expert testimony (coupled with a judge's reminder) reduced the rate of guilty verdicts from nearly three in five (58%) to almost as low as two in five (41%) in Experiment 1 and by 15 percentage points (74%–59%) in Experiment 2. If we combine the two studies, two of every three jurors (66%) would have entered a deliberation room preferring to convict if there was no eyewitness expert, whereas one of every two (49%) would have preferred conviction after receiving the testimony of an eyewitness expert (and a judge's reminder) in a case that was otherwise identical. A test comparing these percentages for the combined samples ($n = 72$ from Experiment 1 and $n = 68$ from Experiment 2), which has substantially more power than tests of only the single samples, revealed that this difference was, in fact, statistically significant, $\chi^2(1, N = 140) = 4.02$, $p < .05$, $\phi = .17$. The difference was significant in another way as well. Research suggests that juries entering deliberation evenly split (e.g., six leaning toward guilty, six leaning toward acquittal) are considerably more likely than not to acquit, whereas two-thirds splits (e.g., eight–four) in favor of guilty are more likely to convict than to acquit (MacCoun & Kerr, 1988; Stasser, Kerr, & Bray, 1982).

General Discussion

It appears from our results that expert testimony that provides only a general overview of theory and research on eyewitness memory can be sufficient under some circumstances to decrease both belief in the eyewitness and the likelihood of a guilty verdict. This occurred when the prosecution case aside from the eyewitness testimony was relatively weak and also when the prosecution case was considerably stronger. The exonerating influence of the expert witness, however, did not occur anytime the expert took the stand. In the trial used in the present research, the expert's influence was limited to the trial condition in which the expert gave testimony after all other evidence had been presented and the judge reminded jurors about the expert testimony by summarizing its points during final instructions. In the absence of either of these features, the expert testimony had little or no impact on guilty votes or any other juror impressions. Overall, then, it appears that expert testimony may have an impact under a limited set of conditions, and when it does have an impact, that impact serves to reduce the advantage to the prosecution of having eyewitness evidence.

On a very different dimension, we found support for our hypothesis that highly thoughtful jurors, as evidenced by very high NC scores, could represent an advantage for the defense when it has the weaker case because their careful scrutiny of evidence would likely find reasonable doubt in even relatively strong prosecution cases. By breaking the NC variable into three categories—low, moderate, and high—rather than the more typical median-split breakdown that divides respondents into lower and higher groups, Experiment 1 uncovered a curvilinear relationship between NC and perceptions of guilt. Consistent with the usual expectation regarding NC, moderately higher NC mock jurors perceived more guilt in a strong prosecution case than did those with low NC scores. However, NC mock jurors who scored in the highest third of NC scores perceived less guilt than did moderate NC mock jurors when the prosecution case was relatively strong, judging the overall weight of the evidence against the defendant to be weaker. A similar finding was obtained in the no-expert-testimony condition of Experiment 2, in which the prosecution case was even stronger. The highest NC mock jurors were the least likely to vote guilty.

We also found the opposite of this leniency effect among the highest NC mock jurors. These very thoughtful individuals proved helpful to the prosecution when its case was the somewhat weaker one, in that they found more evidence for guilt than did moderate NC mock jurors in whatever the prosecution did have in the weaker case condition.

Before discussing these results, we should note that there may be limits to the generality of our findings associated with the fact that ours were simulation studies that used a written (vs. live or videotaped) presentation of the trial, only excerpts of the trial (albeit lengthy, realistic ones), and only college students as mock jurors. Moreover, no serious consequences (e.g., the imprisonment of the defendant) were expected to be associated with mock jurors' decisions, and jurors did not deliberate. With these caveats in mind, we should also note that the material that mock jurors did receive, including opening arguments, the question-and-answer exchanges involved in direct and cross-examinations, closing arguments, judge's instructions, and, of course, the expert testimony, quite closely resembled—at least in verbal content—what they might have heard in an actual trial. In addition, as Cooper and Hall (2000) have observed, research suggests that college-student and community/jury-eligible samples respond quite similarly to simulated trials involving eyewitness testimony (cf. Cutler, Dexter, &

[13] As in Experiment 1, NC scores did not vary by experimental condition, $t(65) = 0.07$.

Penrod, 1990; Nunez, McCoy, Clark, & Shaw, 1999) and that predeliberation verdict choices of individual mock jurors are relatively strong predictors of deliberated final verdicts (Bray & Kerr, 1979; Hastie, 1993a). Moreover, in a recent comprehensive review of jury-simulation studies that compared different samples of mock jurors (e.g., college students vs. jury-eligible community citizens) and manipulated the medium of trial presentation (e.g., live, video, audio, written transcript, or written summary), Bornstein (1999) concluded that there is "little research that has obtained differences between different mock juror samples or different media" (p. 75). Thus, there are reasons to believe that our results have a reasonable level of practical generality, in addition to offering conceptual insight and directions for further research.

Expert Testimony: Type, Timing, and a Two-Step Process of Impact on Jurors

The results regarding eyewitness expert testimony, in part, are consistent with the typical finding of past research; namely, that such testimony may result in decreased belief in the guilt of the defendant. In past studies that have found this exonerating effect, the expert testimony has comprised a prominent and salient portion of the trial or witness information received by the triers-of-fact and has almost always followed the eyewitness testimony. The current results make especially good sense in light of these aspects of past research, given that the exonerating effects of the expert testimony occurred only when the expert testimony occurred late in the trial (after the eyewitness) and the judge offered a reminder and review of the expert's statements in final instructions—in other words, when the expert testimony was prominent and salient to mock jurors as they considered a verdict. When the expert testimony is less salient during verdict consideration, its effects may be muted.

A novel feature of Experiment 1 was that it directly compared the effects of pre- and postevidence expert testimony. As far as we know, it was the first whole-trial jury simulation to present the eyewitness expert in advance of the evidence and the eyewitness. Accordingly, the failure of the preevidence expert to affect mock jurors' perceptions and verdict is especially noteworthy as a new finding. As discussed later, however, it would be premature to conclude that preevidence expert testimony is unlikely to ever have an impact on jurors.

Interestingly, it was not only the preevidence expert testimony that produced no skepticism effect, but postevidence expert testimony that was not followed by a judge's reminder was also inconsequential. Although it was predicted that there would be less of a skepticism effect in this no-reminder condition than in the postevidence/reminder condition, we did not expect there to be no decrease in guilt perceptions when a reminder was absent. On the face of it, this null result is inconsistent with other studies, which generally have observed a skepticism effect when presenting expert testimony after the eyewitness evidence. The reason for the discrepancy, however, may be found in the salience dimension that we have highlighted. In past studies, the eyewitness expert has invariably been quite prominently featured to mock jurors. This was true even in the videotaped trial used by Cutler and his colleagues (e.g., Cutler et al., 1989), in which the eyewitness expert gave elaborate testimony and was subjected to considerable cross-examination in which he noted limitations of research and conclusions. In contrast, the expert in the present studies was presented and questioned without much fanfare and specificity.

Salience, leading to heightened availability during decision making, is central to our conceptualization of how and when eyewitness expert testimony may impact jurors. In general, the results support a recency interpretation of how eyewitness expert testimony is likely to have an impact on jurors. When the time comes to translate what they have learned at trial into a decision regarding guilt, jurors are more likely to spontaneously recall the expert testimony if it—and a reminder about it—came late (vs. early) in the trial and therefore closer in time to their deliberation. In turn, to the extent that the expert testimony signified a cause for skepticism about the eyewitness evidence, that evidence should be at least partially discounted or given less weight in jurors' decision-making calculus. In the present case, the judge's reminder just before deliberation served, in part, to increase the likelihood of activation of the expert testimony to a threshold point at which it was applied to the case. This repetition, in essence, of expert testimony in close temporal proximity—the expert's on-the-stand minilecture and the judge's reminder/summary of it—may have been necessary for sufficient likelihood of activation of memory for the expert testimony and its implications. Consistent with this idea, in a study of the effects of expert testimony about child sexual abuse, Kovera, Gresham, Borgida, Gray, and Regan (1997) found that guilty verdicts were only influenced in the anticipated direction (increased convictions) in a "repetitive testimony" condition in which the expert provided a summary of research findings and then reiterated this summary.

Alternatively, or in addition to an effect on accessibility through repetition, the testimony-plus-reminder sequence in the present studies may have worked in two steps. The recency of the expert testimony may have made more available to jurors what to apply in judging retrospectively the eyewitness evidence, and the judge's reminder may have provided an authorization to apply it. This two-step notion is especially plausible given the fact that neither the postevidence presentation of the expert testimony nor the judge's reminder was sufficient by itself to lead to a reduction in perceptions of guilt or believability of the eyewitness. That is, expert testimony had no exculpatory effects in either the postevidence-expert/no-reminder or the preevidence-expert/reminder conditions. Salience/availability and authorization, therefore, both seem to be required. The combined impact of these two factors may reflect attempts to "correct for bias" on the part of mock jurors (Wegener & Petty, 1997). On hearing the expert testimony, mock jurors may have become worried that they were overly swayed by the eyewitness, and this worry may have been acted on by a downward adjustment in the credibility accorded the eyewitness testimony only when the judge's instructions gave authoritative credence to that worry about a pro-eyewitness bias. In this regard, it is important to note that specific memory for the eyewitness's testimony may have faded by this time to the point where it could be now remembered as weak along a number of dimensions the expert discussed. In contrast, the eyewitness testimony may have held up better on those dimensions in the preevidence conditions when it could be evaluated online after hearing the expert testimony.

The role of salience is also evident in the responses mock jurors gave when questioned about the most important reason for their verdict. The eyewitness was significantly more likely to be mentioned in the postevidence-expert/reminder condition than in the other conditions, and, typically, the mention of the eyewitness was

accompanied by a critique of the eyewitness testimony that resembled factors the expert mentioned (see footnote 11 for examples).

As noted earlier, our results do not rule out entirely the possibility that preevidence eyewitness expert testimony will influence trial verdicts. It is possible, for example, that receiving expert testimony beforehand is likely to affect judgments of eyewitness testimony primarily when the eyewitness testifies live and provides nonverbal expressions of confidence and other demeanor that might be discounted based on what the expert said. Preevidence expert testimony might also be effective when it is made explicitly relevant to the facts of the case at trial or to the on-stand demeanor or statements of the eyewitness. Indeed, in a study of expert testimony about the battered woman syndrome, preevidence testimony that focused on the emotional reactions of victims and situational factors that make escape perceived as impossible was found to have a stronger effect on verdicts than postevidence testimony (Schuller & Cripps, 1998). Finally, there was much other evidence and numerous witnesses in the trial used in the current research. Perhaps if the eyewitness evidence had been more "front and center"—as the main and pivotal evidence in the trial—preevidence expert testimony would have aroused more skepticism about the eyewitness.

Of course, the very fact that the eyewitness evidence existed amid a plethora of other evidence makes it somewhat remarkable that any eyewitness expert testimony significantly reduced judgments of guilt. Additionally, the fact that it was general-knowledge testimony and that the expert was presented as a friend of the court (i.e., not as part of an adversarial strategy by the defense) make the exonerating effect in the postevidence/reminder condition all the more remarkable. At this point, the conclusion must be that general, non-case-specific eyewitness expert testimony, if it is salient and endorsed by the judge, somehow cues jurors to be more skeptical of eyewitnesses, the prosecution case in general, or both. An important focus of further research should be to build on these findings by discerning what about the expert testimony communicates this message and how it enters the decision-making calculus of jurors. For example, does it alter the incriminating value of the eyewitness evidence or the weight attached to it? Alternatively, or in addition, does it "raise the bar" for the evidence necessary to convict?

It bears mentioning that the court-appointed aspect of the expert testimony in the present studies is an uncommon quality of eyewitness expert testimony in actual trials. For a number of reasons, including the adversarial nature of North American justice systems, the preferences of litigating attorneys given an adversarial system, and the difficult logistics of governing ex parte communications with court-appointed experts, eyewitness experts are usually retained and called by one side (usually the defense) of the case. Despite its rarity, inclusion of a court-appointed eyewitness expert is permissible under the law, and a number of psycho-legal researchers have studied and recommended such a nonadversarial practice (see, e.g., Cooper & Hall, 2000; Cutler et al., 1990; Leippe, 1995; Lindsay, MacDonald, & McGarry, 1990). The relevant issue here is the extent to which the present findings using court-appointed testimony generalize to the situation in which the expert is called by the defense and placed in, at least, a perceived adversarial role. In reviewing the research on the effects of eyewitness expert testimony, Leippe noted that "across studies, both types of expert created similar levels of skepticism" (p. 949), an observation that bodes well for ready generalization. On the one hand, Cutler et al. found that mock jurors made less use of court-appointed expert testimony in that, unlike mock jurors who heard an adversarial expert, their verdicts were insensitive to variations in the quality of witnessing and testing conditions that the expert mentioned as important. Cutler et al. conjectured that mock jurors were less likely to systematically process and learn from the court-appointed expert because, owing to their knowledge that the judge had appointed the expert, they accepted the message without scrutiny. Presumably, through their superficial processing and assumption of the expert's credibility, mock jurors registered only a "take-home" message that calling such an expert means there is reason to be skeptical of the eyewitness evidence. If this were in fact the case, it suggests that the skepticism effect may be greater for court-appointed compared with adversarially presented eyewitness experts. A countervailing force on skepticism, however, is that an adversarial expert called by the defense is probably more likely to slant his or her testimony toward matters that suggest problems with the specific eyewitness evidence in the case at hand, thereby increasing skepticism. Thus, it could be that the present results overestimate, underestimate, or appropriately estimate (e.g., if the countervailing forces mentioned cancel each other out) the impact on jurors' skepticism an adversarial (vs. a court-appointed) expert might have when the expert follows the eyewitness evidence (as must necessarily be the case for an adversarial expert). This is an empirical question that future research could pursue.

In terms of practical implications, perhaps the most important finding is that general-knowledge eyewitness expert testimony served, under some conditions, to reduce perceptions of guilt even when the prosecution's case was rather strong and included considerable evidence in addition to the positive eyewitness identification of the defendant. This result gives some credence to the concern that eyewitness expert testimony can have the unwanted effect of creating sufficient juror skepticism about the entire case that leads to an acquittal in the face of strong circumstantial evidence. To be sure, we have yet to test for this impact of expert testimony in an extremely strong prosecution case, and it can be argued that an increase in overall skepticism in a moderately strong case is not so inappropriate. Yet, if our results do reflect a tendency that extends to strong extra-eyewitness cases in general, it puts the onus on judges to decide whether to admit eyewitness expert testimony in light of how much other incriminating evidence besides the eyewitness(es) that the prosecution is prepared to offer. Such decisions are often not easy given the subjectivity of evidence and the fact that all evidence is not fully "in play" until the trial is underway. Nevertheless, it would seem appropriate to add case strength to the factors that need to be factored into the admission of an eyewitness expert into the courtroom (cf. Leippe, 1995). At the same time, the significant impact of expert testimony in rather strong cases may also have implications for whether a general overview of scientific findings about eyewitness memory should become a standard feature of trials that involve eyewitnesses. If such a practice is ever to be implemented, there first needs to be more research into how useful information about evaluating eyewitness evidence can be presented without sending an implicit message to distrust the actual eyewitness(es) and with effective instructions that make clear that other evidence may exist that is independent of the eyewitness evidence. It is not yet known if these objectives can be accomplished.

NC and Reasonable Doubt

The moderating role of jurors' NC in the present research is interesting for both theoretical and practical reasons. At the conceptual level, the curvilinear relationship between jurors' NC and guilt judgments in Experiment 1 suggests that NC may interact with processing goals in producing judgments that require thoughtfulness. Thus, if the goal is to make a decision about what is the best attitude to have based on the available evidence, increasing thinking about a strong persuasive message with largely unassailable arguments should lead to linear increases in the persuasive impact of that message as a function of increasing NC. This is the typical result of persuasion studies in which recipients' NC has been examined (cf. Cacioppo et al., 1996), although these studies have been limited in obtaining evidence for a linear-only effect because recipients have been divided into only two groups (higher and lower NC). However, if, for example, recipients are asked to adopt an "editorial" or "critical" set to be sensitive to any flaws in the arguments or to try to refute the arguments, the NC/persuasion relationship might well turn out to be curvilinear, with the high NC recipients evincing less attitude change than moderate NC recipients in response to a strong but imperfect message. NC might correctly reflect enjoyment of thinking and thus predict cognitive effort in thinking-dependent tasks, but this may not always imply a greater likelihood of making an attitudinal judgment that is most validly sensitive to the data at hand.

At the practical level of the jury, the current results suggest that selecting for those individuals who display the highest levels of enjoyment of thinking—those who seem highest in NC—would generally be a good strategy for the defense counsel facing a superior prosecution case. High NC individuals, it seems from our data, are more likely to find reasonable doubt in a pretty strong case for guilt, owing to the greater cognitive effort they put into studying the evidence. However, achieving this defense advantage through jury selection carries some danger. First, it may not be easy to discern high NC individuals from moderately high ones—and the latter were most swayed toward guilty verdicts by a stronger case in the present research. In effect, when the prosecution case is at least moderately strong, moderately high NC jurors may be advantageous to the prosecution and very high NC jurors advantageous to the defense. Second, just as they may be more likely to uncover reasonable doubt, extremely high NC jurors may also be more likely to see through irrelevant or strained defense arguments. Third, if the defense is mistaken about the strength of the prosecution's case, and it is actually pretty weak, the highest NC jurors may see it as a stronger case than will moderate NC jurors. Finally, as implied earlier, the reactions of high NC jurors may depend on the decision rule they are working with. In a criminal case that uses a reasonable doubt standard, high (vs. lower) NC jurors may be less swayed by a relatively strong case. However, in a civil case that uses a preponderance of evidence standard, they may be more swayed by a relatively strong case because they are more likely to discern the strength of the case and can discount any of the greater doubts they are likely to notice, so long as such evidence does not become preponderant.

Our interpretation of the NC results centers on evidence scrutiny and the discovery and construal of exculpatory evidence as reasonable doubt. It is possible that, instead, or in addition, the NC results reflect a relationship between jurors' NC and definitions of reasonable doubt. Possibly, extremely high NC individuals require a great deal of certainty that the defendant is guilty before they will convict. This account, however, would have difficulty explaining why high NC mock jurors acquitted less (vs. more) often than moderate NC jurors when the case for guilt was weak. Still, this rival hypothesis might be addressed in future research that assesses directly how individuals who differ in NC subjectively define reasonable doubt (cf. Hastie, 1993b).

Conclusions

We began this article by noting two apparent asymmetries in the psychology of evidence evaluation. Eyewitness testimony tends to impress jurors in a fashion that provides an advantage to the prosecution. The reasonable doubt standard may lead jurors who are internally motivated to think about the evidence (high NC) to be biased toward acquittals in relatively strong cases for guilt, and thus provide some advantage for the defense. Our results confirm the second of these biases and offer no remedy other than selecting for less thoughtful jurors or trying to build an iron-clad case for guilt. Regarding the first bias, it appears that postevidence eyewitness expert testimony endorsed by the judge may lessen the prosecution-favorable impact of eyewitness testimony. Interestingly, having a pretty strong case in addition to the eyewitness does not seem to help the prosecution protect the high credibility of their eyewitnesses or their case from eyewitness expert testimony of this sort.

References

Bornstein, B. H. (1999). The ecological validity of jury simulations: Is the jury still out? *Law and Human Behavior, 23,* 75–92.

Bray, R. M., & Kerr, N. L. (1979). Use of the simulation method in the study of jury behavior: Some methodological considerations. *Law and Human Behavior, 3,* 107–119.

Brigham, J. C., & Bothwell, R. K. (1983). The ability of prospective jurors to estimate the accuracy of eyewitness identifications. *Law and Human Behavior, 7,* 19–30.

Brigham, J. C., Wasserman, A. W., & Meissner, C. A. (1999). Disputed eyewitness identification evidence: Important legal and scientific issues. *Court Review, 36,* 12–25.

Buckhout, R. (1974, December). Eyewitness testimony. *Scientific American, 231*(6), 23–31.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42,* 116–131.

Cacioppo, J. T., Petty, R. E., Feinstein, J., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119,* 197–253.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48,* 306–307.

Cacioppo, J. T., Petty, R. E., & Morris, K. J. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology, 45,* 805–818.

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

Cooper, J., Bennett, E. A., & Sukel, H. L. (1996). Complex scientific testimony: How do jurors make decisions. *Law and Human Behavior, 20,* 379–394.

Cooper, J., & Hall, J. (2000). Reactions of mock-jurors to testimony of a court-appointed expert. *Behavioral Sciences and the Law, 18,* 719–729.

Cutler, B. L., Dexter, H. R., & Penrod, S. D. (1990). Nonadversarial methods for sensitizing jurors to eyewitness evidence. *Journal of Applied Social Psychology, 20,* 1197–1207.

Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1989). The eyewitness, the expert psychologist, and the jury. *Law and Human Behavior, 13,* 311–332.

Ellsworth, P. C. (1993). Some steps between attitudes and verdicts. In R. Hastie (Ed.), *Inside the juror: The psychology of juror decision making* (pp. 42–64). New York: Cambridge University Press.

Fox, S. G., & Walters, H. A. (1986). The impact of general vs. specific expert testimony and eyewitness confidence upon mock juror judgment. *Law and Human Behavior, 10,* 215–228.

Friedman, H. (1968). Magnitude of experimental effect and a table for its rapid estimation. *Psychological Bulletin, 70,* 245–251.

Goldstein, A. G., Chance, J. E., & Schneller, G. R. (1989). Frequency of eyewitness identification in criminal cases: A survey of prosecutors. *Bulletin of the Psychonomic Society, 27,* 71–74.

Greene, E. (1988). Judge's instructions on eyewitness testimony: Evaluation and revision. *Journal of Applied Social Psychology, 18,* 252–276.

Haber, R. N., & Haber, L. (2001, November). *A meta-analysis of research on eyewitness lineup identification accuracy.* Paper presented at the annual meeting of the Psychonomics Society, Orlando, FL.

Hastie, R. (1993a). *Inside the juror: The psychology of juror decision making.* New York: Cambridge University Press.

Hastie, R. (1993b). Algebraic models of juror decision processes. In R. Hastie (Ed.), *Inside the juror: The psychology of juror decision making* (pp. 84–115). New York: Cambridge University Press.

Horowitz, I. A., Bordens, K. S., Victor, E., Bourgeois, M. J., & ForsterLee, M. J. (2001). The effects of complexity on jurors' verdicts and construction of evidence. *Journal of Applied Psychology, 86,* 641–652.

Kaplan, M. F., & Kemmerick, G. D. (1974). Juror judgment as information integration: Combining evidential and nonevidential information. *Journal of Personality and Social Psychology, 30,* 493–499.

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.

Kassin, S. M., Tubb, V. Z., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research: A new survey of the experts. *American Psychologist, 56,* 405–416.

Kassin, S. M., Reddy, M. E., & Tulloch, W. F. (1990). Juror interpretations of ambiguous evidence: The need for cognition, presentation order, and persuasion. *Law and Human Behavior, 14,* 25–42.

Kerr, N. L., & Bray, R. M. (Eds.). (1982). *The psychology of the courtroom.* New York: Academic Press.

Kovera, M. B., Gresham, A. W., Borgida, E., Gray, E., & Regan, P. C. (1997). Does expert psychological testimony inform or influence juror decision making? A social cognitive analysis. *Journal of Applied Psychology, 82,* 178–191.

Leippe, M. R. (1980). Effects of integrative memorial and cognitive processes on the correspondence of eyewitness accuracy and confidence. *Law and Human Behavior, 4,* 261–274.

Leippe, M. R. (1994). The appraisal of eyewitness testimony. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 385–417). New York: Cambridge University Press.

Leippe, M. R. (1995). The case for expert testimony about eyewitness memory. *Psychology, Public Policy, and Law, 1,* 909–959.

Leippe, M. R., Seib, H. M., Eisenstadt, D., & Alia-Klein, N. (1998, August). *Effects of eyewitness expert testimony in strong and weak cases.* Paper presented at the 106th Annual Convention of the American Psychological Association, San Francisco, CA.

Leippe, M. R., Wells, G. L., & Ostrom, T. M. (1978). Crime seriousness as a determinant of accuracy in eyewitness identification. *Journal of Applied Psychology, 63,* 345–351.

Lindsay, R. C. L. (1994). Expectations of eyewitness performance: Jurors' verdicts do not follow from their beliefs. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness memory: Current trends and developments* (pp. 362–384). New York: Cambridge University Press.

Lindsay, R. C. L., MacDonald, P., & McGarry, S. (1990). Perspectives on the role of the expert witness. *Behavioral Sciences and the Law, 8,* 457–464.

Loftus, E. F. (1974, December). The incredible eyewitness. *Psychology Today, 8*(7), 116–119.

Loftus, E. F. (1980). Impact of expert psychological testimony on the unreliability of eyewitness identification. *Journal of Applied Psychology, 65,* 9–15.

Loftus, E. F. (1983). Silence is not golden. *American Psychologist, 38,* 564–572.

Maass, A., Brigham, J. C., & West, S. G. (1985). Testifying on eyewitness reliability: Expert advice is not always persuasive. *Journal of Applied Social Psychology, 15,* 207–229.

MacCoun, R. J., & Kerr, N. L. (1988). Asymmetric influences in mock jury deliberations: Jurors' bias toward leniency. *Journal of Personality and Social Psychology, 54,* 21–33.

Moore, P. J., & Gump, B. B. (1995). Information integration in juror decision making. *Journal of Applied Social Psychology, 25,* 2158–2179.

Neil v. Biggers, 409 U.S. 188 (1972).

Nunez, N., McCoy, M. L., Clark, H. L., & Shaw, L. A. (1999). The testimony of elderly victim/witnesses and their impact on juror decisions: The importance of examining multiple stereotypes. *Law and human Behavior, 23,* 413–423.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology, 62,* 189–206.

Petty, R. E., Tormala, Z. L., Hawkins, C., & Wegener, D. T. (2001). Motivation to think and order effects in persuasion: The moderating role of chunking. *Personality and Social Psychology Bulletin, 27,* 332–344.

Saunders, D. M., Vidmar, N., & Hewitt, E. C. (1983). Eyewitness testimony and the discrediting effect. In S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating witness evidence: Recent psychological research and new perspectives* (pp. 57–78). Chichester, England: Wiley.

Scheck, B., Neufeld, P., & Dwyer, J. (2000). *Actual innocence: Five days to execution and other dispatches from the wrongly convicted.* New York: Cambridge University Press.

Schuller, R. A., & Cripps, J. (1998). Expert evidence pertaining to battered women: The impact of gender of expert and timing of testimony. *Law and Human Behavior, 22,* 17–31.

Schum, D. A. (1993). Argument structure and evidence evaluation. In R. Hastie (Ed.), *Inside the juror: The psychology of juror decision making* (pp. 175–191). New York: Cambridge University Press.

Smith, S. M., & Petty, R. E. (1996). Message framing and persuasion: A message processing analysis. *Personality and Social Psychology Bulletin, 22,* 257-268.

Snyder, M., & Uranowitz, S. W. (1978). Reconstructing the past: Some cognitive consequences of person perception. *Journal of Personality and Social Psychology, 36,* 941–950.

Stasser, G., Kerr, N. L., & Bray, R. M. (1982). The social psychology of jury deliberations: Structure, process, and product. In N. L. Kerr & R. M. Bray (Eds.), *The psychology of the courtroom* (pp. 221–256). New York: Academic Press.

United States v. Telfaire, 469 F.2d 552 (D. C. Cir. 1972).

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naïve theories of bias in bias correction. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp.141–208). San Diego, CA: Academic Press.

Wells, G. L. (1984). How adequate is human intuition for judging eyewitness testimony. In G. L. Wells & E. L. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 256–272). New York: Cambridge University Press.

Wells, G. L. (1986). Expert psychological testimony: Empirical and conceptual analysis of effects. *Law and Human Behavior, 10,* 83–95.

Wells, G. L., Lindsay, R. C. L., & Tousignant, J. P. (1980). Effects of expert psychological advice on human performance in judging the validity of eyewitness testimony. *Law and Human Behavior, 4,* 275–285.

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,* 581–598.

Wells, G. L., & Murray, D. M. (1983). What can psychology say about the Neil v. Biggers criteria for judging eyewitness accuracy? *Journal of Applied Psychology, 68,* 347–362.

Wells, G. L., & Murray, D. M. (1984). Eyewitness confidence. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 155–170). New York: Cambridge University Press.

Wells, G. L., & Wright, E. F. (1983). Unpublished raw data.

Whitley, B. E., Jr., & Greenberg, M. S. (1986). The role of eyewitness confidence in juror perceptions of credibility. *Journal of Applied Social Psychology, 16,* 387–409.

Woocher, F. D. (1977). Did your eyes deceive you? Expert psychological testimony on the unreliability of eyewitness identification. *Stanford Law Review, 29,* 969–1030.

Received July 15, 2002
Revision received June 16, 2003
Accepted June 20, 2003 ■

BEYOND THE KEN? TESTING JURORS' UNDERSTANDING OF EYEWITNESS RELIABILITY EVIDENCE

Richard S. Schmechel, Timothy P. O'Toole, Catharine Easterly, and Elizabeth F. Loftus*

ABSTRACT: Over the past thirty years, researchers have made substantial strides in understanding the workings and limitations of human memory. However, the application of these scientific advances to eyewitness identifications in the criminal justice system, though increasing, has been limited. Trial judges in most jurisdictions exercise their discretionary powers to exclude expert testimony about the reliability of eyewitness identifications. The most common rationale for excluding eyewitness identification expert witnesses is that their findings are not "beyond the ken" of the average juror.

To empirically test this "beyond the ken" rationale, an independent survey of potential jurors in the District of Columbia was designed to investigate whether jurors understand, as a matter of common sense, what makes some eyewitness identifications more or less reliable than others. The survey results, presented in this article, demonstrate that jurors misunderstand how memory generally works and how particular factors, such as the effects of stress or the use of a weapon, affect the accuracy of eyewitness testimony. In light of these findings, judicial practices of excluding expert testimony on the reliability of eyewitness identifications should be reexamined. Wrongful convictions, of which eyewitness identification error is the leading cause, will inevitably continue to result unless jurors can be better educated about these scientific findings.

CITATION: Richard S. Schmechel, Timothy P. O'Toole, Catharine Easterly, and Elizabeth F. Loftus, Beyond the Ken? Testing Jurors' Understanding of Eyewitness Reliability Evidence, 46 Jurimetrics J. 177–214 (2006).

---

*Richard S. Schmechel is Soros Advocacy Fellow with the Public Defender Service for the District of Columbia. Timothy P. O'Toole is Chief, Special Litigation Division with the Public Defender Service for the District of Columbia. Catharine Easterly is Staff Attorney, Special Litigation Division with the Public Defender Service for the District of Columbia. Elizabeth F. Loftus is Distinguished Professor at the University of California, Irvine.

HeinOnline -- 46 Jurimetrics 177 2005-2006

Over 75,000 people a year become criminal defendants on the basis of eyewitness identifications.[1] Court officers, law enforcement, and jurors are repeatedly called upon to assess the accuracy of witnesses' claims that they remember having seen the accused committing a criminal offense or other significant details about a crime. Yet, without prior training in assessing eyewitness reliability—or, in the case of jurors, the chance to learn how to make this assessment in the courtroom—these key actors in the criminal justice system have only their common sense intuitions as guides when facing a host of difficult questions about witnesses' claims. These questions include the following: Is an eyewitness more reliable if the witness expresses absolute confidence in her identification in trial testimony? Do people remember faces better or worse when they are under stress? When do people's memories begin to fade? Does it matter if a witness is of a different race from the perpetrator? What is the significance of an eyewitness learning new information about the alleged culprit after the event? Is one kind of police identification procedure better than another?

In the late 1970s pioneering psychologists created a new field of social science research dedicated to the study of such questions about eyewitness reliability. They and their successors built upon and extended core research regarding memory and visual perception that dates from the foundation of psychology as a science in the early twentieth century. After thirty years of scientific investigation, the field has found empirical answers to the above questions and many others.

Such discoveries frequently are relevant to criminal trials, for example:

- Accuracy-Confidence Correlation: an eyewitness' stated confidence is not a good predictor of identification accuracy;
- Stress Effects: highly stressful situations may make an experience seem especially vivid, but such stressors can reduce the ability to recall details about a person's face;
- Time Estimates: eyewitnesses typically overestimate how long an event took to unfold;
- Cross-race Bias: eyewitnesses are more accurate at identifying members of their own race than members of other races;
- Postevent Information: eyewitness testimony about an event often reflects not only what a person actually saw, but also information learned later on that unconsciously becomes part of the memory;
- Presentation Format: witnesses are more likely to misidentify someone when they view all the suspects simultaneously in a group rather than one at a time, in sequence.[2]

---

1. Press Release, Nat'l Sci. Found., False Identification: New Research Seeks to Inoculate Eyewitnesses Against Errors (Jan. 3, 1997), available at http://www.nsf.gov/pubs/stis1997/pr971/pr971.txt.

2. Several excellent resources are available that introduce this research to nonexperts in more detail. *See* BRIAN L. CUTLER, EYEWITNESS TESTIMONY: CHALLENGING YOUR OPPONENT'S WITNESS (2002); BRIAN L. CUTLER & STEVEN D. PENROD, MISTAKEN IDENTIFICATION: THE EYEWITNESS, PSYCHOLOGY AND THE LAW (1995); ELIZABETH F. LOFTUS, EYEWITNESS TESTIMONY (1996). The American Psychology-Law Society has assembled a comprehensive research bibliography at http://www.ap-ls.org/links/publishingEyewitness.html.

HeinOnline -- 46 Jurimetrics 178 2005-2006

Eyewitness reliability research uses methods accepted in all sciences. Researchers form hypotheses based on prevailing theories of human memory and cognition, and these hypotheses are tested—sometimes by archival studies (such as of police records) or surveys but mostly by controlled experiments. Data from the experiments are analyzed to see if the results could be explained by chance. Results are peer reviewed before publishing, and it often takes several confirmatory studies of the same hypothesis for a research result to become generally accepted in the scientific community.

For example, in 2004 researchers at Yale University School of Medicine published the results of an experiment on the effects of high and low levels of stress on eyewitness memory.[3] The study tested the popular assumption that people are better at remembering events that are stressful and raise a physiological alarm response.[4] The scientists hypothesized, based on anecdotal evidence from returning veterans, that this theory may not be true for high stress situations.[5] Working with commanders in charge of armed forces prisoner-of-war training, researchers designed an experiment where over 500 healthy military personnel underwent low stress interrogation exercises (using verbal pressure over a forty-minute period) and high stress interrogation (including both verbal and physical confrontation over the same period). Twenty-four hours later, personnel from both groups were asked if their interrogators were present in a photo array of sixteen pictures presented simultaneously, a "live" lineup of fifteen people, or a sequence of sixteen photos shown one at time.[6] If participants identified an individual as their interrogator, they were asked their level of confidence in that determination.[7] The experimental results show that almost three-fourths of the high stress group and a fourth of the low-stress group could not identify their interrogators, that a sequential identification procedure produced the most accurate results, and that subjects' stated confidence levels were not reliable indicators of accuracy. This large, realistic study has already spurred further reviews of the negative effects of stress on eyewitness reliability, a finding about which there appears to be general consensus emerging among experts.[8]

In fact, relative to other scientific research that enters courtrooms, the lack of controversy in the field of eyewitness identification is remarkable. A 2001 survey of established eyewitness researchers found near unanimity that the above-mentioned findings and many others were reliable and established in scientific literature.[9] If eyewitness experts are thought of as researchers who offer fact-finders information that helps educate their thinking about the reliability of a

---

3. Charles A. Morgan, III et al., *Accuracy of Eyewitness Memory for Persons Encountered During Exposure to Highly Intense Stress*, 27 INT'L J.L. & PSYCHIATRY 265, 265–67 (2004).

4. *Id.* at 265.

5. *Id.*

6. *Id.* at 269–70.

7. *Id.*

8. *See* Kenneth A. Deffenbacher et al., *A Meta-Analytic Review of the Effects of High Stress on Eyewitness Memory*, 28 LAW & HUM. BEHAV. 687, 699 (2004).

9. Saul Kassin et al., *On the "General Acceptance" of Eyewitness Testimony Research: A New Survey of the Experts*, 56 AM. PSYCHOLOGIST 405, 413–14 (2001).

HeinOnline -- 46 Jurimetrics 179 2005-2006

witness' testimony, experts may be analogized to other psychological and mental health experts whose testimony is routinely heard by jurors throughout the country. However, the level of consensus and the quantity of research in the eyewitness identification arena (well over 2,000 studies have been done)[10] vastly exceeds the research conducted with respect to most other mental health evidence. If one analogizes further and tries to compare the empirical bases and expert consensus in eyewitness research with fields sometimes considered to be "junk science," like local narcotics trafficking pattern experts who are admitted in most courtrooms, the reliability of eyewitness experts is even more obvious.

Of course, there are limits to what eyewitness research can do. Typically, eyewitness experts are prepared to testify in court about the extent to which the research literature explains how a particular factor, considered alone or in combination with others, likely would affect the reliability of an identification.[11] For example, all other things being equal, a victim's identification of an assailant is likely to be less reliable when a weapon was used in the commission of the crime or when the victim views a simultaneous lineup (compared to a crime without a weapon or a sequential lineup showing suspects one-by-one). Experts do not comment on the reliability of a particular witness' memory.

In short, eyewitness reliability research today is an established body of knowledge. It uses well-accepted methodologies. It is part of the research agenda at major universities throughout the world. It is a subject of thousands of peer-reviewed publications. It has existed for decades. There is nearly unanimous consensus among researchers about the field's core findings, and the conclusions of eyewitness reliability research provide an empirical basis for deciding some of the most difficult and pressing questions in this nation's courts.

I. THE TIMELINESS OF RELIABILITY RESEARCH

Given the well-established status of eyewitness research, one would expect such findings to have made their way into courtrooms around the country. Without access to research findings, jurors, judges, and attorneys have only their common sense intuitions with which to gauge an eyewitness who pauses on the stand, points, and says "that's him." Because such in-court identifications are often the most emotionally moving evidence heard in criminal trials, common sense intuitions usually favor crediting the eyewitness' word. Yet, common sense is often wrong.[12]

---

10. CUTLER & PENROD, *supra* note 2, at 68.

11. *Id.* at 19.

12. The experience of Jennifer Thompson, a twenty-two-year-old college student who was raped in 1984 in her own home, provides one example. *See* Jennifer Thompson, Op-Ed., *I Was Certain, But I Was Wrong*, N.Y. TIMES, June 18, 2000, § 4, at 15. She spoke at length with her assailant. She took every opportunity to study his face as they moved by several locations in the house where lights were on. She studied the man's face, his hairline, his body. A few days after the attack she positively identified a suspect from a photo array. Investigators arrested the individual, and she again picked him out of a live lineup. At both identifications and at trial she stated with complete certainty that the suspect, named Ronald Cotton, was the man who raped her. She was wrong. Eleven years after being incarcerated, Ronald Cotton was exonerated by DNA evidence that also inculpated a man with similar

HeinOnline -- 46 Jurimetrics 180 2005-2006

Prompted by DNA exonerations, in the last five years, executive and legislative branches at all levels of government have begun to reform eyewitness identification procedures used by law enforcement officials so as to minimize the risk of altering or biasing eyewitness' memories.

A major catalyst was a 1999 guide to eyewitness evidence published by the National Institute of Justice[13] that was followed by a training manual in 2003.[14] This guide, drafted by a multidisciplinary panel of eyewitness experts and law enforcement members, was given to every state and federal agency in the country. It presents best-practice recommendations for how to gather identification evidence (such as via a photo array, a lineup, or interviews) in a manner that is nearly as rigorous as rules governing the collection and handling of physical evidence.

Numerous police agencies have adopted new eyewitness procedures in Boston, Minneapolis, and other cities.[15] In New Jersey and North Carolina, procedural reforms have been undertaken by the state-level executive branch. Illinois, Virginia, and thirteen other states have completed or pending legislative efforts to mandate state-level reform.[16]

The judicial system, too, has become somewhat more receptive to expert testimony and the use of social science in trials. The Federal Rules of Evidence (FRE) were adopted in 1975, immediately preceding the explosion of research on eyewitness reliability. They, and parallel state rules of evidence,[17] directly address the admissibility of scientific research by qualified experts in three provisions. Most importantly, FRE 702 and similar rules provide that:

> If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as

---

features named Bobby Poole. Poole pled guilty to the crime in 1995. Since that time, Jennifer Thompson has become an outspoken voice for scrutinizing eyewitness identifications with all possible care. Still, despite the evidence and her acceptance of the fact that he was not the one who raped her, Jennifer Thompson's false memories are so real that when she pauses and thinks of her attacker she says she still sees Ronald Cotton. *Frontline: What Jennifer Saw* (PBS television broadcast Feb. 25, 1997). (transcript available at http://www.pbs.org/wgbh/pages/frontline/shows/dna/interviews/thompson.html).

13. NATIONAL INSTITUTE OF JUSTICE, EYEWITNESS EVIDENCE: A GUIDE FOR LAW ENFORCEMENT (1999).

14. NATIONAL INSTITUTE OF JUSTICE, EYEWITNESS EVIDENCE: A TRAINER'S MANUAL FOR LAW ENFORCEMENT (2003).

15. Scott Ehlers, *Eyewitness Identification: State Law Reform*, THE CHAMPION, Apr. 2005, at 34.

16. The extent of reform being considered in these jurisdictions varies widely. Some follow recommendations from scientists laid out in the American Psychology-Law Society (APLS) recommendations (a division of the American Psychological Association), others use modifications of the APLS recommendations by legal organizations such as the American Bar Association or Department of Justice guidelines.

17. Today, all but two states have adopted rules of evidence identical or substantially similar to FRE 702. For a survey of state court rules of scientific admissibility through December 15, 1997, *see* Heather G. Hamilton, Note, *The Movement from Frye to Daubert: Where Do the States Stand?*, 38 JURIMETRICS J. 201, 210–13 (1998). Since Ms. Hamilton's note was published, Connecticut and Pennsylvania have adopted comparable language. *See* CONN. CODE EVID. § 7-2 and PA. R. EVID. 702. Only New York and Massachusetts have substantially different rules.

HeinOnline -- 46 Jurimetrics 181 2005-2006

> an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.[18]

Other key rules include Rule 704 (allowing expert testimony even when it touches on an ultimate issue to be decided by the fact finder) and Rule 403 (allowing judges to reject any testimony when its probity is "substantially outweighed" by unfair prejudice or confusion of the issues).

More than mere codifications of existing practices and case law, FRE 702, 704, and 403 were liberalizing reforms intended to encourage the use of expert witnesses. As expected, overall use of expert witnesses has greatly increased under the rules.[19] The Supreme Court's 1995 *Daubert* ruling has opened the door even wider to expert testimony by letting judges apply evidentiary rules even when a field lacks general consensus among experts. Precisely how these rules of evidence have been applied to judicial decisions about eyewitness experts is further explained below. However, the broad trend to increase the use of scientific evidence in courts bears emphasis—such evidence is to be admitted if it is scientifically sound, helpful to laypersons, and not substantially outweighed by any prejudice or confusion.

Another legal trend supports admission of expert testimony regarding eyewitness identification. Over the last thirty years, under the due process clause and the Sixth Amendment, courts around the country have reaffirmed and strengthened the right of an accused to present a complete and zealous defense, including presentation of expert testimony. In *Chambers v. Mississippi*,[20] the Supreme Court held that a defendant's right to offer witness testimony on his behalf is one of the most fundamental constitutional rights.[21] Building upon this foundation, the Supreme Court, in cases like *Ake v. Oklahoma*[22] and *Crane v.*

---

18. FED. R. EVID. 702. The Advisory Committee characterizes the ability to "assist the trier of fact" as a question whether the "untrained layman would be qualified to determine intelligently and to the best possible degree the particular issue . . . ." FED. R. EVID. 702 advisory committee's note. Prior to amendment in 2000, the rule stated: "If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise." FED. R. EVID. 702 (1999). The amendment was not intended to make any substantive change to the rule as interpreted by the federal courts. *See* FED. R. EVID. 702 advisory committee's note.

19. *See, e.g.*, FED. R. EVID. 702 advisory committee's note on proposed revision stating:

> This revision is intended to limit the use, but increase the utility and reliability, of party-initiated opinion testimony bearing on scientific and technical issues. The use of such testimony has greatly increased since enactment of the Federal Rules of Evidence. This result was intended by the drafters of the rule, who were responding to concerns that the restraints previously imposed on expert testimony were artificial and an impediment to the illumination of technical issues in dispute.

Id. (citations omitted).

20. 410 U.S. 284 (1973).

21. *Id.* at 302.

22. 470 U.S. 68 (1985).

HeinOnline -- 46 Jurimetrics 182 2005-2006

Kentucky,[23] has emphasized that the adversary system depends on defendants' ability to contest prosecution theories with expert testimony and to challenge the circumstances under which government evidence was gathered. *Ake* held that state funding of expert psychiatric testimony for defendants was constitutionally required whenever sanity is a significant factor at trial because lay jurors would run the risk of an inaccurate weighing of the facts without expert advice.[24] *Crane* held that where an inculpatory confession was nearly the only evidence against the defendant, testimony must be allowed regarding details of the physical and psychological environment present during a confession because these go to the reliability and credibility of the evidence.[25] Such cases on the right to present defense witnesses dovetail neatly with the recent *Apprendi v. New Jersey*[26] decision that emphasizes the unique role of jurors, rather than judges, as the ultimate fact finders. The overall message is that due process requires that jurors be allowed to consider and resolve contested factual matters, and—if those factual matters involve questions in which expert testimony can be helpful—to hear expert testimony when offered by an accused in a criminal trial.

The powerful stories of wrongful convictions based on flawed eyewitness evidence, the efforts of other government branches and other organizations to reform eyewitness procedures, and broader judicial trends to liberalize evidence rules all combine to make the introduction of eyewitness research into the courts especially timely. Indeed, because these trends all point in the same direction, toward introduction of eyewitness science into the courtroom, one would expect that an eyewitness expert "revolution" would have taken place over the past thirty years. Such an expectation, however, would be misguided.

II. JUDICIAL RESISTANCE TO RELIABILITY RESEARCH

All but a few courts have chosen to ignore clear evidence that denial of defense requests to present expert testimony about eyewitness reliability leads to wrongful convictions. A poignant example of this phenomenon is the Maryland case of Kirk Bloodsworth,[27] the namesake of a section of Congress's Innocence Protection Act that provides for postconviction DNA testing.[28] Mr. Bloodsworth was twice convicted of the murder and rape of a young girl, chiefly on the basis of one young eyewitness' composite sketch and two eyewitnesses' lineup and in-court identifications. At trial, counsel for Bloodsworth requested that jurors be allowed to hear the findings of an eyewitness expert. Defense counsel's motion

23. 476 U.S. 683 (1986).
24. *Ake*, 470 U.S. at 82.
25. *Crane*, 476 U.S. at 691.
26. 530 U.S. 466 (2000).
27. Bloodsworth v. State, 512 A.2d 1056 (Md. 1986).
28. Justice for All Act of 2004, 42 U.S.C. § 14136(e) (2004).

HeinOnline -- 46 Jurimetrics 183 2005-2006

was denied by the trial court on grounds that an expert would not be helpful,[29] and Bloodsworth then spent nine years in prison before DNA evidence exonerated him.[30]

Such resistance is all the more surprising because long before DNA exoneration could prove what is now known—that mistaken eyewitness identifications are the leading cause of wrongful convictions—courts recognized that eyewitness evidence is not as reliable as many jurors think it is. The Supreme Court, first in the 1927 case of Sacco and Vanzetti[31] and again in the 1967 case of *United States v. Wade*,[32] emphasized that eyewitness identification testimony can be deeply problematic if not carefully monitored.[33]

Despite the judiciary's general awareness that eyewitness identifications are problematic, most courts are hostile to the introduction of eyewitness social science in criminal trials. Perhaps this has something to do with the judiciary's own limited understanding of the social science research. One of the largest studies to date on nonexperts' knowledge about eyewitness reliability has shown that individual judges' knowledge of the factors indicating reliable eyewitness claims is flawed.[34] In that survey more than two-thirds of respondent judges gave answers contrary to established research in three of the six questions asked about basic issues such as the link between eyewitness confidence and accuracy.[35] It may be the very familiarity of judges with a few of the problems with eyewitness testimony that leads many to wrongly assume they need not learn more and, more importantly, that jurors need not learn more either. Such assumptions result in courts around the country refusing to allow criminal defendants to provide jurors with information that could help place eyewitness testimony into a more informed context.

---

29. *Bloodsworth*, 512 A.2d at 1063:

> I am concerned that the possibility of admitting the evidence would tend to confuse or mislead the jury. This is not just a matter of usurping the province of the jury, although it is in my judgment most certainly that, it is also that such testimony is of little value in aiding the jury in this case. I'm not persuaded that the testimony will be helpful to the jury in understanding the evidence in this case.

30. Shockingly, the case affirming Mr. Bloodsworth's conviction remains controlling law on the admissibility of eyewitness expert testimony in Maryland and is an egregious example of the widespread judicial resistance to this particular sort of expert testimony. Further details about Mr. Bloodsworth's wrongful conviction may be found in TIM JUNKIN, BLOODSWORTH: THE TRUE STORY OF THE FIRST DEATH ROW INMATE EXONERATED BY DNA (2004).

31. Sacco v. Massachusetts, 275 U.S. 574 (1927).

32. 388 U.S. 218 (1967).

33. Justice Brennan, writing for the majority, quoted Justice Frankfurter:

> The vagaries of eye-witness identification are well known: the annals of criminal law are rife with instances of mistaken identification. Mr. Justice Frankfurter once said: "What is the worth of identification testimony even when uncontradicted? The identification of strangers is proverbially untrustworthy. The hazards of such testimony are established by a formidable number of instances in the records of English and American trials.

Id. at 228 (citations omitted).

34. Richard A. Wise & Martin A. Safer, *A Survey of Judges' Knowledge and Beliefs About Eyewitness Testimony*, 40 CT. REV. 6 (2003).

35. *Id.* at 8.

HeinOnline -- 46 Jurimetrics 184 2005-2006

A. Current Case Law

Appellate courts' reluctance to admit eyewitness research usually shows itself in the practice of routinely upholding lower court denials of defendants' motions to allow expert researchers to testify.[36] In reaching their decisions, appellate courts have used four main approaches.[37]

1. *Pure Discretion Regarding Experts*

The majority rule in approximately twenty-eight state and federal appellate courts is one of pure deference: appellate courts simply say they will defer to trial courts' discretionary decisions.[38] The decisions of these majority rule courts offer no standing rules to lower courts as to what fact patterns require eyewitness experts to be admitted. A few such courts, without more, require trial judges to examine the unique facts of the case when eyewitness identifications are key to

---

36. Occasionally there are skirmishes to get jury instructions about factors relevant to eyewitness reliability; however, very few of these decisions have received appellate review. *But see* State v. Cromedy, 727 A.2d 457, 458–49 (N.J. 1999) (dealing with cross-racial identification jury instruction).

37. As of fall 2005, approximately ten appellate courts, in addition to the United States Supreme Court, had not given separate consideration to eyewitness identification experts and relied on rulings concerning expert witnesses more generally. There have been no rulings by the highest appellate courts in the following states: Hawaii, Michigan; Mississippi; New Hampshire; New Mexico; North Carolina (but one recent appellate case used a pure discretionary standard, State v. Cole, 556 S.E.2d 666, 670 (N.C. Ct. App. 2001)); Oklahoma (but eyewitness experts have been admitted by a lower appellate court, Bristol v. State, 764 P.2d 887, 890 (Okla. Crim. App. 1988)); Oregon (but admission appears to be discretionary, State v. Fox, 779 P.2d 197, 197–98 (Or. Ct. App. 1989)); Virginia (but appears to be admitted in narrow circumstances, Currie v. Commonwealth, 515 S.E.2d 335, 338–39 (Va. Ct. App. 1999)); Wisconsin (but appears not to be admitted except in narrow circumstances, State v. Sullivan, No. 02-2775-CR, 2003 WL 22091937, at *3 (Wis. Ct. App. Sept. 3, 2003).

38. As of fall 2005, approximately five federal circuits and twenty-three states and the District of Columbia held that the admissibility of eyewitness identification experts depends on the exercise of trial court discretion without giving any general rules as to what facts might constitute an abuse of discretion: United States v. Stokes, 388 F.3d 21 (1st Cir. 2004), *rev'd on other grounds*, 125 S. Ct. 1678 (2005); United States v. Welch, 368 F.3d 970 (7th Cir. 2004), *rev'd on other grounds*, 125 S. Ct. 1063 (2005); Hager v. United States, 856 A.2d 1143 1147–48 (D.C. 2004); United States v. Smith, 156 F.3d 1046, 1053 (10th Cir. 1998); United States v. Jackson, 50 F.3d 1335,1340 (5th Cir. 1995); United States v. Rincon, 28 F.3d 921, 923 (9th Cir. 1994); Parker v. State, 968 S.W.2d 592, 596–97 (Ark. 1998) (further narrowing Utley v. State, 826 S.W.2d 268 (Ark. 1992)); Campbell v. People, 814 P.2d 1, 7 (Colo. 1991) (en banc); State v. McClendon, 730 A.2d 1107, 1114 (Conn. 1999); Garden v. State, 815 A.2d 327 (Del. 2003), *rev'd on other grounds*, Garden v. State, 844 A.2d 311 (Del. 2004); McMullen v. State, 714 So. 2d 368, 372 (Fla. 1998); Johnson v. State, 526 S.E.2d 549, 552–53 (Ga. 2000); State v. Alger, 764 P.2d 119, 127–28 (Idaho Ct. App. 1988) (distinguishing the only state supreme court ruling on the subject that occurred prior to a change in evidence rules, State v. Hoisington, 657 P.2d 17 (Idaho 1983)); People v. Enis, 743 N.E.2d 1 (Ill. 2000); State v. Schutz, 579 N.W.2d 317 (Iowa 1998); State v. Kelly, 752 A.2d 188, 191 (Me. 2000); Bloodsworth v. State, 512 A.2d 1056, 1064 (Md. 1986); Commonwealth v. Zimmerman, 804 N.E.2d 336, 340 (Mass. 2004); State v. Miles, 585 N.W.2d 368, 371 (Minn. 1998); State v. Long, 575 A.2d 435, 463 (N.J. 1990); People v. Lee, 750 N.E.2d 63, 65 (N.Y. 2001); State v. Fontaine, 382 N.W.2d 374, 377 (N.D. 1986); State v. Werner, 851 A.2d 1093, 1100 (R.I. 2004); State v. McCord, 505 N.W.2d 388, 391 (S.D. 1993); Weatherred v. State, 15 S.W.3d 540, 543 (Tex. Crim. App. 2000); State v. Percy, 595 A.2d 248, 252–53 (Vt. 1990); State v. Cheatam, 81 P.3d 830, 840 (Wash. 2003); State v. Taylor, 490 S.E.2d 748, 754 (W. Va. 1997); Engberg v. Meyer, 820 P.2d 70, 80 (Wy. 1991).

HeinOnline -- 46 Jurimetrics 185 2005-2006

the case—a kind of minimal rule that discretion should be exercised on the facts, but one free of any guidance as to how facts should be weighed.[39] This "pure discretionary" standard is typically announced by appellate courts with little explanation and often seems to be hollow. For instance, some jurisdictions have said they apply a discretionary standard while at the same time categorically stating that juries do not need expert testimony.[40]

With such deferential standards of review, trial court decisions in majority rule jurisdictions are *de facto* final. Review for "abuse of discretion" is vacuous. Since most cases coming up on appeal are denials of defense motions to admit experts, most appellate case law in these "purely discretionary" jurisdictions simply recite trial-level rationales about why experts are excluded and so appear to discourage the admission of eyewitness reliability experts. The appearance that this expert testimony is generally inadmissible is enhanced by the practical barriers to appellate review of any trial court decisions permitting expert testimony. Because the prosecution cannot appeal from an acquittal, cannot generally raise an interlocutory challenge, and has no reason to raise such claims in an appeal from a conviction, a trial judge searching through appellate decisions on the subject will likely look in vain for decisions affirming a decision to admit expert testimony. This is not because such rulings never occur, but because such decisions never get challenged on appeal and published.

Against this majority rule stand state and federal jurisdictions that have circumscribed trial court discretion somewhat or completely.

2. *Limited Discretion to Deny Experts*

Of the twenty-two jurisdictions opposing the pure discretion approach, about eleven purport to require the admission of eyewitness identification experts in limited circumstances, generally where an identification by a stranger is uncorroborated and occurs many weeks after the initial incident.[41] Early rulings in Arizona, California, and the Third Circuit started this so-called "modern trend" toward requiring the admission of eyewitness experts when the entire case turns

---

39. United States v. Smith, 156 F.3d 1046, 1053 (10th Cir. 1998); Campbell v. People, 814 P.2d 1, 8–9 (Colo. 1991) (en banc); Johnson v. State, 526 S.E.2d 549, 552–53 (Ga. 2000); State v. Cheatam, 81 P.3d 830, 842 (Wash. 2003).

40. *See, e.g.*, McMullen v. State, 714 So. 2d 368, 371–72 (Fla. 1998) (upholding an earlier decision by that court that espoused a discretionary approach while saying a jury is never in need of an eyewitness expert's advice).

41. One federal circuit and ten states have articulated some kind of fixed rules regarding what factual circumstances require the admission of eyewitness experts: United States v. Mathis, 264 F.3d 321, 340 (3d Cir. 2001); Skamarocius v. State, 731 P.2d 63, 67 (Alaska Ct. App. 1987); State v. McCutcheon, 781 P.2d 31, 34 (Ariz. 1989) (narrowing State v. Chapple, 660 P.2d 1208 (Ariz. 1983)); People v. Jones, 70 P.3d 359, 374 (Cal. 2003) (following People v. McDonald, 690 P.2d 709 (Cal. 1984)); Cook v. State, 734 N.E.2d 563, 570 (Ind. 2000); Commonwealth v. Christie, 98 S.W.3d 485, 490 (Ky. 2002); State v. DuBray, 77 P.3d 247, 255 (Mont. 2003); White v. State, 926 P.2d 291, 293 (Nev. 1996); State v. Buell, 489 N.E.2d 795, 803 (Ohio 1986); State v. Frazier, 592 S.E.2d 621, 623 (S.C. 2004) (following State v. Whaley, 406 S.E.2d 369 (S.C. 1991)); State v. Maestas, 63 P.3d 621, 626 (Utah 2002).

HeinOnline -- 46 Jurimetrics 186 2005-2006

upon eyewitness evidence.[42] However, the "modern trend" has only been followed by a few of the more recent appellate decisions. In fact, Arizona retreated from its broad language in 1989,[43] and only one state supreme court followed suit in the 1990s.[44]

3. *Limited Discretion to Admit Experts*

A third approach, also in opposition to the pure discretion trend, makes the exclusion of eyewitness experts a default rule except in special circumstances. Approximately six courts have precedent where the denial of experts is required except in circumstances similar to those where admission is required in other minority jurisdictions, such as no corroboration and a significant passage of time between sighting and identification.[45] On those unique facts, admission is discretionary.

4. *Prohibition On Experts*

The fourth approach is not to allow eyewitness experts to testify under any circumstance. Until the 1980s, few appellate jurisdictions had given separate attention to whether eyewitness experts could be admitted to testify and a flat prohibition on such testimony was the norm among trial courts. Prohibition of eyewitness expert testimony was largely abandoned in the 1990s when a wave of almost two dozen of the highest state and federal appellate courts adopted the above-mentioned "pure discretion" standard deferring to trial courts.[46] However, five high courts, including the Eleventh Circuit, still have a rule of total exclusion.[47]

In sum, very few jurisdictions encourage lower courts to provide jurors with eyewitness identification research when sought by criminal defendants. At best, a few jurisdictions require the admissibility of this evidence when its relevance and importance is indisputable. Most jurisdictions do not do even this much; the "pure discretionary" standard used by many courts operates to discourage judges from admitting eyewitness expert testimony, while other courts flatly prohibit the

---

42. State v. Chapple, 660 P.2d 1208,1223–24 (Ariz. 1983); People v. McDonald, 690 P.2d 709, 723 (Cal. 1984), *superceded by* 738 P.2d 764 (1987); United States v. Downing, 753 F.2d 1224, 1241 (3d Cir. 1985).

43. State v. McCutcheon, 781 P.2d 31, 34 (Ariz. 1989).

44. White v. State, 926 P.2d 291 (Nev. 1996).

45. Three federal circuits and three states have indicated that eyewitness experts should be excluded except in special circumstances: United States v. Martin, 391 F.3d 949, 954 (8th Cir. 2004) (following United States v. Kime, 99 F.3d 870 (8th Cir. 1996)); United States v. Bellamy, 26 Fed. Appx. 250, 259 (4th Cir. 2002) (following United States v. Harris, 995 F.2d 532 (6th Cir. 1993)); United States v. Lumpkin, 192 F.3d 280, 289 (2d Cir. 1999); *Ex parte* Williams, 594 So. 2d 1225, 1227 (Ala. 1992); State v. Chapman, 436 So. 2d 451, 453 (La. 1983); State v. Whitmill, 780 S.W.2d 45, 47 (Mo. 1989).

46. *See* cited cases, *supra* note 38.

47. United States v. Smith, 122 F.3d 1355, 1357–59 (11th Cir. 1997); State v. Gaines, 926 P.2d 641, 649 (Kan. 1996); State v. George, 645 N.W.2d 777, 790 (Neb. 2002); Commonwealth v. Simmons, 662 A.2d 621, 631 (Pa. 1995); State v. Coley, 32 S.W.3d 831, 833–34 (Tenn. 2000).

HeinOnline -- 46 Jurimetrics 187 2005-2006

use of such testimony altogether.[48] Although the appellate case law has arguably moved in the direction of admissibility over the past thirty years, the movement has been slow and the current standards have the effect of keeping this critical information away from jurors.

B. Rationales for Current Case Law

Why have courts moved so grudgingly in the area of eyewitness experts when the general trend in nearly all other scientific areas has been toward greater admissibility—particularly when evidence is offered by an accused in his own defense? Typically, appellate opinions discussing the admissibility of eyewitness expert testimony provide two reasons: (1) the reliability of the identification is not deemed a critical issue because of corroborating evidence; and (2) the proffered testimony does not pertain to a subject that courts believe is "beyond the ken" of the average juror and therefore "helpful" to the trier of fact.[49] Often courts rely on both rationales in explaining why it was permissible to keep the jury from learning about the eyewitness identification research.

1. *The Corroborating Evidence Rationale*

The first of these rationales—that expert testimony about reliability factors is unnecessary because the particular eyewitness identification at issue was not critical to the outcome or was sufficiently corroborated—is a fallacious standard for trial or appellate courts.

A primary problem with this argument is that judicial reliance on corroborating evidence seems to rest on a false premise—namely, that, where other corroborating evidence of guilt exists, defense testimony attacking the reliability of a government eyewitness is of such little importance as to be "irrelevant." If that were true, the proffered eyewitness testimony itself should have been excluded as irrelevant. This obviously never happens because eyewitness testimony is always important evidence.[50]

The opportunity to defend oneself before a jury against criminal charges is a bedrock foundation of our adversarial system. A regime where a trial judge decides whether a defense response to prosecution evidence is "important" enough to allow jurors to hear it resembles the Confrontation Clause regime recently condemned by the Supreme Court in *Crawford v. Washington*.[51] In *Crawford*, the question was whether the Sixth Amendment would permit judges to apportion confrontation rights by first deciding whether prosecution evidence

---

48. Discretionary standards do not inherently work this way. But the combination of two factors lead to this result: (1) lack of guidance about how to exercise discretion; and (2) no finding of abuse in cases excluding experts.

49. *See, e.g.*, People v. Lee, 750 N.E.2d 63, 66–67 (N.Y. 2001).

50. *See* Kyles v. Whitley, 514 U.S. 419, 446 (1995) (noting the clear importance of eyewitness testimony even in cases where substantial amounts of other evidence exists: "[t]he effective impeachment of one eyewitness can call for a new trial even though the attack does not extend to others, as we have said before.").

51. 541 U.S. 36 (2004).

HeinOnline -- 46 Jurimetrics 188 2005-2006

was "reliable" enough to make confrontation unnecessary. In rejecting this form of analysis, the Court explained that "dispensing with confrontation because testimony is obviously reliable is akin to dispensing with jury trial because a defendant is obviously guilty. This is not what the Sixth Amendment prescribes."[52] Similar language could be used to describe a regime in which the judge, and not the defendant, decides whether prosecution evidence is sufficiently corroborated to render any defense through the presentation of expert testimony unnecessary. Even if there is an independent DNA match identifying the defendant, it simply is not the proper role of the trial judge to exclude proffered expert testimony on government eyewitnesses.

But even if our adversarial system allowed judges to take issues away from juries, looking to corroborating evidence to determine whether an eyewitness identification expert should be permitted is an unsound way to identify those cases where testimony could be significant. First, trial courts deciding admissibility questions in a pretrial posture are poorly situated to resolve questions about whether the "corroborating" evidence is so substantial as to effectively eliminate any potential importance of the expert's proffered testimony to the fact finder. The danger is that prior to hearing all the evidence and without a complete understanding of the case, the trial judge will be swayed by the persuasive power of an eyewitness identification when she views the constellation of facts in a case. The result may be the creation of corroborating evidence out of nothing.[53]

Second, particularly in a pretrial posture, it is difficult for a trial court to determine whether corroborating evidence arose independently of the eyewitness identification. In a criminal investigation, evidence emerges from a dynamic context in which each item is affected by the establishment of other evidence. Knowledge of one eyewitness' identification can raise a second person's confidence in their identification or lead police to use more suggestive interrogation tactics with other potential witnesses.[54]

Reliance on corroborating evidence, thus, improperly prevents a defendant from responding to important government proof. Because any assessment of corroborating evidence must be made before all the evidence is in, trial courts may prohibit a defense response to prosecution evidence that jurors find to be critical.

---

52. *Id.* at 62.

53. In the case of exoneree Ronald Cotton, the government introduced a pair of black shoes and a flashlight found at his home into evidence at trial to corroborate Jennifer Thompson's eyewitness identification and description of a man wearing black shoes and possessing a flashlight. Apart from that eyewitness identification, Ronald Cotton's possession of such ordinary items would never have been deemed identification evidence.

54. Kirk Bloodsworth's case exemplifies how the corroboration of several eyewitnesses that seemed to mutually reinforce each other, in retrospect, may not be trustworthy when the identifications were not truly independent of one another. The same flawed interrogation procedures, investigator bias (or fraud), or postevent information supplied to eyewitnesses may create a false appearance of independent corroboration.

HeinOnline -- 46 Jurimetrics 189 2005-2006

Ending reliance on corroborating evidence will admittedly lead to the admission of expert testimony in many more cases. But this is a natural product of the adversarial system in which juries and not judges should resolve disputed questions about the reliability of a piece of government proof. Courts need not fear that such a regime would necessarily invalidate convictions in cases where judges erroneously exclude expert testimony. Harmless error review by appellate courts can effectively ensure that convictions will survive in those rare cases where any expert testimony would truly have been unimportant to the jury's verdict. This appears to be the intent of some trial and appellate courts when applying a corroboration rationale to exclude experts in cases where the evidence is in fact overwhelming; the courts are effectively saying that any exclusion of the expert testimony is harmless because there was so much independent evidence supporting the guilt of the defendant.

The harmless error rule has become well established through many years of application by appellate courts. Appellate courts applying a harmless error standard have a distinct advantage over trial courts in evaluating the significance of corroborating evidence because they have access to more complete information posttrial and because they have clear standards to guide their review. Harmless error standards, moreover, are strongly weighted to ensure that an error will not be deemed "harmless" unless the government can convince the court that the error could not have played any role in the verdict, thus reducing the likelihood that defense expert testimony will be excluded just because a judge, as opposed to a jury, has some intuition that the identification is correct.

Finally, the use of a harmless error standard has one other substantial benefit over a standard blankly excluding experts where there is thought to be corroborating evidence. Harmless error review begins with the determination that a ruling of the trial court, such as the improper exclusion of an expert response to eyewitness testimony, was in fact "error," and then proceeds to determine whether that error mattered in the context of a particular case. This form of analysis ensures that, in the next case, a trial court will permit the expert response because an appellate court has already held that the failure to allow the testimony was error, thus eliminating any sort of speculation about whether the expert testimony would have affected the verdict.

By contrast, where corroborating evidence is cited by appellate courts as the rationale for upholding the exclusion of an eyewitness expert without distinct error and harm analyses, trial courts receive the signal that it is up to them to assess how important the evidence is and whether it will matter to the jury when deciding whether to exclude expert testimony.[55]

In sum, whether the evidence corroborating an eyewitness identification is strong or weak, rarely can a reasonable basis be found to disallow expert testimony about the reliability of that identification on grounds that it is merely "irrelevant" or "sufficiently corroborated."

---

55. Manson v. Brathwaite, 432 U.S. 98, 118 & n.* (1977) (Stevens, J., concurring) (warning against reliance on corroborating evidence when making decisions about the admissibility of identification testimony).

HeinOnline -- 46 Jurimetrics 190 2005-2006

2. *The "Beyond the Ken" Rationale*

The second rationale proffered by courts in current case law—that eyewitness expert testimony is "not beyond the ken" of jurors—is slightly more defensible on its face because it at least applies the proper standard for the admissibility of a particular item of expert testimony. The "beyond the ken" rationale derives from the "helpfulness standard" of Rule 702. The rule requires that the expert testimony "will assist the trier of fact to understand the evidence."[56] Even evidence that only marginally improves jurors' ability to determine issues meets this standard. Because of their short opinions, it is not always clear whether courts that refer to the helpfulness standard mean to say that, given corroborating evidence in the case, experts would add nothing more to jurors' ability to judge eyewitness reliability, or whether judges simply believe that jurors know the findings of eyewitness experts as a matter of common sense. In either case, courts say eyewitness research is "not beyond the ken" of jurors.

The problem with this analysis, however, is that (a) it is not apparent from their decisions that trial or appellate judges fully understand what the proffered eyewitness expert testimony would have been; and (b) appellate and trial judges speculate about what jurors understand about eyewitness expert testimony without any hard evidence.

There is considerable evidence that courts do not adequately understand the very expert research they deem to be of no assistance to jurors. Most court opinions, especially the majority following the "pure discretion" approach, contain no literature review and do not parse out which particular aspects of the proffered expert testimony they deem to be common knowledge. In these cases, one can only conclude judges are relying on their experience and intuitions, which, as Wise and Safer demonstrated, are wrong half the time.[57] For example, without any expert guidance as to what the research says, only 31% of surveyed judges believed a description of the "forgetting curve" to be true, a research finding that memory often deteriorates quickly shortly after an event, then forgetting becomes more gradual.[58] Nonetheless, expert testimony about the "forgetting curve" is typically excluded as common knowledge when proffered by defendants.[59] Of course, a few court opinions, particularly those following the modern trend limiting discretion to deny experts, do take care to address particular eyewitness research findings. However, even these cases rarely cite to research and end up with unscientific and contrary holdings, disagreeing, for example, on whether the lack of significant correlation between a witness' confidence and their accuracy is commonly known.[60] Stress, too, and its

---

56. FED. R. EVID. 702.

57. Wise & Safer, *supra* note 34, at 8.

58. *Id.*

59. *See, e.g.*, Hager v. United States, 856 A.2d 1143, 1148–49 (D.C. 2004); Skamarocius v. State, 731 P.2d 63, 66 (Alaska Ct. App. 1987); Commonwealth v. Christie, 98 S.W.3d 485, 490 (Ky. 2002).

60. *Compare* Garden v. State, 815 A.2d 327 (Del. 2003) (*rev'd on other grounds*, Garden v. State, 844 A.2d 311 (Del. 2004)) (holding that expert testimony on the correlation of confidence and

HeinOnline -- 46 Jurimetrics 191 2005-2006

debilitating effects on memory, is thought to be common sense to some of those few courts who have considered the matter counterintuitive to others.[61] In short, while most appellate courts do not address specific eyewitness research findings and do not examine the research, those that do have generated conflicting precedent that gives trial courts no sure guidance in exercising their discretion.

Most importantly, courts also do not understand what jurors know about eyewitness reliability. Until now, no one has appropriately measured what jurors know about eyewitness reliability. Research into laypersons' knowledge of eyewitness reliability factors has been sparse. Eyewitness researchers, in the course of conducting more targeted experiments, have repeatedly found their subjects to be unaware of basic scientific findings of the field. There also have been some small, nonrandomized surveys about juror understanding of eyewitness reliability that indicated a general lack of knowledge.[62] Based on these small surveys and their laboratory experience, scientists studying eyewitness reliability have come to believe that ordinary lay jurors know little about the subject as a matter of common sense.

These researchers' opinions are in sharp contrast to judicial opinions. For example, recent surveys have shown that mug-shot-induced bias—the fact that prior exposure to mugshots of a suspect increases the likelihood that a witness will later choose the suspect from a lineup—was thought to be a matter of common sense by only 13% of researchers compared with 74% of judges.[63] The rift between the courts' and researchers' beliefs about what jurors understand about eyewitness reliability has been unresolved because no one has performed a large, empirical study of jurors' knowledge. Thus, judicial rulings about what is "beyond the ken" of jurors are, at best, speculation based solely on judges' experience and intuition.[64]

Neither judicial rationale for excluding eyewitness experts—that their testimony is unnecessary because of corroborating evidence or not helpful because it is not beyond the ken of jurors—holds up under examination. The first rationale is logically and legally flawed, and the second is mere speculation.

To better ground their assertions about what is "beyond the ken," judges would have to understand both what eyewitness research says and whether

---

accuracy should be permitted to the same extent as other admissible topics such as cross-racial bias) *and* Skamarocius, 731 P.2d at 66 (finding that confidence is commonly, but mistakenly, thought to correlate with accuracy) *and* United States v. Downing, 753 F.2d 1224, 1230 n.6 (3d Cir. 1985) (noting that confidence-accuracy studies are often counterintuitive to jurors) *with* United States v. Mathis, 264 F.3d 321, 340 (3d Cir. 2001) (holding that the correlation between confidence and accuracy in identifications was not a proper subject for expert testimony).

61. *Compare* Commonwealth v. Christie, 98 S.W.3d 485, 490 (Ky. 2002) (finding stress to be among the narrow circumstances sufficient to support eyewitness expert admission) *and* United States v. Mathis, 264 F.3d 321, 340 (3d Cir. 2001) (finding stress is among the proper subjects for expert testimony) *with* Green v. United States, 718 A.2d 1042, 1053 (D.C. 1998) (upholding trial finding that effects of stress on eyewitnesses are not beyond the ken of jurors).

62. *See infra* note 69.

63. Wise & Safer, *supra* note 34, at 10.

64. At worst, such opinions are being disingenuous about other motives, such as keeping out evidence that almost always benefits defendants.

HeinOnline -- 46 Jurimetrics 192 2005-2006

laypersons' common sense understanding of these issues parallels the research findings. If jurors can be shown to understand eyewitness research findings as a matter of common sense, then the judicial rationale for excluding experts as "unhelpful" is valid. But, if eyewitness research findings are not known to jurors, the only plausible rationale for excluding eyewitness experts is wrong and courts should reevaluate their positions on eyewitness expert admissibility.

III. EMPIRICAL FINDINGS

In the winter of 2004, lawyers from the Public Defender Service for the District of Columbia (PDS) began investigating whether jurors do understand as a matter of common sense what makes some eyewitness identifications more or less reliable than others.[65] Preliminary investigation showed that despite extensive academic research into the reliability of eyewitnesses, there was little direct research on the extent to which nonexperts' intuitions match experts' findings.[66] The surveys indicate that while many attorneys have sound common-sense opinions about factors like cross-racial identification, the effects of stress or violence, and the importance of unbiased lineup instructions, most lack an accurate understanding of eyewitness confidence or bias in simultaneous (versus sequential) lineups.[67] Surveys of judges (mentioned above) indicate they have correct intuitions at roughly the same rates as attorneys.[68] However, there have been only a handful of surveys about the knowledge of nonexperts outside the legal system, and most of these have been nonrandomized studies using undergraduate or law students.[69] These latter surveys have used a diagnostic test called the Knowledge of Eyewitness Behavior Questionnaire (KEBQ) that poses scenarios with multiple factors relevant to eyewitness identifications. As with the judicial and attorney surveys, a significant portion of students taking the KEBQ were found to have some awareness of factors, such as cross-racial identification, but grossly misunderstand others, like the fact that training does not improve a witness' accuracy.[70]

---

65. This inquiry was an outgrowth of PDS's duty to give the best representation possible to indigent clients and was conducted in the context of several ongoing cases involving purported eyewitness identifications.

66. A few studies have assessed attorneys' understanding of eyewitness reliability variables. All of these studies, however, involve small samples and some date from the early 1980s. For an overview, see Jennifer Devenport et al., *Eyewitness Identification Evidence*, 3 PSYCHOL. PUB. POL'Y & L. 338 (1997).

67. *Id.* at 343.

68. *Compare* Wise & Safer, *supra* note 34, *with* CUTLER & PENROD, *supra* note 2, at 159–68.

69. Kenneth A. Deffenbacher & Elizabeth F. Loftus, *Do Jurors Share a Common Understanding Concerning Eyewitness Behavior?*, 6 LAW & HUM. BEHAV. 15 (1982); Kevin M. McConkey & Suzanne M. Roche, *Knowledge of Eyewitness Memory*, 24 AUSTL. PSYCHOL. 377 (1989); Elizabeth Noon & Clive R. Hollin, *Lay Knowledge of Eyewitness Behaviour: A British Survey*, 1 APPLIED COGNITIVE PSYCHOL. 143 (1987).

70. CUTLER & PENROD, *supra* note 2, at 101, 175 Table 11.1 (1995) (noting that even those factors most often recognized by test takers, like cross-racial bias, were missed by 25% or more of the participants).

HeinOnline -- 46 Jurimetrics 193 2005-2006

There has been significantly more "indirect" research into the attitudes of nonexperts about eyewitness reliability, such as research that does not attempt to survey a population's self-described understanding of reliability but instead relies upon smaller studies of how laypersons reason about eyewitness identification factors.[71] For example, postdiction studies required respondents to read summaries of eyewitness identification experiments and then predict the rate of accurate identifications in the experiment. Similarly, judgment studies put nonexperts through mock trials and assessed jurors' sensitivity to different aspects of the case through questionnaires and their final verdict. Considered together, the quantity of this "indirect" research into laypersons' knowledge of eyewitness research findings suggests jurors' common sense reasoning on these matters is often flawed. However, this kind of research lacks the persuasive value of direct surveys assessing the knowledge of the jury pool.[72]

To describe the Washington, D.C. juror pool, PDS decided to conduct its own research on the issue.[73] PDS accordingly commissioned a Washington, D.C. polling firm to survey approximately 1,000 potential District of Columbia jurors. The survey results, discussed in detail below, demonstrate that judicial assertions about jurors' ability to appraise eyewitness identifications are wrong. Jurors suffer from a basic misunderstanding of how memory generally works and do not understand how particular factors, such as the effects of stress or the use of a weapon, affect the accuracy of eyewitness testimony.

A. Mechanics of the PDS Survey

In the Fall of 2003 and early Winter of 2004, Peter D. Hart Research Associates, Inc. (Hart Research) worked together with attorneys at PDS and Dr. Elizabeth Loftus to prepare a survey that would discern what potential District of Columbia jurors understand about memory in general and the reliability of eyewitness identification evidence in particular. The questionnaire consisted of approximately twenty questions about whether jurors believe eyewitness testimony is generally reliable and also about the specific factors jurors believe would make an eyewitness identification more or less reliable.[74]

---

71. *Id.* at 346–49.

72. Admittedly, surveys are subject to the critique that they cannot assess how jurors are actually able to apply their knowledge about eyewitness identifications in the courtroom—a question to which postdiction and judgment studies may be better suited.

73. The project was initiated by Edward J. Ungvarsky, Public Defender Service for the District of Columbia (PDS) Special Counsel and a Vice Chair of the NACDL Forensic Evidence Committee, and PDS's Special Litigation Division, a unit that works exclusively on systemic criminal justice issues such as forensic evidence challenges and issues surrounding the suppression of exculpatory evidence by prosecutors. The motivation behind this project was simple: If jurors understand as a matter of common sense what makes some eyewitness identifications more reliable than others, it would not make sense for PDS to continue to devote resources toward educating already-informed jurors on this topic. On the other hand, if jurors actually have an incomplete understanding of the factors that affect eyewitness reliability, it would remain imperative to continue to seek to provide jurors with the tools necessary to ensure that they could intelligently evaluate the evidence presented to them by the government.

74. *See infra* Appendix.

HeinOnline -- 46 Jurimetrics 194 2005-2006

Between February 18, 2004, and February 23, 2004, Hart Research conducted a telephone survey in the District of Columbia. Residential phone numbers were chosen at random from the District of Columbia area code. One thousand two hundred ninety-six potential jurors were identified as U.S. citizens who were at least eighteen years of age and not currently on probation or parole. Participants were also asked demographic questions about education, neighborhood, age, employment, party affiliation, prior jury service, language, race, income and exposure to the criminal justice system. Ultimately, 1,007 potential jurors completed the survey. The margin of error (95% confidence) for the survey was plus-or-minus 3.1 percentage points.

1. *Juror Misunderstandings of Memory in General*

Human memory does not record events like a video recorder.[75] In the first place, human memory is more selective than a video camera. The sensory environment contains a vast amount of information, but the memory process perceives and accurately records only a very small percentage of that information. Second, because the act of remembering is reconstructive,[76] akin to putting puzzle pieces together, human memory can change in dramatic and unexpected ways because of the passage of time or subsequent events, such as exposure to "post-event" information like conversations with other witnesses or media reports. Third, memory can also be altered through the reconstruction process. Questioning a witness about what he or she perceived and requiring the witness to reconstruct the experience can cause the witness' memory to change by unconsciously blending the actual fragments of memory of the event with information provided during the memory retrieval process.

These characteristics of human memory have profound implications with regard to the accuracy of eyewitness claims that they "remember" seeing the accused or other key details about the crime. For a memory to be reliable, the witness must have accurately perceived the event, and the witness' memory must not have degraded over time or been polluted by postevent information and questioning. But more importantly, for jurors to fairly assess whether this claim is accurate, they must understand memory's complexity, selectivity, and malleability. Jurors must also understand what specific factors affect perception and encoding of memories, what factors can pollute memory, and what factors in the re-creation process can distort a witness' "memory" of an event.

PDS's survey of potential jurors in the District of Columbia suggests that juror understanding of these subjects fails at even the most basic level. Several survey questions, for example, were designed to test jurors' general understandings of the workings of human memory. One question asked whether

---

75. EDWARD ARNOLDS ET AL., EYEWITNESS TESTIMONY: STRATEGIES AND TACTICS 14–15 (1984); David M. Shofi, *The New York Courts' Lack of Direction and Discretion Regarding the Admissibility of Expert Identification Testimony*, 13 PACE L. REV. 1101, 1104 (1994); Gary L. Wells, *Eyewitness Behavior*, 4 LAW & HUM. BEHAV. 237, 238 (1980).

76. KARL HABERLANDT, HUMAN MEMORY: EXPLORATION AND APPLICATION 4 (1999) (defining and describing reconstructive memory).

HeinOnline -- 46 Jurimetrics 195 2005-2006

"the act of remembering a traumatic event [was] like a video recording in that one can recall details as if they had been imprinted or burned into one's brain." Over half of the respondents (52%) either thought this statement was true or did not know whether it was true. Indeed, 46% of potential jurors believe that the witness on the stand is effectively narrating a video recording of events that she can see in her "mind's eye" for jurors.

Other results demonstrated similar deficits of knowledge on the most basic level about how memory works. The survey asked potential jurors to assess the reliability of their own memories. Almost two-thirds of the respondents (66%) thought the statement "I never forget a face" applied "very well" or "fairly well" to them. Likewise, more than three out of four respondents (77%) thought that the phrase "I have an excellent memory" applied "very well or "fairly well" to them. The fact that such large majorities of people tend to believe their memories are above average suggests that potential jurors may begin each trial with unwarranted confidence in memory and the ability to identify faces generally. Such confidence may well be transferred to the testifying witnesses and cause jurors to overestimate the accuracy of witness memories as well.

2. *Specific Reliability Factors—Weapon Focus*

For more than twenty-five years, social scientists have posited that the presence of a weapon during a crime attracts the attention of the witness to the weapon, reducing attention to the culprit's facial and physical characteristics. This phenomenon is often referred to as "weapon focus."[77] A series of controlled studies have now validated the existence of this effect.[78] For example, experiments have been conducted involving videotaped robberies with some of the culprits brandishing a handgun and others concealing the gun. These studies have repeatedly demonstrated that witness accuracy was better when the gun was concealed than when the gun was brandished.

---

77. *See, e.g.*, ELIZABETH F. LOFTUS & JAMES M. DOYLE, EYEWITNESS TESTIMONY: CIVIL AND CRIMINAL, § 2.10 (2d ed. 1992); Elizabeth F. Loftus et al., *Some Facts About "Weapon Focus,"* 11 LAW & HUM. BEHAV. 55 (1987); Anne Maass & Gunther Kohnken, *Eyewitness Identification: Simulating the "Weapon Effect,"* 13 LAW & HUM. BEHAV. 397 (1989).

78. An archival study of actual police cases demonstrated that the presence of a weapon "did not reduce the quantity of descriptive information the witness was subsequently able to provide to police about the culprit, but it did impair witnesses' subsequent ability to recognize the culprit." *See* Donald P. Judges, *Two Cheers for the Department of Justice's Eyewitness Evidence: A Guide for Law Enforcement*, 53 ARK. L. REV. 231, 243 n.37 (2000) (citing Patricia A. Tollestrup et al., *Actual Victims and Witnesses to Robbery and Fraud: An Archival Analysis, in* ADULT EYEWITNESS TESTIMONY: CURRENT TRENDS AND DEVELOPMENTS 144, 158 (David F. Ross et al. eds., 1994)); *see also* Nancy Mehrkens Steblay, *A Meta-Analytic Review of the Weapon Focus Effect*, 16 LAW & HUM. BEHAV. 413 (1992). Experiments in which eye movements are monitored while subjects witness a scene where a weapon is involved also demonstrate the existence of the phenomenon of "weapon focus." Brian L. Cutler et al., *The Reliability of Eyewitness Identification*, 11 LAW & HUM. BEHAV. 233, 240, 244 (1987); Elizabeth F. Loftus et al., *Some Facts About "Weapon Focus,"* 11 LAW & HUM. BEHAV. 55, 57–61 (1987).

HeinOnline -- 46 Jurimetrics 196 2005-2006

Some courts have speculated that "weapon focus" is among those features of memory that jurors understand as a "matter of common sense."[79] The PDS survey asked potential jurors whether they thought that the fact that "a weapon is involved in the crime" tends to make "an eyewitness' memory about the details of the crime more reliable, less reliable or [would have] no effect." Thirty-seven percent of the respondents actually thought the presence of a weapon would make a witness' memory for event details *more* reliable, while thirty-three percent of the respondents thought that the presence of a weapon either would have no effect or were not sure of what effect a weapon would have. Only three out of ten potential jurors correctly understood that the presence of a weapon tends to make an eyewitness' memory for details less reliable.

3. *Specific Reliability Factors—Presence of Violence or Stress*

Social science studies have also shown that a person's ability to recall details of an event is likely worse if a witness has observed a violent event as opposed to a nonviolent one.[80] This is particularly true for the peripheral details.

To find out whether jurors understand the effect that violence has on witness identifications, the PDS survey asked potential jurors whether they thought that the fact that "a crime is violent" tends to make "an eyewitness' memory about the details of the crime more reliable, less reliable or [would have] no effect." Thirty-nine percent of the respondents thought that event violence would make a witness' memory for event details more reliable,[81] while 33% of the respondents thought that event violence either would have no effect or were not sure of what effect event violence would have. Only three out of ten potential jurors correctly understood that event violence tends to make an eyewitness' memory for details less reliable.

4. *Specific Reliability Factors—Duration of the Incident*

For over a century, social scientists have been conducting experiments concerning people's ability to estimate the duration of a stressful incident. Those studies have consistently shown not only that most people have enormous difficulty estimating the length of these events but also that the vast majority of these errors are in the same direction—toward overestimating the duration of a

---

79. Taylor v. United States, 451 A.2d 859, 866–67 n.9 (D.C. 1982).

80. *E.g.*, Morgan, III et al., *supra* note 3. In another study, it was observed that "[p]erceptual abilities actually decrease in a highly stressful situation, and the person under stress is less reliable than he or she would be otherwise. Such a witness becomes less capable of remembering details, less accurate in reading dials, and less certain in detecting signals." Thomas J. Feeney, *Expert Psychological Testimony on Credibility Issues*, 115 MIL. L. REV. 121, 146 (1987). "One theory indicates that moderate levels of stress or arousal increase performance up to a point. Under this theory, known as the Yerkes-Dodson law, perceptual performance follows a U-shaped curve. At very low levels of arousal, the senses are not yet functioning fully. Performance peaks at moderate levels of arousal and then declines as the stress increases further." *Id.* at 146 n.163 (citation omitted).

81. This is unsurprising. Researchers have postulated that lay people share "a common" but mistaken "belief that stress heightens a witness' observation powers and 'burns' an image of the scene into the mind." Feeney, *supra* note 80, at 146.

HeinOnline -- 46 Jurimetrics 197 2005-2006

stressful event. Overestimation can vary substantially depending upon the amount of stress accompanying the event.[82]

Thus, the longer an eyewitness has to observe a particular face, the more accurate her identification becomes. Although jurors understand this principle as a matter of common sense, they do not also understand the unreliability of witnesses' subjective time estimates. Often, the only piece of information a juror has about the length of exposure time is the witness' estimate. This estimate exaggerates the exposure time.

The PDS survey shows that jurors do not understand this phenomenon as a matter of common sense. Over 40% of survey respondents either thought that witness time estimates were accurate or were not sure whether such estimates were accurate. While that alone indicates that a significant number of jurors would be likely to overestimate exposure times and thus witness reliability, the beliefs of those who are more skeptical of witness time estimates are even more troubling. Of those who correctly understood that witnesses themselves are not good at evaluating how long an event took to unfold, a sizeable portion (about 25%) believed that witnesses *underestimate* the actual time. In all, 63% of the survey respondents do not understand what scientific research has demonstrated about a witness' ability to gauge the duration of an event. The jurors either believed witnesses' subjective time estimates or thought that witnesses tended to actually see a face for longer than claimed. Only 37% of the total respondents correctly understood events unfold faster than witnesses think they do.

5. *Specific Reliability Factors—Confidence*

With regularity, a witness in court will express 100% confidence in an identification. But is the identification of a "confident" witness more reliable than the identification of a less certain one? Witnesses who are highly confident in their identifications are only slightly more likely to be correct as compared to witnesses who are less sure of the identifications. In other words, the correlation between confidence and accuracy is weak.[83]

Eyewitness accuracy is a complex product of many factors. Factors such as the witness' eyesight and concentration, the amount of lighting, the length of

---

82. *See, e.g.*, ELIZABETH F. LOFTUS & JAMES M. DOYLE, EYEWITNESS TESTIMONY: CIVIL AND CRIMINAL 26–27 (1987) (explaining that a witness tends to overestimate the duration of an especially stressful or violent event); JAMES MARSHALL, LAW AND PSYCHOLOGY IN CONFLICT 41–81 (1966) (on average, viewers estimated that a 42-second film in which man rocks a baby carriage and then flees when a woman approaches lasted a minute and a half). Another study reported that in a survey of over 100 experimental psychologists, potential jurors, judges, law students, and lawyers, 95% of the experts agreed that witnesses generally overestimate the duration of crimes, whereas fewer than half of the potential jurors did so. A. Daniel Yarmey & Hazel P. Tressillian Jones, *Is the Psychology of Eyewitness Identification a Matter of Common Sense?*, *in* EVALUATING WITNESS EVIDENCE: RESEARCH AND NEW PERSPECTIVES 33 (Sally M. Lloyd-Bostock & Brian R. Clifford eds., 1983). *See also* Elizabeth F. Loftus et al., *Time Went by So Slowly: Overestimation of Event Duration by Males and Females*, 1 APPL. COGNITIVE PSYCHOL. 1 (1987).

83. In some instances, studies have shown no meaningful correlation between confidence and accuracy. *See, e.g.*, Cutler et al., *supra* note 78, at 233, 234; Kenneth A. Deffenbacher, *Eyewitness Accuracy and Confidence*, 4 LAW & HUM. BEHAV. 243, 258 (1980).

HeinOnline -- 46 Jurimetrics 198 2005-2006

exposure, the quality of the view, whether a weapon or violence was involved, and the procedures used by police to obtain the identification all help to determine whether a particular identification is an accurate one.

By contrast, confidence is a product of personality and social factors of which accuracy of observation is only a minor part. A witness' confidence will also depend on how self-confident the witness is to begin with and what interactions the witness has had with others to boost or undermine that confidence. For example, studies have shown that confidence is highly malleable and can be substantially increased by many postevent factors, including confirming feedback.[84] Witness confidence can accordingly increase after the incident through the use of procedures that do not in any way enhance the accuracy of the original identification and may undermine it. The weak correlation between confidence and accuracy that may have existed immediately after the incident is thus often destroyed after a witness' confidence level is raised or lowered through "contaminating" exposure to feedback.

The PDS survey results demonstrate that jurors do not understand the relationship between confidence and accuracy. For example, one survey question asked respondents to compare the reliability of a witness who was "absolutely certain" of an identification with that of a witness who was not. A plurality of respondents, 31%, found the "absolutely certain" witness to be "much more reliable." Moreover, only 17% of the respondents correctly understood the slight correlation between confidence and accuracy. Thus, a majority of the respondents demonstrate a fundamental misunderstanding about the confidence-accuracy correlation.[85]

Another confidence-accuracy survey question confirmed these findings. Nearly 40% of survey respondents agreed that "an eyewitness' level of confidence in his or her identification is an excellent indicator of that eyewitness' reliability." Thus, four out of ten potential jurors, absent education on this subject, would assess witness testimony under the mistaken impression that there is a very strong correlation between witness confidence and witness accuracy. Moreover, even though 55% of polled jurors correctly reject the notion that confidence is an "excellent indicator" of accuracy, the responses to the earlier question demonstrate that these jurors do not understand whether any confidence-accuracy correlation exists and, if so, what that correlation is. Thus, these survey results also make clear that jurors have no meaningful idea of how to evaluate witness statements of confidence and are likely to substantially overestimate the reliability of a confident witness.

---

84. *E.g.*, Amy L. Bradfield et al., *The Damaging Effect of Confirming Feedback on the Relation Between Eyewitness Certainty and Identification Accuracy*, 87 J. APPLIED PSYCHOL. 112 (2002).

85. Even if survey respondents who said that an identification of a confident witness is "equally reliable" as one by a less confident witness are treated as having correctly understood this phenomenon, 59% of the respondents—more than half of those on any given jury—demonstrate a fundamental misunderstanding of the confidence accuracy relationship.

HeinOnline -- 46 Jurimetrics 199 2005-2006

6. *Specific Reliability Factors—Police Officers*

Studies have shown that police officers are sometimes able to provide more detailed accounts of the event and that they are sometimes less susceptible to the effects of postevent misinformation when compared to laypersons. But research has also shown that police officers perform no better at identifying faces than other citizens,[86] which means that their eyewitness identifications are generally no more reliable than anyone else's.[87]

To test whether jurors understood this fact, the PDS survey asked respondents to compare the reliability of an eyewitness identification by "a police officer" with the reliability of an identification by "an average citizen." Sixty percent of the respondents failed to understand that the two identifications were equally reliable, and many of these respondents (22% of the total) believed police officer testimony to be "much more reliable."

7. *Specific Reliability Factors—Cross-Racial Impairment*

A variety of studies have shown that eyewitnesses experience a "cross-racial impairment" when identifying members of another race. Eyewitnesses are better at identifying members of their own race and have difficulty identifying members of other races.[88]

When asked to compare the reliability of an identification by an eyewitness "of the same race as the person being identified" with the reliability of an identification by an eyewitness "of a different race" from the suspect, almost two-thirds of jurors surveyed indicated that they are ill-informed about the inaccuracy of cross-racial identification: A large plurality of the survey respondents (48%) thought cross-race and same-race identifications are of equal reliability, and many of the other respondents either did not know or thought a cross-racial identification would be more reliable (11%). Only 36% of the survey respondents understand that a cross-racial identification may be less reliable.

8. *Specific Reliability Factors—Show-Ups*

A show-up occurs when police display a single suspect to the witness and ask whether that suspect is the culprit. Both the Supreme Court and the Attorney General have recognized the inherently suggestive nature of show-up

---

86. Elizabeth F. Loftus, *Eyewitnesses: Essential But Unreliable*, 18 PSYCHOL. TODAY 22, 24 (1984).

87. Brian Clifford, *Police as Eyewitnesses*, 36 NEW SOC'Y 176 (1976) (demonstrating that police officers perform more poorly than civilians as witnesses because police officers often have a biased interpretation of events).

88. *See generally* ELIZABETH LOFTUS, EYEWITNESS TESTIMONY § 4.9 (3d ed. 1997 & Supp. 1999); GARY L. WELLS & ELIZABETH LOFTUS, PSYCHOLOGICAL PERSPECTIVES 1 (1984); Sheri Lynn Johnson, *Cross-Racial Identification Errors in Criminal Cases*, 69 CORNELL L. REV. 934 (1984); Stephanie J. Platz & Harmon M. Hosch, *Cross-Racial/Ethnic Eyewitness Identification: A Field Study*, 18 J. APPLIED SOC. PSYCHOL. 972 (1988); John P. Rutledge, *They All Look Alike: The Inaccuracy of Cross-Racial Identifications*, 28 AM. J. CRIM. L. 207 (2001).

HeinOnline -- 46 Jurimetrics 200 2005-2006

procedures,[89] and the Supreme Court has expressly acknowledged the greater risk of misidentification involved in show-ups as compared to lineups.[90] There is strong empirical evidence that show-ups are more likely to yield false identifications than properly constructed lineups. The risk of generating an unreliable identification by using a show-up identification procedure, moreover, grows substantially more pronounced over time[91] especially if the suspect is wearing clothing similar to the culprit's.

The PDS survey indicates substantial confusion among potential jurors about the reliability of show-up procedures. While this was one of the few areas where a majority of potential jurors appear to have intuitions similar to research findings, a substantial minority of jurors still did not understand this concept. A quarter of potential jurors believed that a show-up is either *more* reliable than a lineup procedure or that the two procedures are equally reliable.

9. *Specific Reliability Factors—Lineup Instructions*

A lineup procedure occurs when a witness is shown a group of people and asked to attempt an identification. Research demonstrates that during a lineup procedure eyewitnesses tend to engage in a comparative analysis, that is, they identify the person from the lineup who, in the opinion of the eyewitness, *looks most like* the culprit relative to other members of the lineup.[92] Where the actual culprit is not in the lineup, this relative judgment process poses a grave danger of misidentification because there will always be someone who looks more like the culprit than the remaining lineup members.

Research has demonstrated that a witness' willingness to make a relative judgment about the culprit, in a lineup that does not include the culprit, is considerably less when the witness is warned that the culprit may not be present

---

89. *See* Stovall v. Denno, 388 U.S. 293, 302 (1967) ("the practice of showing suspects singly to persons for the purpose of identification, not as part of a lineup, has been widely condemned."); UNITED STATES DEPARTMENT OF JUSTICE, EYEWITNESS EVIDENCE: A TRAINER'S MANUAL FOR LAW ENFORCEMENT 30 (2003) (discussing how to reduce the inherent suggestiveness of show-up procedures).

90. *See* Moore v. Illinois, 434 U.S. 220, 229 (1977) ("Indeed, a one-on-one confrontation generally is thought to present greater risks of mistaken identification than a lineup."); Simmons v. United States, 390 U.S. 377, 383 (1968) ("[The danger of misidentification] will be increased if the police display to the witness only the picture of a single individual who generally resembles the person he saw, or if they show him the pictures of several persons among which the photograph of a single such individual recurs or is in some way emphasized.").

91. A. Daniel Yarmey et al., *Accuracy of Eyewitness Identifications in Showups and Lineups*, 20 LAW & HUM. BEHAV. 469, 459–77 (1996).

92. For studies supporting the relative judgment phenomenon, see Gary L. Wells & Eric P. Seelau, *Eyewitness Identification: Psychological Research and Legal Policy on Lineups*, 1 PSYCHOL. PUB. POL'Y. & L. 765 (1995); Gary L. Wells, *What Do We Know about Eyewitness Identification?*, 48 AM. PSYCHOL. 553 (1993); Gary L. Wells, *The Psychology of Lineup Identifications*, 14 J. APPLIED SOC. PSYCHOL. 89 (1984).

HeinOnline -- 46 Jurimetrics 201 2005-2006

in the lineup.[93] Moreover, this instruction results in the same number of accurate identifications when the culprit is actually present in the lineup.

Potential jurors do not appear to understand the role or importance of such instructions. Over half the respondents in the PDS survey (51%) thought that an identification would be *more* reliable if the eyewitness was *not* instructed about the culprit's potential absence, and an additional 21% either thought it did not matter whether such an instruction was given or were not sure. Only 28% of the respondents thought that a lineup would be more reliable if the eyewitness *was* instructed that the actual culprit "may or may not be in the lineup." These numbers were virtually the same for photographic arrays.[94] The survey thus demonstrates substantial juror confusion on the importance of lineup and photo array instructions.

10. *Specific Reliability Factors—Sequential Identification*

Another way to avoid the relative-judgment problem is the use of a sequential lineup. Unlike a traditional lineup in which an eyewitness views the lineup members simultaneously, in the sequential lineup the eyewitness views the members one at a time and decides individually whether each person is the culprit before viewing the next member. Studies have shown that this procedure reduces the number of mistaken identifications, particularly when the witness does not know how many members will be viewed and thus cannot anticipate when the process will end. The sequential procedure produces fewer mistaken identifications in lineups that do not contain the actual perpetrator, but it does not significantly impair eyewitnesses' abilities to identify the perpetrator in perpetrator-present lineups.[95]

---

93. Wells & Seelau, *supra* note 92, at 765, 772–73. *See also* CUTLER & PENROD, *supra* note 2, at 101, 115–23 (concluding that "the research shows that biased instructions substantially increase the likelihood of false identifications"); Roy S. Malpass & Patricia G. Devine, *Eyewitness Identification: Lineup Instructions and the Absence of the Offender*, 66 J. APPLIED PSYCHOL. 482, 482–89 (1981) (failure to give explicit instructions to the eyewitness that explain that the perpetrator might not be in the lineup leads subjects to select someone from the lineup regardless of whether the perpetrator is present; even when these instructions are given, eyewitnesses tend to make relative judgments.).

94. An equal number (51%) thought an identification from a photo array would be *more* reliable if the eyewitness was *not* instructed about the culprit's potential absence, while an additional 19% either thought it did not matter whether such an instruction was given or were not sure. Similarly, only 30% of the respondents correctly understood that an identification from a photo array is more reliable if the witness is instructed that the actual culprit "may or may not be in the photo array."

95. *See* R.C.L. Lindsay et al., *Sequential Lineup Presentation: Technique Matters*, 76 J. APPLIED PSYCHOL. 741, 744 (1991); Brian L. Cutler & Steven D. Penrod, *Improving the Reliability of Eyewitness Identification: Lineup Construction and Presentation*, 73 J. APPLIED PSYCHOL. 281, 288–89 (1988); R.C.L. Lindsay & Gary L. Wells, *Improving Eyewitness Identifications from Lineups: Simultaneous Verses Sequential Lineup Presentations*, 70 J. APPLIED PSYCHOL. 556, 562 (1985); *see also* Nancy M. Steblay et al., *Eyewitness Accuracy Rates in Sequential and Simultaneous Lineup Presentations: A Meta-Analytic Comparison*, 25 LAW & HUM. BEHAV. 459, 471 (2001) (meta-analysis of twenty-three papers showed that a sequential lineup, as opposed to a simultaneous lineup, reduced the rate of mistaken identification in culprit-absent lineups but was also associated with a reduction in accurate identification rates in culprit-present lineups).

HeinOnline -- 46 Jurimetrics 202 2005-2006

Potential jurors are unaware of the benefits of sequential identifications procedures. When asked to compare the reliability of a procedure in which a "witness views a lineup of potential suspects standing next to one another" with a procedure in which the witness "views potential suspects one at a time," over three-quarters of the respondents (76%) either thought the reliability of a simultaneous live lineup was better than or equal to that of the sequential lineup, or were not sure which process was better. With respect to photo lineups, the numbers are similar—61% of the respondents either thought the reliability of a simultaneous photo lineup was better than or equal to that of a sequential photo lineup or were not sure which process was better. In both cases, therefore, potential jurors seemed not to understand the importance of a sequential lineup in securing an accurate identification.

11. *Specific Reliability Factors—"Double Blind" Procedures*

The behavior of the person who administers the identification procedure can influence the reliability of eyewitness testimony itself.[96] If the person who administers the live or photographic lineup is the case detective or some other investigator who knows the identity of the suspect, a substantial danger exists that the person conducting the lineup will communicate the suspect's identity to the witness.[97] This can happen consciously or unconsciously, and through both verbal and nonverbal means. Moreover, as noted above, the administrator of the lineup can also cause mistaken eyewitnesses to develop high levels of false certainty by providing postidentification feedback about the supposed accuracy of the identification. All of these problems can be substantially reduced through the use of "double blind" procedures, that is, when lineups are administered by someone who does not know which lineup member is the suspect and which ones are fillers.

With respect to live lineups, a bare majority of potential jurors (55%) appears to have at least some grasp of the importance of conducting "a lineup where the police officer running the lineup is unaware who the suspect is," although a substantial minority of jurors (45%) do not understand the importance of this concept at all. Specifically, one fifth of the jurors incorrectly believe that a live lineup where police know the identity of the suspect is *more* reliable than a "double blind" procedure, and another 27% of respondents believe either that the two procedures are equally reliable or were unsure of the difference between the two procedures. This means that approximately 5 jurors in a panel of 12 will start

---

96. *See, e.g.*, Ronald P. Fisher, *Interviewing Victims and Witnesses of Crime*, 1 PSYCHOL., PUB. POL'Y, & L. 732, 753–58 (1995) (discussing how interviewers' expectations affect subjects' performance); Gary L. Wells & Amy L. Bradfield, *"Good, You Identified the Suspect": Feedback to Eyewitnesses Distorts Their Reports of the Witnessing Experience*, 83 J. APPLIED PSYCHOL. 360 (1998) (providing studies demonstrating confirming feedback given at the time of an initial identification made subjects significantly more confident of their identification of a suspect).

97. Gary L. Wells et al., *Eyewitness Identification Procedures: Recommendations for Lineups and Photospreads*, 22 LAW & HUM. BEHAV. 603, 627–29 (1998) (police officers aware of the perpetrator's identity can unconsciously bias an eyewitness' lineup selection).

HeinOnline -- 46 Jurimetrics 203 2005-2006

a trial with a basic misunderstanding about the importance of "double blind" procedures, a misunderstanding that will obviously continue throughout the trial and into jury deliberations in the absence of accurate, authoritative information.

This problem is even more pronounced with photographic arrays. A majority of potential jurors (52%) do not understand the importance of conducting "a photo array where the police officer running the photo array is unaware who the suspect is." Specifically, 30% of the jurors incorrectly believed that a photo array where police know the identity of the suspect is *more* reliable than a "double blind" procedure, and another 22% of respondents believed either that the two procedures are equally reliable or were unsure of the difference.

B. Summary of Survey Findings

The PDS survey documents that a substantial number of jurors come to each trial with basic misunderstandings about the way memory works in general and about specific factors that can affect the reliability of eyewitness identifications. The findings undermine previous judicial speculation about what jurors know as a "matter of common sense." As an empirical matter, the PDS poll shows that significant numbers of jurors (often substantial majorities) do not understand concepts like weapon focus, the effects of stress, the tendency of witnesses to overestimate exposure time, and the lack of meaningful correlation between witnesses' stated confidence and accuracy in making an identification. Jurors also place unwarranted stock in the eyewitness abilities of police officers, they overestimate the reliability of cross-racial identifications, and they have minimal understanding of how police procedures can affect the accuracy of an eyewitness identification.

C. Potential Limitations of Survey Findings

Any survey is limited in its ability to exactly determine the knowledge of laypersons.[98] However, if anything, it is likely that the survey methodology used in this instance *overestimates* the juror understanding of eyewitness reliability factors. First, the questions asked often singled out the relevant variables (rather than the juror having to analyze the facts of an identification and say that, for example, the presence of a weapon is significant). Second, there is a base rate of accuracy that comes from survey participants guessing among the multiple-choices, despite being given the option to say they did not know. Therefore, survey participants' correct responses to the survey questions likely overestimate the extent to which a juror would take proper account of that variable in an actual case.

The precise applicability of these survey results to other jurisdictions cannot be determined, though the pool of potential District of Columbia jurors surveyed is likely similar to many other urban jury pools nationwide. The complete results

98. *See* EYEWITNESS TESTIMONY: PSYCHOLOGICAL PERSPECTIVES (Gary L. Wells & Elizabeth F. Loftus eds. 1984).

HeinOnline -- 46 Jurimetrics 204 2005-2006

of the survey list the detailed demographic information of survey participants. Notably, the education level of participants was quite high (45% received a bachelors or higher degree) and reflected a wide range of ages and incomes.

Over the past 30 years, researchers have made substantial strides in understanding the workings and limitations of human memory. However, the application of these scientific advances to eyewitness identifications in the criminal justice system, though increasing, has been limited. Trial judges in most jurisdictions routinely exercise their discretionary powers to exclude expert testimony about the reliability of eyewitness identifications. This is so despite changes to the rules of evidence, Supreme Court opinions acknowledging defendants' rights to put on testimony in their defense, and attempts by executive and legislative branches to give greater scrutiny to eyewitness identifications.

The most common rationale offered for excluding eyewitness identification expert witnesses is that their findings are not "beyond the ken" of the average juror. To test this rationale, a survey of potential jurors in the District of Columbia was designed to investigate whether jurors understand, as a matter of common sense, what makes some eyewitness identifications more or less reliable than others. The survey results presented above demonstrate that jurors misunderstand how memory generally works and how particular factors, such as the effects of stress or the use of a weapon, affect the accuracy of eyewitness testimony. Jurors also misunderstand how eyewitnesses' stated levels of certainty correspond with accuracy. These findings demonstrate that there are flaws in judicial intuitions about what jurors do and do not understand.

Judges, prosecutors, and defense attorneys alike need to be educated about the perils of eyewitness identification. The common sense assumptions of these groups may be somewhat better than the average juror by virtue of greater experience with the criminal justice system, but even these groups lack an adequate understanding of eyewitness reliability. Defense attorneys will not introduce experts or file suppression motions, prosecutors will not discern credible allegations, and judges will not hear counsel's arguments on eyewitness reliability—unless they all reach beyond their commonsense intuitions. Tens of thousands of convictions each year are made on the basis of eyewitness evidence about which jurors' common sense is mistaken. To safeguard against wrongful convictions on eyewitness evidence, the justice system must replace false intuitions with scientifically based opinions.

HeinOnline -- 46 Jurimetrics 205 2005-2006

APPENDIX

SURVEY QUESTION AND ANSWER DATA

All results are shown as percentages unless otherwise noted. Correct answers to knowledge-based questions are in bold.

1a. What is the last grade that you completed in school?

Grade school	1
Some high school	7
High school graduate	27
Some college, no degree	13
Vocational training/2-year college	6
4-year college/bachelor's degree	22
Some postgraduate work, no degree	3
2-3 years postgraduate work/master's degree	14
Doctoral/law degree	6
Not sure/refused	1

1b. In which quadrant of Washington, D.C., do you live—Southeast, Southwest, Northeast, or Northwest?

Southeast	22
Southwest	4
Northeast	26
Northwest	48
Not sure/refused	-

2. As you may know, to be eligible for jury duty in Washington, D.C., you must be a U.S. citizen, at least eighteen years old, and a resident of the District of Columbia. You must read, speak, and understand English, and you must NOT be on probation or parole. You DO NOT have to be registered to vote. Given these requirements, are you currently qualified for jury duty in Washington, D.C.?

Yes, qualified to serve on a jury	100	CONTINUE
No, not qualified to serve on a jury	-	**TERMINATE**
Not sure	-	

HeinOnline -- 46 Jurimetrics 206 2005-2006

3. For each of the following phrases, please tell me whether it applies to you very well, fairly well, just somewhat well, or not very well.

	Applies Very Well	Applies Fairly Well	Applies Just Somewhat Well	Does Not Apply Very Well	Not Sure
I never forget a face	38	28	24	9	1
I have an excellent memory	44	33	17	6	-

4a. In some criminal trials, an eyewitness may testify that the person on trial, who is also called the defendant, was the person the eyewitness saw committing the crime. In general, do you consider this sort of eyewitness identification to be a very reliable, somewhat reliable, not very reliable, or unreliable form of evidence?

Very reliable	18
Somewhat reliable	65
Not very reliable	11
Unreliable	4
Not sure	2

4b. And suppose that an eyewitness has no motivation to lie, and genuinely believes that his or her identification of the defendant is accurate. In those cases, do you consider eyewitness identifications in criminal trials to be very reliable, somewhat reliable, not very reliable, or unreliable forms of evidence?

Very reliable	25
Somewhat reliable	62
Not very reliable	8
Unreliable	3
Not sure	2

HeinOnline -- 46 Jurimetrics 207 2005-2006

5. Now I'm going to describe several pairs of eyewitnesses who might testify in a criminal trial. I'd like you to tell me whether you would consider the first or second eyewitness to be the more reliable ("MR") eyewitness, or whether you would consider both eyewitnesses to be equally reliable ("ER"), or whether neither would be reliable ("N"). **(If necessary ask:)** Would the (first/second) eyewitness be "much more" reliable ("MMR") or only "slightly more" reliable ("SMR")?

	The First Eyewitness		**The Second Eyewitness**		**ER**	**N**	**Not Sure**
	MMR	SMR	SMR	MMR			
5a. One eyewitness says they are absolutely certain of their identification of the criminal defendant, AND another eyewitness does NOT say they are absolutely certain of their identification of the criminal defendant.							
	31	**17**	12	5	26	4	5
5b. One eyewitness is a police officer, AND another eyewitness is an average citizen.							
	22	15	7	9	**40**	3	4
5c. One eyewitness is shown a group of people including the defendant and identifies the defendant as the culprit AND another eyewitness is shown only one person, the defendant, and the eyewitness identifies the defendant as the culprit.							
	50	18	5	4	18	2	3
5d. One eyewitness is of the same race as the person being identified as the culprit, AND another eyewitness is of a different race from the person being identified as the culprit.							
	21	**15**	4	3	48	5	4

6a. Eyewitnesses often are asked to estimate how much time elapsed during the commission of the crime. Do you think eyewitness estimations of the duration of crimes are usually accurate or usually NOT accurate?

Usually accurate	29	**Skip to Q.7**
Usually not accurate	**58**	**CONTINUE**
Not sure	13	**Skip to Q.7**

(Ask only of respondents who say eyewitness estimates of crime duration are usually not accurate in Question 6a.)

6b. What do you think happens more often—that an eyewitness UNDERESTIMATES the time that elapsed during the commission of the crime and says that the event took LESS time than it actually did, or that an eyewitness OVERESTIMATES the time that elapsed during the commission of the crime and says that the event took MORE time than it actually did?

More likely that an eyewitness underestimates the time	14
More likely that an eyewitness overestimates the time	37
Not sure	7
Respondents who say eyewitness estimates of crime duration are usually accurate/not sure (Q.6a)	42

HeinOnline -- 46 Jurimetrics 208 2005-2006

7. In situations in which **(READ ITEM)**, do you think this makes an eyewitness' memory about the details of the crime more reliable, less reliable, or has no effect on the eyewitness' memory about the details of the crime?

	More Reliable	Less Reliable	No Effect On Reliability	Not Sure
A weapon is involved in a crime	37	**30**	25	8
A crime is violent	39	**30**	22	9

8. Sometimes police talk to an eyewitness and generate a sketch of the culprit based on that eyewitness' description. Do you think such a sketch is a reliable tool in identifying the culprit or not?

Yes, sketch is a reliable tool	65
No, sketch is not a reliable tool	**25**
Not sure	10

9. Do you think an identification made by an eyewitness who participated in the creation of such a sketch makes that eyewitness' identification more reliable, less reliable, or just as reliable as an identification provided by an eyewitness who did NOT participate in the creation of such a sketch?

More reliable	34
Less reliable	**12**
Just as reliable	**48**
Not sure	6

HeinOnline -- 46 Jurimetrics 209 2005-2006

Ask Questions 10a and 10b to one-half of the respondents respectively.

10a. There are different ways of conducting identification procedures. One type of identification procedure involves showing an eyewitness a lineup, which may involve a number of individuals standing in a line facing the eyewitness. For each of the following pairs of lineup procedures that I describe to you, procedure **A** and procedure **B**, please tell me which one you think is more reliable, or whether you think both are equally reliable.

	A is more reliable	**B** is more reliable	**A & B** are equally reliable	Not Sure
Procedure A: A lineup where the eyewitness is instructed that the criminal suspect may or may not be included in the lineup, OR **Procedure B:** A lineup where the eyewitness is NOT instructed that the criminal suspect may or may not be included in the lineup.				
	28	51	15	6
Procedure A: A witness views a lineup of potential suspects standing next to one another, OR **Procedure B:** A witness views potential suspects one at a time.				
	53	**24**	16	7
Procedure A: A lineup where the police officer running the lineup is AWARE of who the suspect is, OR **Procedure B:** A lineup where the police officer running the lineup is UNAWARE of who the suspect is.				
	18	**55**	21	6

10b. There are different ways of conducting identification procedures. One type of identification procedure involves showing the eyewitness photographs, which is usually called a photo array. For each of the following pairs of photo array procedures that I describe to you, procedure **A** and procedure **B**, please tell me which one you think is more reliable, or whether you think both are equally reliable.

	A is more reliable	**B** is more reliable	**A & B** are equally reliable	Not Sure
Procedure A: A photo array where the eyewitness is instructed that the criminal suspect may or may not be included in the lineup, OR **Procedure B:** A photo array where the eyewitness is NOT instructed that the criminal suspect may or may not be included in the lineup.				
	30	51	16	3
Procedure A: A witness views a photo array of potential suspects all at the same time, OR **Procedure B:** A witness views a photo array of potential suspects one picture at a time.				
	38	**39**	19	4
Procedure A: A photo array where the police officer running the photo array is AWARE of who the suspect is, OR **Procedure B:** A photo array where the police officer running the photo array is UNAWARE of who the suspect is.				
	30	**48**	18	4

HeinOnline -- 46 Jurimetrics 210 2005-2006

11. Now I am going to read you a series of statements and I'd like you to tell me whether you believe each one is true or false.

	T	F	Not Sure
11a. An eyewitness' level of confidence in his or her identification is an excellent indicator of that eyewitness' reliability.	39	**55**	6
11b. The act of remembering a traumatic event is like a video recording in that one can recall details as if they had been imprinted or burned into one's brain.	46	**48**	6
11c. Once an eyewitness learns from police that the person they identified as the culprit is the suspect the police believe committed the crime, that eyewitness is more likely to express greater confidence in their identification than they did beforehand.	**85**	11	4
11d. An eyewitness who identifies the same culprit in a number of identification procedures can still be mistaken.	**89**	9	2
11e. If more than two eyewitnesses identified the defendant as the culprit, it always means the eyewitnesses picked the right person.	19	**79**	2
11f. Eyewitnesses can believe they remember details about a crime that they actually learned about later from someone else, such as the police.	**80**	16	4
11g. Eyewitnesses will sometimes identify a person as the culprit because they have seen that person somewhere before and the face is familiar, even though the person was not who they actually saw committing the crime.	**73**	21	6
11h. Before an eyewitness identifies a defendant at trial, if they learn that someone else has also identified the defendant as the culprit, that eyewitness is more likely to express greater confidence in their identification when they testify in front of the jury.	**86**	10	4
11i. Generally, eyewitnesses are equally accurate when identifying someone of a different race as when they are identifying someone of their own race.	27	**66**	7
11j. If an eyewitness was under high stress at the time of the crime, the eyewitness will have better recall for the details of the event.	14	**80**	6

HeinOnline -- 46 Jurimetrics 211 2005-2006

Factual Questions:

F1. What is your age? I am going to read you some age categories. Stop me when we get to your category.

18 – 24	11
25 – 29	9
30 – 34	13
35 – 39	9
40 – 44	12
45 – 49	6
50 – 54	8
55 – 59	8
60 – 64	7
65 – 69	6
70 – 74	5
75 and over	4
Refused	2

F2. Are you currently employed? **(If "currently employed," ask:)** What type of work do you do? **(If "not currently employed," ask:)** Are you a student, a homemaker, retired, or unemployed and looking for work?

Currently Employed	
Professional, manager	29
White-collar worker	25
Blue-collar worker	10
Farmer, rancher	-
Not Currently Employed	
Student	5
Homemaker	3
Retired	20
Unemployed, looking for work	6
Other	-
Not sure	2

HeinOnline -- 46 Jurimetrics 212 2005-2006

F5. Generally speaking, do you think of yourself as a Republican, a Democrat, an independent, or something else? **(If "Republican" or "Democrat," ask:)** Do you consider yourself a strong (Republican/Democrat) or a not so strong (Republican/Democrat)? **(If "Independent," ask:)** Would you say that you lean more toward the Republicans or more toward the Democrats?"

Strong Republican	4
Not so strong Republican	3
Independent, leans Republican	2
Independent	15
Independent, leans Democrat	7
Not so strong Democrat	12
Strong Democrat	47
Something else/Other	7
Not sure	3

F6. Have you ever been a juror in a CRIMINAL trial? **(If "Yes," ask:)** Was that trial in Washington, D.C.?

Yes, have been a criminal juror in Washington, D.C.	37
Yes, have been a criminal juror outside Washington, D.C.	4
No, have not been a criminal juror	58
Not sure/refused	1

F7. Are you from a Hispanic or Spanish-speaking background?

Yes, Hispanic	5
No, not Hispanic	93
Not sure/refused	2

F8. What is your race—white, black, Asian, or something else?

White	32
Black	61
Asian	2
Other	1
Hispanic (VOL)	2
Not sure/refused	2

HeinOnline -- 46 Jurimetrics 213 2005-2006

F9. Last year, what was your total family income from all sources, before taxes? Just stop me when I get to the right category.

Less than $20,000	12
Between $20,000 and $30,000	14
Between $30,000 and $40,000	11
Between $40,000 and $50,000	9
Between $50,000 and $75,000	13
Between $75,000 and $100,000	8
More than $100,000	13
Not sure/refused	20

F10. Have you or a close family member ever been the victim of a crime? **(If "Yes," ask:)** Was the crime a violent or a nonviolent crime?

Yes, have been the victim of a violent crime	27
Yes, have been the victim of a nonviolent crime	25
Yes, have been the victim of both a violent and a nonviolent crime (VOL)	6
No, have not been the victim of a crime	40
Not sure/refused	2

F11. Have you or a close personal friend or a family member ever been arrested? **(If "Yes," ask:)** And was that you, your friend, or your family member?

Yes, respondent has been arrested	12
Yes, close personal friend has been arrested	12
Yes, family member has been arrested	23
No, have not been arrested	56
Not sure/refused	5

HeinOnline -- 46 Jurimetrics 214 2005-2006

On the "General Acceptance" of Eyewitness Testimony Research

A New Survey of the Experts

Saul M. Kassin — *Williams College*
V. Anne Tubb and Harmon M. Hosch — *University of Texas at El Paso*
Amina Memon — *University of Aberdeen*

In light of recent advances, this study updated a prior survey of eyewitness experts (S. M. Kassin, P. C. Ellsworth, & V. L. Smith, 1989). Sixty-four psychologists were asked about their courtroom experiences and opinions on 30 eyewitness phenomena. By an agreement rate of at least 80%, there was a strong consensus that the following phenomena are sufficiently reliable to present in court: the wording of questions, lineup instructions, confidence malleability, mug-shot-induced bias, postevent information, child witness suggestibility, attitudes and expectations, hypnotic suggestibility, alcoholic intoxication, the cross-race bias, weapon focus, the accuracy–confidence correlation, the forgetting curve, exposure time, presentation format, and unconscious transference. Results also indicate that these experts set high standards before agreeing to testify. Despite limitations, these results should help to shape expert testimony so that it more accurately represents opinions in the scientific community.

In recent years, and with increasing frequency, psychologists have served as expert witnesses in trials that contain possible erroneous eyewitness identifications. To assess the extent to which there is "general acceptance" of various eyewitness phenomena within the scientific community (a criterion for the admissibility of scientific evidence, initially enunciated in *Frye v. United States*, 1923), Kassin, Ellsworth, and Smith (1989) surveyed 63 eyewitness experts for their views on the reliability of 19 propositions. Their survey revealed that whereas certain research findings were judged to be reliable by most experts (e.g., the effects of exposure time, lineup instructions, the wording of questions, preevent expectations, postevent information, and the accuracy–confidence correlation), others did not elicit high levels of consensus (e.g., the effects of stress, event violence, gender, weapon focus, hypnotic retrieval techniques, and training in eyewitness observation).

By providing empirical evidence of the consensus within the community of experts, this survey has proved useful to judges ruling on the admissibility of expert witnesses; psychologists needing to determine the appropriate contents of their testimony; and cross-examiners seeking to discredit experts who overstate, understate, or in other ways misrepresent the literature. In light of the kinds of substantive disputes likely to erupt in the courtroom (Leippe, 1995; Penrod, Fulero, & Cutler, 1995), the net effect, it was hoped, was to encourage expert testimony that more accurately reflects the consensus of opinions within the scientific community.

The time has come for Kassin et al.'s (1989) survey of experts to be updated. Since its publication, there has been a surge of new and important research developments in the area (for reviews, see Cutler & Penrod, 1995; Ross, Read, & Toglia, 1994; Sporer, Malpass, & Koehnnken, 1996; Thompson et al., 1998; Wells, 1993). There has also been a surge of activity directed at the application of eyewitness research findings within the legal system. In 1996, the National Institute of Justice reported on 28 wrongful convictions, cases in which convicted felons were exonerated by DNA evidence after varying numbers of years in prison. Remarkably, all of these cases contained one or more false identifications (Connors, Lundregan, Miller, & McEwen, 1996). More recently, the number of DNA exoneration cases had climbed to 62 (including 8 individuals who had been sentenced to death)—52 of which contained identifications from 77 confident but mistaken eyewitnesses (Scheck, Neufeld, & Dwyer, 2000). In response to these newly documented miscarriages of justice, a group of eyewitness researchers—as part of an initiative of the American Psychology–Law Society (Division 41 of the American

Editor's note. David Faigman served as action editor for this article.

Author's note. Saul M. Kassin, Department of Psychology, Williams College; V. Anne Tubb and Harmon M. Hosch, Department of Psychology, University of Texas at El Paso; Amina Memon, Department of Psychology, University of Aberdeen, Aberdeen, Scotland.

We thank LavEarl Moore for help in compiling lists of experts and posting questionnaires and the University of Texas at El Paso Research Incentive Program for supporting this research.

Correspondence concerning this article should be addressed to Saul M. Kassin, Department of Psychology, Williams College, Bronfman Science Center, Williamstown, MA 01267. Electronic mail may be sent to skassin@williams.edu.

Copyright 2001 by the American Psychological Association, Inc. 0003-066X/01/$5.00
Vol. 56, No. 5, 405–416 DOI: 10.1037//0003-066X.56.5.405

Psychological Association)—wrote a scientific review article that contained specific procedural rules for how to minimize errors made in the collection of eyewitness identifications (Wells et al., 1998). In 1999, the National Institute of Justice—at the request of former Attorney General Janet Reno—assembled a multidisciplinary working group of police, prosecutors, defense attorneys, and research psychologists, who published a "how to" manual, the first of its kind, entitled *Eyewitness Evidence: A Guide for Law Enforcement* (for a discussion of how these guidelines were developed, see Wells et al., 2000).

In addition to recent activity in the eyewitness arena, there have also been significant changes in the legal criteria by which judges admit or exclude expert testimony at trial. At the time the 1989 survey was conducted, most courts had relied heavily on the *Frye* test—that scientific testimony is admissible only if it is based on a theory or research finding that has "general acceptance in the particular field in which it belongs" (*Frye v. United States*, 1923, p. 1014). This criterion—which was specifically applied to eyewitness research in *United States v. Amaral* (1973)—provided the motivational impetus for the original survey of experts.

Since that time, the evidentiary landscape has changed in significant ways. In *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993), the U.S. Supreme Court urged trial judges to serve as more active gatekeepers by ascertaining whether an expert proffers testimony that is scientific and will assist the trier of fact. The Court thus shifted from the prior emphasis on general acceptance to the broader question of whether the testimony would be based on information that is not only relevant but reliable and valid—and obtained through sound scientific methods.[1] The Court then extended this opinion in *Kumho Tire Co., Ltd. v. Carmichael et al.* (1999), in which it stated that *Daubert*'s basic principles and the gatekeeping function assigned to trial judges may also be applied, albeit flexibly, to engineers and other nonscientific experts who proffer "technical" or "other specialized knowledge."

The *Daubert* ruling applies only to federal courts. Indeed, although many states followed suit, some did so only in part, and still others retained their existing standards—with 17 states continuing to use the *Frye* test (Berger, 2000; Faigman, Kaye, Saks, & Sanders, 1997). In addition, as the majority opinion in *Daubert* makes clear, trial judges may still consider opinions within the community of experts in determining the admissibility of expert testimony ("general acceptance can yet have a bearing on the inquiry"; *Daubert*, 1993, p. 2797). Thus, as Leippe (1995) noted, Kassin et al.'s (1989) survey can have an important bearing on the extent to which different propositions about eyewitness testimony, on an item-by-item basis, are deemed reliable enough to be presented at trial. In *Daubert*, the U.S. Supreme Court majority asserted with confidence "that federal judges possess the capacity to undertake this review" (p. 2796). Lacking the training and experience necessary to evaluate psychological theories, research methods, error rates, and even the quality of peer-reviewed journals, judges would thus benefit greatly from the opinions of experts, whose collective wisdom presumably takes such considerations into account.

The present study was conducted with three goals in mind. First, we sought to update expert opinions on the classic eyewitness propositions tested in the original survey. In discussing the forensic usefulness of their results, Kassin et al. (1989) noted that as research accumulates, and as the effects of known factors are modified, expert opinion is likely to change. Their survey thus needs to be updated periodically to account for new research developments. To illustrate the point, Kassin et al. (1989) noted that when their data were collected, there was little direct evidence for the proposition that "the presence of a weapon impairs a witness's ability to accurately identify the perpetrator's face." No doubt aware of this issue, their experts split about 50–50 on the reliability of this statement. Yet, shortly thereafter, this weapon focus effect was demonstrated in several independent studies (see studies analyzed in Steblay, 1992).

Our second objective was to assess expert opinions on relatively new phenomena not previously tested. After the initial survey, for example, new studies were conducted on such topics as child witnesses (Ceci & Bruck, 1995; Poole & Lamb, 1998), repressed and/or false memories of trauma (Loftus, 1993; Pezdek & Banks, 1996; Read & Lindsay, 1997), the effects of alcohol (Yuille & Tollestrup, 1990), the processes by which eyewitnesses make identifications (Dunning & Stern, 1994; Sporer, 1993), sequential versus simultaneous presentations of photographic arrays and lineups (R. C. L. Lindsay, Lea, & Fulford, 1991; Wells, 1993), the malleability of confidence and other retrospective reports of the eyewitnessing experience (Luus & Wells, 1994; Shaw, 1996; Wells & Bradfield, 1998, 1999), factors that moderate the correlation of accuracy and confidence (Kassin, Rigby, & Castillo, 1991; D. S. Lindsay, Read, & Sharma, 1999; Penrod & Cutler, 1995; Robinson & Johnson, 1999; Sporer, Penrod, Read, & Cutler, 1995), and the commonsense assumptions about eyewitnesses held by laypersons and members of the legal profession (Devenport, Penrod, & Cutler, 1997; Kassin & Barndollar, 1992; Stinson, Devenport, Cutler, & Kravitz, 1996, 1997).

Our third goal was to reestimate the level of expert witness activity occurring in today's criminal and civil courtrooms. To compare these data with those obtained in the 1989 survey, we posed the same set of questions. Thus, we asked respondents to indicate separately how often they had been asked to testify, had agreed to testify, and had actually testified at trial as an eyewitness expert for the plaintiff or the defense. In an open-ended question, we also asked respondents to indicate whether they had ever declined a request to testify and the reasons for that decision.

[1] In *Daubert*, the Supreme Court offered four questions as potentially helpful—though not necessary or definitive—to this inquiry: (a) whether the theories or methods used by the expert to formulate an opinion can be tested, (b) whether they have been subjected to peer review, (c) whether there is a measurable error rate, and (d) whether the theories or methods are generally accepted within the expert's community.

In light of the criteria articulated in *Daubert*, we also asked respondents to indicate for each item whether their opinion was based on published, peer-reviewed, scientific research.

Method

The Experts

The following sources and resources were used to generate a list of experts: the membership rosters of Division 41 of the American Psychological Association, also known as the American Psychology–Law Society; the Society of Applied Research on Memory and Cognition; and the attendee lists of the 1995 and 1997 European Association of Psychology and Law biennial meetings. We identified members of these associations who had conducted eyewitness research. This list was then supplemented by a PsycINFO search for individuals who had published an article, book chapter, or other paper on eyewitness identifications during the previous 10 years. Finally, the names of eyewitness experts were solicited from subscribers to the PSYCHLAW listserve, an electronic communication network.

Questionnaires were mailed to a total of 197 prospective respondents. Of these respondents, 53% were employed in the United States. The others were from the United Kingdom, Canada, Germany, Australia, the Netherlands, Spain, New Zealand, Israel, Italy, Sweden, Denmark, and France. Six individuals returned the questionnaire, saying they did not have time to complete it or no longer considered themselves to be experts. In 5 other instances, the envelopes were sent to the wrong name, to an incorrect address, or to individuals who had died. From the remaining population of 186 prospective participants, 64 returned data in usable form, yielding a 34% response rate. With regard to their areas of specialization, 34 respondents described their primary area as cognitive psychology, 17 as personality/social, 6 as child/developmental, and 3 as clinical/counseling. Four respondents identified themselves by combinations of two or more specific areas.

For confidentiality purposes, we did not require respondents to identify themselves on their questionnaires, so it was not possible to precisely assess the ways in which our sample resembled and differed from the population of 186 eligible participants. We do know that 62 of the 64 respondents had a Ph.D. in psychology (4 had also earned a J.D.; 1 had only a D.S., and another had only an M.A.) as did all but 3 members of the total population. We also compared the numbers of eyewitness publications self-reported by respondents with the actual publication numbers for the list of all prospective participants, as derived from PsycINFO, and we found that respondents constituted a highly prolific subgroup ($M = 17.98$) of the total population ($M = 7.92$), $Z = 3.66$, $p < .001$. It appears that the experts in our study could be described as a blue-ribbon group of leading researchers. The specific magnitude of the difference cannot be known, however, because the respondent data were based on self-reports—not on actual publication numbers.

The Questionnaire

Our questionnaire contained 30 statements concerning the accuracy of eyewitness testimony. Seventeen of the 21 items from the 1989 survey were retained. To minimize the length of our instrument, 2 of the original items (pertaining to sex differences and the tendency to overestimate event durations) were dropped, and the 2 items on cross-race identifications were combined into a single generally stated principle (in the original instrument, cross-race effects on Black and White witnesses were separately assessed). In light of the research literatures and controversies that developed in the 1990s, 13 new items were added. These items addressed simultaneous versus sequential lineups, prelineup exposure to mug shots, child witnesses, the diagnostic value of identification reaction times, the repression of traumatic experiences, and the recovery of repressed memories. The complete set of 30 statements is presented in Table 1.

Five questions accompanied each statement. First, respondents were asked to characterize the reliability of the proposition described. As in the original survey, there were seven response alternatives: (a) "the evidence suggests the reverse is probably true," (b) "the evidence does not support it," (c) "the evidence is inconclusive," (d) "the evidence tends to favor it," (e) "the evidence is generally reliable," (f) "the evidence is very reliable," or (g) "I don't know." Next, respondents were asked to indicate, yes or no, whether they thought the phenomenon "was reliable enough for psychologists to present it in courtroom testimony" The third question asked, yes or no, "Under the right circumstances, would *you* be willing to testify in court that this phenomenon is reliable?" Fourth, we asked respondents to reflect on whether their opinion on the issue was "based on published, peer reviewed, scientific research." Fifth, we asked respondents for their opinion on whether "most jurors believe this statement to be true as a matter of common sense."

Following the specific questions on the 30 eyewitness phenomena, respondents were asked several questions concerning their personal backgrounds and experiences. Specifically, we asked about their professional credentials (degrees obtained, primary areas of specialization, memberships in the American Psychological Association and the American Psychological Society), relevant scholarly achievements (number of eyewitness publications in scientific journals, law reviews, books, chapters, and magazines or newsletters), and courtroom experience (the estimated number of times they were asked to testify, the number of times they agreed to testify, and the number of times they actually testified). For this last set of questions, we asked respondents to specify how often they had been called on criminal or civil cases and whether it was by the prosecution (plaintiff) or the defense. In parallel with Kassin et al.'s (1989) survey, two more general questions were then asked: (a) "What do you see as the primary role of the eyewitness expert: to educate the jury, assist a particular party, or other (please specify)?" and (b) "In general, would

Table 1
Eyewitness Topics and Statements

Topic	Statement
1. Stress	Very high levels of stress impair the accuracy of eyewitness testimony.
2. Weapon focus	The presence of a weapon impairs an eyewitness's ability to accurately identify the perpetrator's face.
3. Showups	The use of a one-person showup instead of a full lineup increases the risk of misidentification.
4. Lineup fairness	The more members of a lineup resemble the suspect, the higher is the likelihood that identification of the suspect is accurate.
5. Lineup instructions	Police instructions can affect an eyewitness's willingness to make an identification.
6. Exposure time	The less time an eyewitness has to observe an event, the less well he or she will remember it.
7. Forgetting curve	The rate of memory loss for an event is greatest right after the event and then levels off over time.
8. Accuracy–confidence	An eyewitness's confidence is not a good predictor of his or her identification accuracy.
9. Postevent information	Eyewitness testimony about an event often reflects not only what they actually saw but information they obtained later on.
10. Color perception	Judgments of color made under monochromatic light (e.g., an orange streetlight) are highly unreliable.
11. Wording of questions	An eyewitness's testimony about an event can be affected by how the questions put to that witness are worded.
12. Unconscious transference	Eyewitnesses sometimes identify as a culprit someone they have seen in another situation or context.
13. Trained observers	Police officers and other trained observers are no more accurate as eyewitnesses than is the average person.
14. Hypnotic accuracy	Hypnosis increases the accuracy of an eyewitness's reported memory.
15. Hypnotic suggestibility	Hypnosis increases suggestibility to leading and misleading questions.
16. Attitudes and expectations	An eyewitness's perception and memory for an event may be affected by his or her attitudes and expectations.
17. Event violence	Eyewitnesses have more difficulty remembering violent than nonviolent events.
18. Cross-race bias	Eyewitnesses are more accurate when identifying members of their own race than members of other races.
19. Confidence malleability	An eyewitness's confidence can be influenced by factors that are unrelated to identification accuracy.
20. Alcoholic intoxication	Alcoholic intoxication impairs an eyewitness's later ability to recall persons and events.
21. Mugshot-induced bias	Exposure to mug shots of a suspect increases the likelihood that the witness will later choose that suspect in a lineup.
22. Long-term repression	Traumatic experiences can be repressed for many years and then recovered.
23. False childhood memories	Memories people recover from their own childhood are often false or distorted in some way.
24. Discriminability	It is possible to reliably differentiate between true and false memories.
25. Child witness accuracy	Young children are less accurate as witnesses than are adults.
26. Child suggestibility	Young children are more vulnerable than adults to interviewer suggestion, peer pressures, and other social influences.
27. Description-matched lineup	The more that members of a lineup resemble a witness's description of the culprit, the more accurate an identification of the suspect is likely to be.
28. Presentation format	Witnesses are more likely to misidentify someone by making a relative judgment when presented with a simultaneous (as opposed to sequential) lineup.
29. Elderly witnesses	Elderly eyewitnesses are less accurate than are younger adults.
30. Identification speed	The more quickly a witness makes an identification upon seeing the lineup, the more accurate he or she is likely to be.

Note. The first 18 items were retained from the original survey. In the present instrument, Item 14 was stated positively rather than negatively, and Item 18 combined initially separate items for Black and White witnesses.

you say that juries are better equipped to evaluate eyewitness testimony with or without the aid of a competent expert (or is there no difference)?" Finally, we asked respondents to list any eyewitness topics about which they had testified that were not covered in the questionnaire and if they had ever declined to testify, to list the reasons for that decision.

Results and Discussion

The Experts

As we noted earlier, the experts in our sample were a prolific group with regard to research productivity. Ninety-two percent had published one or more books, chapters, or articles on the psychology of eyewitness identification. On average, respondents had authored or edited 2.15 books, 6.54 chapters, 13.22 scientific journal articles, 1.42 law review articles, and 5.38 magazine or newsletter articles.

In addition to being active researchers and writers, many of our respondents were also actively involved in the judicial system. Seventy-eight percent had been asked to testify as eyewitness experts on at least one occasion—and an estimated average of 33.38 times. In those cases, they said they agreed to testify an estimated 78% of the time. And in cases in which they did agree, they actually testified an estimated 92% of the time. In an estimated 29% of these latter cases, our respondents were countered in court by an opposing expert.

Consistent with the results of the 1989 survey, the requests for expert testimony were not equally distributed across parties in criminal and civil proceedings. As presented in Table 2, our experts were asked to testify more often in criminal cases (n = 3,150; 93% of all requests) than in civil cases (n = 220; 7% of requests). By far, the lion's share of requests for expert assistance were from criminal defendants (n = 3,016; 89% of all requests).

It is striking to compare the number of experts who agreed to testify with those who did so in the 1989 survey. In absolute numbers, our respondents were called on to testify more frequently than in the past (Ns = 3,369 and 1,268, respectively), but they were less likely to agree to do so (the agreement rates were 74% in 1989 and 41% in the present survey, $Z = 16.89$, $p < .001$). When they did agree to testify, a significantly greater percentage actually did so than was the case in the previous survey (70% and 51%, respectively), $Z = 6.97$, $p < .001$.

Consistent with the earlier finding that eyewitness experts agreed to become involved more often in civil than criminal cases (the 1989 agreement rates were 91% and 71%, respectively), we found that the agreement rate was higher (76%) when experts were sought by civil plaintiffs than by civil defendants, criminal prosecutors, and criminal defendants (43%, 48%, and 40%, respectively; ps $< .001$). In contradiction to the charge that eyewitness experts are liberally biased toward criminal defendants, it is important to note that respondents were just as likely to agree to assist the prosecution as they were the defense in criminal cases, $Z = 0.47$, *ns*.

There are two possible explanations for the finding that experts are less agreeable today than in the recent past. One is that today's experts testify as often in absolute terms but that they agree proportionally fewer times because they receive so many more requests—about three times more per expert. A second possibility is rooted in the fact that our experts were more active and productive researchers than those in Kassin et al.'s (1989) sample, and so perhaps they had less time or inclination to become involved as consultants in actual cases. Of the 49 experts who had been asked to testify at some time, all of them reported that they had declined a request for expert testimony on at least one occasion. There were various reasons for this decision. Nineteen respondents cited moral or ethical concerns (e.g., "I could not say what the attorney wanted me to say") and the personal belief that the defendant was guilty of the crime charged. Ten said they had declined to testify be-

Table 2
Estimated Number of Times Respondents Were Asked to Testify, Agreed to Testify, Actually Testified, and Were Opposed in Court

	Criminal				Civil					
	Prosecution		Defense		Plaintiff		Defense		Total	
Action	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Asked to testify	134		3,016		62		158		3,370	
Agreed to testify	65	48[a]	1,193	40[a]	47	76[a]	68	43[a]	1,373	41[a]
Actually testified	56	86[b]	837	70[b]	27	57[b]	40	44[b]	960	70[b]
Opposed	25	45[c]	30	4[c]	9	33[c]	12	30[c]	76	8[c]
Total yield		42[d]		28[d]		44[d]		25[d]		28[d]

[a] Agreement rate (i.e., percentage of times experts agreed to testify when asked). [b] Percentage of experts who after agreeing to testify actually did so. [c] Percentage of experts whose testimony was opposed in court. [d] Percentage of experts initially asked who ultimately testified.

cause they were busy or were pressed for time. Eight declined because they did not feel competent to address the specific phenomenon pertinent to the case (many of these requests were for testimony on the effects of hypnosis on the recovery of repressed memories). Other reasons for not testifying were that the phenomenon in question was not important relative to other evidence (e.g., the defendant's fingerprints) or was not sufficiently reliable or that the respondent did not think that expert testimony would assist the trier of fact.

To examine individual differences in generalized assessments of eyewitness phenomena, we next compared respondents who were (a) high versus low in research productivity and (b) high versus low in courtroom experience (both categorizations were derived by median split). For each respondent, we summed across the 30 eyewitness items the number of affirmative responses they gave to the critical question, "Do you think that this phenomenon is reliable enough for psychologists to present in courtroom testimony?" A two-way analysis of variance (Research Productivity × Courtroom Experience) on these 0–30 scores revealed a nonsignificant tendency for respondents with a high number of publications to judge more phenomena as "reliable enough" compared with those with a lower number of publications, $F(1, 54) = 3.03$, $p < .10$ (Ms = 20.86 and 18.28, respectively). There was no significant difference on this measure as a function of courtroom experience, $F(1, 54) < 1$ (Ms = 20.03 and 18.96 for high- and low-frequency experts, respectively), nor was there a significant interaction between the two factors, $F(1, 54) < 1$.

A similar two-way analysis of variance was conducted on the critical question, "Would *you* be willing to testify that this phenomenon is reliable?" Again, for each respondent, we summed across all items the number of affirmative responses, yielding scores that ranged from 0 to 30. On this measure, a significant and interesting main effect for research productivity indicated that high-publication respondents were willing to testify about more items than were low-publication respondents (Ms = 18.08 and 12.99, respectively), $F(1, 54) = 4.87$, $p < .05$. There were no significant effects for courtroom experience, $F(1, 54) = 1.45$, $p < .25$, and no interaction, $F(1, 54) < 1$.

Judgments of Eyewitness Phenomena

For each of the 30 eyewitness propositions, we sought to assess (a) how the experts as a group characterized the reliability of the phenomenon; (b) whether they saw the phenomenon as sufficiently reliable to present in court; (c) whether they would personally be willing to testify about it; (d) whether they saw their opinion on reliability as being based on published, peer-reviewed, scientific research; and (e) whether they thought that jurors were aware of the phenomenon as a matter of common sense. Table 3 presents the complete distribution of responses on the reliability question for each of the propositions. More important from a practical standpoint, Table 4 shows for each item the percentage of experts who answered "yes" to the "reliable enough," "would you testify," "research basis," and "juror commonsense" questions.

Comparisons with 1989 survey. For the 17 propositions that were retested, we compared the reliability assessments of our respondents with those initially provided by Kassin et al.'s (1989) experts. The most striking aspect of these comparisons was the degree of consistency in the two samples. As shown in Table 5, there were no significant shifts in most of the original items, and for some, the numbers were virtually identical. Thus, as before, most experts saw as sufficiently reliable expert testimony on the wording of questions (98%), lineup instructions (98%), attitudes and expectations (92%), the accuracy–confidence correlation (87%), the forgetting curve (83%), exposure time (81%), and unconscious transference (81%). Also as before, there was less, if any, consensus on the effects of color perception in monochromatic light (63%), observer training (61%), high levels of stress (60%), the accuracy of hypnotically refreshed testimony (45%), and event violence (37%).

Our experts saw two phenomena as significantly more reliable than did those who took part in the initial survey. These increases were obtained for the weapon focus effect (87% and 57%), $\chi^2(1, N = 112) = 13.61$, $p < .001$, and hypnotic suggestibility effects (91% and 69%), $\chi^2(1, N = 118) = 9.67$, $p < .005$. The increased levels of acceptance for these latter propositions make sense in light of the substantial bodies of research in these areas over the ensuing decade. For example, Steblay (1992) conducted a meta-analytic review of the weapon focus effect and concluded that it is reliable across a wide range of settings—a conclusion that has been further corroborated in more recent studies (e.g., Pickel, 1999). In a similar manner, a great deal of recent theory and research has shed light on hypnosis in general (Kirsch & Lynn, 1995) and on the risk of suggestibility effects on reported memories (Malinoski & Lynn, 1999; McConkey & Sheehan, 1995). Thus, as noted elsewhere with regard to the first survey (Kassin, Ellsworth, & Smith, 1994), it appears that our experts were highly responsive to changes—and nonchanges—in the status of the scientific literature.

Propositions not previously tested. One of the main purposes of the present study was to assess expert opinion on important and recent eyewitness phenomena that were not previously tested. As described in Table 1, we tested 13 new propositions (Items 18–30). Table 4 shows that, in this subset of items, 6 were viewed by at least 80% of experts as reliable enough to be presented in court: that eyewitness confidence is malleable and influenced by factors unrelated to accuracy (95%), that exposure to mug shots of a suspect increases the likelihood of his or her selection from a subsequent lineup (95%), that young children are more vulnerable than adults to suggestion and other social influences (94%), that alcohol impairs eyewitness performance (90%), that eyewitnesses find it relatively difficult to identify members of a race other than their own (90%), and that the risk of false identification is increased by the use of a simultaneous as opposed to sequential

Table 3
Distribution of Reliability Judgments for the 30 Propositions

Topic	1	2	3	4	5	6	7
1. Stress	2	2	13	17	19	4	5
2. Weapon focus	0	0	3	14	27	15	4
3. Showups	3	3	6	10	16	15	10
4. Lineup fairness	5	5	4	7	15	11	14
5. Lineup instructions	0	0	0	3	17	36	5
6. Exposure time	0	1	7	11	14	25	2
7. Forgetting curve	1	4	5	7	18	24	3
8. Accuracy–confidence	1	1	3	15	16	24	1
9. Postevent information	1	0	2	2	16	42	0
10. Color perception	0	0	1	2	7	10	43
11. Wording of questions	0	0	0	1	14	48	0
12. Unconscious transference	0	1	3	20	16	22	1
13. Trained observers	1	1	29	13	2	1	17
14. Hypnotic suggestibility	1	0	4	5	22	26	5
15. Attitudes and expectations	0	0	0	10	26	27	0
16. Event violence	8	7	17	11	5	1	14
17. Confidence malleability	0	0	0	5	18	37	2
18. Cross-race bias	0	0	1	16	19	25	2
19. Hypnotic accuracy	28	28	5	0	0	0	2
20. Alcoholic intoxication	0	1	1	13	20	19	9
21. Mug-shot-induced bias	0	0	0	12	31	19	1
22. Long-term repression	3	21	28	7	3	0	0
23. False childhood memories	0	2	7	22	14	16	3
24. Discriminability	14	25	15	4	1	2	3
25. Child accuracy	2	11	10	12	17	11	1
26. Child suggestibility	0	0	2	11	22	27	2
27. Description-matched lineup	3	3	4	8	15	10	19
28. Presentation format	1	0	2	6	20	20	13
29. Elderly witnesses	0	6	8	17	11	4	17
30. Identification speed	0	2	7	19	11	2	22

Note. Evaluations of the research evidence were coded as follows: 1 = *the reverse is probably true*, 2 = *no support*, 3 = *inconclusive*, 4 = *tends to favor*, 5 = *generally reliable*, 6 = *very reliable*, and 7 = *I don't know*.

presentation format (81%). Three additional propositions were endorsed by at least two thirds of our experts, these being that identification accuracy is increased by having foils that match the witness's description of the culprit (71%), that young children are less accurate witnesses than adults (70%), and that the memories people recover from childhood are often false or distorted in some way (68%).

Finally, 4 of the new items were patently not endorsed by a majority of experts. On the notion that elderly witnesses are less accurate than younger adults, opinion was split on the "reliable enough" question (50%). Eliciting consensus in a negative direction—that the research evidence was not reliable enough to be presented in court—most of our experts did not endorse the propositions that identification speed is predictive of accuracy (40%), that it is possible to differentiate between true and false memories (32%), or that traumatic experiences can be repressed for many years and then recovered (22%). It is quite clear that respondents made distinctions among the phenomena we had assessed, seeing some but not others as reliable enough for presentation to a judge and a jury.

How Reliable Is "Reliable Enough"?

Over the years, psychologists with opposing views on the role of eyewitness experts in court have debated the question of how clear and convincing a body of research must be before it is worthy enough to be offered in court. To address this issue, we compared each respondent's assessment of each proposition with his or her judgment of whether that proposition was reliable enough for testimony. Table 6 reports percentages of experts who judged each topic to be reliable enough and who would themselves testify about it. Data across respondents and items were combined to identify the percentages of experts who believed that psychologists should testify—and that they would testify—at varying levels of reliability.

This analysis revealed that experts were internally consistent in their assessments. Only 4% were willing to

Table 4
Discrete Judgments and Opinions Concerning the 30 Eyewitness Topics Tested

Topic	Is it reliable?	Would you testify?	Research basis?	Common sense?
Wording of questions	98	84	97	25
Lineup instructions	98	79	95	39
Confidence malleability	95	79	95	10
Mug-shot-induced bias	95	77	97	13
Postevent information	94	83	98	17
Child suggestibility	94	81	100	73
Attitudes and expectations	92	70	94	31
Hypnotic suggestibility	91	76	90	19
Alcoholic intoxication	90	61	76	95
Cross-race bias	90	72	97	65
Weapon focus	87	77	97	34
Accuracy–confidence	87	73	97	5
Forgetting curve	83	73	93	29
Exposure time	81	68	93	97
Presentation format	81	64	93	0
Unconscious transference	81	66	92	19
Showups	74	59	85	30
Description-matched foils	71	48	82	30
Child accuracy	70	59	91	78
Lineup fairness	70	54	78	48
False childhood memories	68	52	87	25
Color perception	63	27	37	41
Stress	60	50	98	37
Older witnesses	50	38	77	66
Hypnotic accuracy	45	34	89	55
Identification speed	40	29	75	61
Trained observers	39	31	76	73
Event violence	37	29	79	14
Discriminability	32	25	89	52
Long-term repression	22	20	87	79

Note. Numbers represent the percentage of experts who responded "yes" to each question. Topics are rank ordered according to their scores on the key question, "Do you think this phenomenon is reliable enough for psychologists to present in courtroom testimony?"

testify when they felt that a body of research was "inconclusive." When they believed there was "no support" for a proposition, 27% said they were willing to testify, presumably to say just that. Respondents were split when, as they saw it, the evidence "tends to favor" the issue or when it "suggests the reverse is probably true" (45% and 44%, respectively). Yet the vast majority of experts were willing to testify when they perceived the research evidence to be "generally reliable" (77%) or, better yet, "very reliable" (91%).

Role of Eyewitness Experts

Central to debates on the nature, scope, and role of expert scientific testimony is the question of whether experts should serve as advocates for a party in the dispute or adopt a more neutral posture in an effort to educate the jury, as one might expect of a court-appointed expert. Thus, we asked whether the primary role of an eyewitness expert is to assist a particular party, educate the jury, or serve some other function. Among our 64 respondents, 49 (77%) said that their primary purpose was to educate the jury, compared with only 3 (5%) who sought to assist a particular party, and 4 (6%) who cited a combination of purposes. An additional 8 (13%) respondents cited other reasons, most notably: to educate trial judges, to train police officers on how to increase identification accuracy while minimizing error, and to influence legal policy on how eyewitness identification evidence should be collected.

A principle criterion for the admission of expert testimony is that it assist the trier of fact. Accordingly, Kassin et al. (1989) proposed a Bayesian-like test: that expert testimony be admitted to the extent that it offers to revise what jurors already believe as a matter of common sense (i.e., by informing them of research findings not intuitively known or correcting misconceptions not supported by research). As one can see in Table 4, respondents made sharp

Table 5
Comparison of Reliability Judgments, 1989 and Present

Topic	Reliable enough to testify? 1989	Present
Stress	71	60
Weapon focus	57	87*
Showups	83	74
Lineup fairness	77	70
Lineup instructions	95	98
Exposure time	85	81
Forgetting curve	83	83
Accuracy–confidence	87	87
Postevent information	87	94
Color perception	66	63
Wording of questions	97	98
Unconscious transference	85	81
Trained observers	59	61[a]
Hypnotic suggestibility	69	91*
Attitudes and expectations	87	92
Event violence	36	37

[a] This item was reverse coded to make the present data comparable with those of 1989.
* $p < .05$ (Significant difference in judged reliability).

distinctions among items along this dimension. For example, some phenomena were seen as falling well within the realm of common sense and as known by the average person (e.g., exposure time), whereas others were seen as highly nonintuitive (e.g., the low accuracy–confidence correlation). In a single question, we asked respondents to speculate as to whether juries were generally better equipped to evaluate eyewitnesses with or without the aid of a competent expert (a no-difference alternative was included). As in the 1989 survey, the result on this item was clear: Sixty-one respondents believed that eyewitness experts generally have a positive impact on juries (95%), and 3 said there was no impact (5%). Not a single respondent thought that juries were adversely affected.

New Developments and Future Issues

To explore new areas in the eyewitness landscape not specifically covered in our survey, respondents were asked if they had ever testified on issues other than those tested. In response to this question, several new areas were mentioned. The most frequently cited issue concerned voice recognition and the extent to which "earwitnesses" can identify people from the sound of their voice (Yarmey, Yarmey, & Yarmey, 1994). Other issues cited concerned the effects on identification accuracy of disguise, retention interval, multiple exposures to a given suspect, cowitnesses, duration estimates, object recognition abilities, distinctive perpetrator characteristics, and use of the cognitive interview with child witnesses.

General Discussion

It has been more than 10 years since the publication of Kassin et al.'s (1989) original survey of eyewitness experts. Since that time, the field has expanded, with more researchers publishing more articles on more performance-relevant variables than in the past (most notably, this literature includes such explosive "new" topics as repressed and recovered memories and suggestibility effects in children). Since that time, parts of the American legal system—following the U.S. Supreme Court's opinions in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993) and *Kumho Tire Co., Ltd. v. Carmichael et al.* (1999)—have altered the criteria that they use to admit scientific expert testimony. Since that time, researchers have identified dozens of settled cases in which innocent people convicted of capital and noncapital crimes have been exonerated by DNA testing (Connors et al., 1996; Scheck et al., 2000). Also since that time, the U.S. Department of Justice—with primary input from research psychologists—published guidelines for law enforcement on how to minimize error in the collection and preservation of eyewitness identification evidence (Technical Working Group for Eyewitness Evidence, 1999; see Wells et al., 2000). In short, the eyewitness landscape has changed a great deal in recent years, making it necessary to update judges, lawyers, and psychologists on the consensus that exists within the scientific community.

The present study revealed some important consistencies and some changes in the opinions of eyewitness experts. Like those sampled in 1989, our respondents judged many eyewitness phenomena as reliable enough for presentation in court. As indicated by an agreement rate of at least 80%, there was a strong consensus, in descending order, on the reliability of the following 10 originally tested propositions: the wording of questions, lineup instructions, postevent information biases, attitudes and expectations,

Table 6
Percentages of Experts Who for Each Assessment of Reliability Judged the Topics to Be Reliable Enough and Would Themselves Testify

Opinion on reliability	Judgment: Reliable enough?	Would you testify?
Reverse is probably true	56_a	44_a
No support	33_b	27_b
Inconclusive	7_c	4_c
Tends to favor	66_a	45_a
Generally reliable	97_d	77_d
Very reliable	100_e	91_e

Note. Within each column, percentages without a common subscript differ at $p < .05$.

hypnotic suggestibility, the accuracy–confidence correlation, weapon focus, the forgetting curve, exposure time, and unconscious transference (by a two-thirds margin, respondents also endorsed the item on showups). Aware of advances in the literature, our respondents, as a group, viewed two propositions more favorably than in the 1989 survey—those pertaining to hypnotic suggestibility and weapon focus. As before, however, they clearly did not endorse the items pertaining to stress, event violence, hypnotic accuracy, observer training, and color perception under monochromatic light (this last item elicited many "I don't know" responses). In short, our respondents were discriminating in their judgments and responsive to changes over time in the research evidence.

We also examined 13 eyewitness propositions that were not previously tested but that had drawn a great deal of recent attention. Being the first expert opinion data collected on these issues, the results were highly informative. Overall, six items were judged reliable by at least 80% of respondents. In order of their endorsement rates, these pertained to the malleability of confidence, exposure to mug shots, suggestibility of young children, alcoholic intoxication, the cross-race bias, and simultaneous versus sequential presentation format. By a lesser two-thirds consensus, respondents also judged as reliable the propositions that lineup foils should match the witness's description of the culprit, that young children are less accurate than adults, and that recovered childhood memories are often false. Again indicating that experts were discriminating in their opinions, most agreed that long-term repression and recovery and the related proposition that it is possible to differentiate true and false memories are not reliable enough for presentation in court. This result represents the first poll of eyewitness experts on this controversial topic (in this regard, it would be important to sample clinical psychologists whose perspectives do not emanate from the eyewitness area). In contrast to the consensus obtained on these various propositions, there was very little agreement (i.e., opinions were relatively split) on the new items pertaining to older witnesses and the correlation between identification speed and accuracy.

In terms of how eyewitness experts perceive and manage their activity in the courtroom, three sets of results indicate that they set relatively high standards for their own involvement. First, by examining the association between each respondent's perceptions of reliability and his or her willingness to testify, we found that most said they would agree to testify only on propositions they considered "generally reliable" and "very reliable." Second, across all propositions, respondents were far more likely to see a phenomenon as reliable enough for testimony (overall M = 19.60) than to indicate a personal willingness to testify (overall M = 15.54). Third, respondents were discriminating in their decisions to serve as expert witnesses. In comparison with the 1989 results, they were called on to testify more often, but they were less likely to agree to do so (the agreement rates were 74% in 1989 and 41% in the present sample). Of 49 respondents who had been asked to testify, every single one had declined the opportunity on at least one occasion—most often for moral or ethical reasons, a lack of time, or a perceived incompetence on the relevant issues.

There are two potential limitations of the present survey. One is that additional research published in years to come will inevitably force experts to revise at least some of their current assessments. Comparisons between the 1989 survey and our own offer the case in point. In the initial survey, for example, only 57% saw the weapon focus effect as reliable enough for expert testimony. Yet in light of the publication of later studies and a meta-analysis, 87% of our respondents judged this proposition in the affirmative. In a related vein, new research developments will make it necessary to add topics to be surveyed. Kassin et al. (1989) solicited opinions on 19 basic propositions. Our questionnaire contained 30 items, including 13 that were new. At a later time, it will similarly become necessary to further expand the scope of this inquiry.

It could be argued that the sampling of experts in this survey constitutes a second limitation. To identify our population of respondents, we sought individuals active in the eyewitness area, which raises an issue that has long plagued the *Frye* test, namely, that individuals with the most expertise in an area may also have the greatest motivation to present it in a favorable light. Thus, in *Kumho Tire Co., Ltd. v. Carmichael et al.* (1999), the Supreme Court anticipated that a trial judge might at times exclude proffered expert testimony because the discipline lacks the potential for reliability—despite a consensus within the community of experts (the only disciplines the Court singled out as examples were astrology and necromancy). This possible confounding of expertise and motivation implies that perhaps our respondents should have been drawn from a broader population of basic experimental psychologists who study noneyewitness processes or do not testify in court.

Such an approach may seem reasonable on its face, but it would create a different and serious problem, namely, that a broader sample of experimental psychologists would lack a necessary degree of familiarity with parts of the literature. Over the past quarter century, eyewitness research has become a specialized and highly technical area of application built on the shoulders of the mock witness paradigm and a focus on correctable "system variables" (Wells, 1993). As a result, there are many essential nuts-and-bolts concepts (e.g., simultaneous and sequential presentation formats, valid vs. blank lineups, biased vs. unbiased instructions, and description-matched vs. suspect-matched foils) that are simply not known to the generalists among us.

Is there any evidence to suggest that our eyewitness experts comprised a *Frye* sample tainted by self-motivated interest? We do not think so. Indeed, close inspection of the data offers four bases of reassurance. The first and simplest is that 28 (44%) of our respondents had never testified in court and so were not motivated in this regard (many were asked and declined the opportunity). Second, there is no

support for a stereotype that some might hold—that a caricature-like distinction exists between pure scientists and forensic consultants. Thus, when we correlated the total number of publications per respondent with his or her courtroom experience, we found that very few fit the pure "scientist" or "consultant" profile. To the contrary, numbers of publications were positively—not negatively—associated with the number of times experts were asked to testify, $r(56) = .38, p < .002$; had agreed to testify, $r(56) = .48, p < .001$; and had actually testified, $r(56) = .53, p < .001$. Third, as we reported earlier, we conducted an internal analysis to test the motivated self-interest hypothesis that experts who frequently testify in court, compared with those who do not, would endorse more of the 30 items surveyed as "reliable enough for psychologists to present in courtroom testimony." In fact, there was no significant difference—a result that was corroborated by the more specific comparison of the 35 respondents who had testified at least once with the 28 who had never testified (Ms = 19.21 and 19.83, respectively). Fourth, although individual respondents disagreed in their assessments of certain principles, and although some were generally more demanding than others, the group as a whole was by no means uniformly accepting of all principles. Respondents discriminated quite reasonably among statements for which there was a wealth of experimental support and those for which there was not. And on an item concerning perceptions of color under monochromatic light, 67% of respondents—very few of whom were trained in vision—candidly admitted, "I don't know."

The present results should provide needed guidance to judges (in their decision making at suppression hearings, trials, *Daubert* hearings, and on appeal), lawyers (in their examination of eyewitnesses and police who collect identification evidence), and psychological experts (who must determine which phenomena are reliable enough to present in court). The information contained herein should thus help to sharpen the direct and cross-examinations of eyewitness experts and shape the content of their testimony so that it more accurately reflects opinions within the scientific community. From a paradigmatic standpoint, similarly conducted expert surveys in other psychological domains (e.g., expert testimony on rape trauma and other abuse syndromes, the polygraph, parental competence in custody disputes, legal insanity) also have the potential to make important practical and forensic contributions.

REFERENCES

Berger, M. A. (2000). The Supreme Court's trilogy on the admissibility of expert testimony. In the Federal Judicial Center (Ed.), *Reference manual on scientific evidence* (2nd ed., pp. 9–38). Washington, DC: Federal Judicial Center.

Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom: A scientific analysis of children's testimony*. Washington, DC: American Psychological Association.

Connors, E., Lundregan, T., Miller, N., & McEwen, T. (1996). *Convicted by juries, exonerated by science: Case studies in the use of DNA evidence to establish innocence after trial*. Washington, DC: References U.S. Department of Justice.

Cutler, B. L., & Penrod, S. D. (1995). *Mistaken identification: The eyewitness, psychology, and the law*. New York: Cambridge University Press.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786 (1993).

Devenport, J. L., Penrod, S. D., & Cutler, B. L. (1997). Eyewitness identification evidence: Evaluating commonsense evaluations. *Psychology, Public Policy, and Law, 3*, 338–361.

Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology, 67*, 818–835.

Faigman, D. L., Kaye, D. H., Saks, M. J., & Sanders, J. (1997, Supp. 2000). *Modern scientific evidence: The law and science of expert testimony*. Eagan, MN: West.

Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

Kassin, S. M., & Barndollar, K. A. (1992). On the psychology of eyewitness testimony: A comparison of experts and prospective jurors. *Journal of Applied Social Psychology, 22*, 1241–1249.

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.

Kassin, S. M., Ellsworth, P. C., & Smith, V. L. (1994). Deja vu all over again: Elliott's critique of eyewitness experts. *Law and Human Behavior, 18*, 203–210.

Kassin, S. M., Rigby, S., & Castillo, S. R. (1991). The accuracy–confidence correlation in eyewitness testimony: Limits and extensions of the retrospective self-awareness effect. *Journal of Personality and Social Psychology, 61*, 698–707.

Kirsch, I., & Lynn, S. J. (1995). Altered state of hypnosis: Changes in theoretical landscape. *American Psychologist, 50*, 846–858.

Kumho Tire Co., Ltd. v. Carmichael et al., 526 U.S. 137 (1999).

Leippe, M. (1995). The case for expert testimony about eyewitness memory. *Psychology, Public Policy, and Law, 1*, 909–959.

Lindsay, D. S., Read, D. J., & Sharma, K. (1999). Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. *Psychological Science, 9*, 215–219.

Lindsay, R. C. L., Lea, J. A., Fulford, J. A. (1991). Sequential lineup presentation: Technique matters. *Journal of Applied Psychology, 76*, 741–745.

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

Luus, C. A. E., & Wells, G. L. (1994). The malleability of eyewitness confidence: Co-witness and perseverance effects. *Journal of Applied Psychology, 79*, 714–723.

Malinoski, P. T., & Lynn, S. J. (1999). The plasticity of early memory reports: Social pressure, hypnotizability, compliance, and interrogative suggestibility. *International Journal of Clinical and Experimental Hypnosis, 47*, 320–345.

McConkey, K. M., & Sheehan, P. W. (1995). *Hypnosis, memory, and behavior in criminal investigation*. New York: Guilford Press.

Penrod, S. D., & Cutler, B. (1995). Witness confidence and accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law, 1*, 817–845.

Penrod, S. D., Fulero, S. M., & Cutler, B. L. (1995). Expert psychological testimony on eyewitness reliability before and after *Daubert*: The state of the law and the science. *Behavioral Sciences and the Law, 13*, 29–259.

Pezdek, K., & Banks, W. P. (Eds.). (1996). *The recovered memory/false memory debate*. San Diego, CA: Academic Press.

Pickel, K. L. (1999). The influence of context on the "weapon focus" effect. *Law and Human Behavior, 23*, 299–311.

Poole, D. A., & Lamb, M. E. (1998). *Investigative interviews of children: A guide for helping professionals*. Washington, DC: American Psychological Association.

Read, J. D., & Lindsay, S. D. (Eds.). (1997). *Recollections of trauma: Scientific evidence and clinical practice*. New York: Plenum.

Robinson, M. D., & Johnson, J. T. (1999). How not to enhance the confidence–accuracy relation: The detrimental effects of attention to the identification process. *Law and Human Behavior, 22*, 409–428.

Ross, D. F., Read, J. D., & Toglia, M. P. (Eds.). (1994). *Adult eyewitness*

testimony: Current trends and developments. New York: Cambridge University Press.

Scheck, B., Neufeld, P., & Dwyer, J. (2000). *Actual innocence: Five days to execution and other dispatches from the wrongly convicted*. New York: Doubleday.

Shaw, J. S., III. (1996). Increases in eyewitness confidence resulting from postevent questioning. *Journal of Experimental Psychology: Applied, 2*, 126–146.

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78*, 22–33.

Sporer, S. L., Malpass, R. S., & Koehnken, G. (Eds.). (1996). *Psychological issues in eyewitness identification*. Mahwah, NJ: Erlbaum.

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

Steblay, N. M. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior, 16*, 413–424.

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.

Stinson, V., Devenport, J. L., Cutler, B. L., & Kravitz, D. A. (1997). How effective is the motion-to-suppress safeguard? Judges' perceptions of the suggestiveness and fairness of biased lineup procedures. *Journal of Applied Psychology, 82*, 211–220.

Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement*. Washington, DC: National Institute of Justice.

Thompson, C. P., Herrmann, D. J., Read, J. D., Bruce, D., Payne, D. G., & Toglia, M. P. (1998). *Eyewitness memory: Theoretical and applied perspectives*. Mahwah, NJ: Erlbaum.

United States v. Amaral, 488 F. 2d 1148 (9th Cir. 1973).

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist, 48*, 553–571.

Wells, G. L., & Bradfield, A. L. (1998). "Good, you identified the suspect": Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology, 83*, 360–376.

Wells, G. L., & Bradfield, A. L. (1999). Distortions in eyewitness recollections: Can the postidentification feedback effect be moderated? *Psychological Science, 10*, 138–144.

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*, 581–598.

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 and Human Behavior, 22*, 603–648.

Yarmey, A. D., Yarmey, A. L., & Yarmey, N. J. (1994). Face and voice identifications in showups and lineups. *Applied Cognitive Psychology, 8*, 453–464.

Yuille, J. C., & Tollestrup, P. A. (1990). Some effects of alcohol on eyewitness memory. *Journal of Applied Psychology, 75*, 268–273.